Short Term Scientific Mission
iV&L COST Action IC1307

Quantity Expressions in Language & Vision

End of Project Scientific Report

16 November 2017

Practical Details

- Applicant: Sandro Pezzelle, PhD Student at CIMeC - University of Trento
- Host: Prof. Jakub Szymanik, University of Amsterdam, iV&L Net MC member
- Period: September, 27th - October, 31st (35 days)
- Requested financial support: EUR 2,500

STMS Schedule

September 28-29 The applicant attended the workshop on "Cognitive Semantics and Quantities" and presented the work "Quantifiers and Proportions in Language and Vision: Insights from Behavioral and Computational Studies" (joint presentation with Raffaella Bernardi). The workshop allowed him to discuss with the other speakers and to get feedback on his work on multi-modal learning of quantifiers

First week The applicant and the hosts (Prof. Szymanik and Dr. Steinert-Threlkeld) had several meetings to think of a project to be started during the STSM. By the end of the first week the applicant started working on building a preliminary dataset and developing the models. The applicant had couple of meetings with some PIs (Prof. Raquel Fernandez), researchers, and PhD students of University of Amsterdam

Second week The applicant kept meeting with the hosts and working on the project. In particular, some decisions about the dataset were taken and some implementation issues were solved. During this week, the applicant had the opportunity to attend two talks by Luciano Serafini and Shane Steinert-Threlkeld

Third week The applicant and the hosts got preliminary results and discussed them. They took some further decisions regarding the dataset and the models. During this week, the applicant attended a talk by Ravi Shekhar and Elia Bruni

Fourth week On Tuesday, October 24, the applicant gave the talk "Learning to Quantify from Language and Vision: Insights from Behavioral and Computational Studies" at the Computational Linguistics seminars in Science Park, ILLC, Amsterdam. He presented some his previous work together with motivation, aim, and preliminary results of the ongoing project in collaboration with Prof. Szymanik and Dr. Steinert-Threlkeld. A fruitful discussion followed. Within the same week, he had a meeting with Prof. Rick Nouwen in Utrecht and he attended a "Mini-Symposium on Deep Generative Models and Unsupervised Machine Learning" (keynote speakers: Aapo Hyvarinen, Yoshua Bengio, David Blei)

October 30-31 The applicant and the hosts had couple of meetings to discuss the most recent outcomes of the project and to plan the next steps. They agreed on some conference deadlines to aim to.

1 Purpose of STSM and Achieved Goals

Overview This STSM allowed the applicant (2nd-year PhD Student at CLIC Lab - CIMeC, University of Trento) with expertise in Computational Linguistics, Cognitive Sciences, and Computer Vision to visit the Institute for Logic, Language and Computation (Department of Linguistics, University of Amsterdam) and, in particular, Prof. Jakub Szymanik and Dr. Shane Steinert-Threlkeld working within the "Cognitive Semantics and Quantities" (CoSaQ) ERC project. The research topics explored by the applicant during the scientific mission were related to the iV&L Net working group "Integrated Modeling of Language and Vision" (WG1).

Purpose The purpose of this STSM was to merge the background of the applicant with the hosts' expertise in Logic, Formal Linguistics, and Cognitive Semantics to explore research issues involving theoretical aspects of quantifiers (*some*, *most*, *all*) from a computational perspective. The motivation, described in detail in the STSM proposal, was that quantifiers, though widely used in everyday communication by speakers of almost every language, are only partially investigated in Computational Linguistics, Computer Vision, and work combining Language & Vision. A proper extraction, evaluation, and learning of quantity information from texts and images, however, seems to be as needed as challenging, as it emerges, for instance, by the poor performance of state-of-the-art Visual Question Answering (VQA) models [Antol et al., 2015] on the so-called *count* questions (e.g. "How many children are wearing hats?"). The shared starting point between the applicant and the host was that any computational model aimed at reproducing speakers' behavior should take into account 'fuzzy' quantification (i.e. quantifiers) besides exact numbers. Moreover, cognitive aspects should be integrated to ensure the plausibility of these models besides their performance.

Achieved Goals Overall, the 35-day STSM allowed the applicant to successfully achieve most of the desired goals: (a) Profiting from the formal background of the host's research group to deepen his knowledge on formal, cognitive and logical frameworks on quantifiers; (b) Presenting and getting feedback on his PhD thesis involving the learning of quantities from language and vision; (c) Designing, implementing and presenting to an expert audience the project started in the host institution; (d) Actively participating in the academic life of the host institution (attending seminars and symposia, giving talks and presentations, scheduling collective and individual meetings).

2 Project

The applicant and the hosts designed a long-term project involving the learning of different classes of quantifiers (e.g. Aristotelian, proportional, numerical) in language and vision. In particular, the project includes a first step focusing on language, a second step focused on vision, and a possible final step focused on merging linguistic and visual representations. The overall idea is that the meaning of different classes of quantifiers might depend more/less on contextual information, either linguistic or visual. For example, it might be conjectured that the use of Aristotelian quantifiers (*some*, *all*) is not heavily dependent on the context, whereas the use of e.g. proportional ones (*less than half*, *many*) would be affected by context to a greater extent. Given the time constraints, during the STMS we decided to start from exploring the issue in language, which we believe is the necessary first block of the project. In what follows, we thus describe the project of modelling quantifiers in the linguistic modality.

2.1 Overview

Background Generalized Quantifier Theory (hence, GQT) [Barwise and Cooper, 1981, Keenan and Stavi, 1986, Van Benthem, 1986] aims at devising a general semantics for expressions of quantity by applying mathematical (or generalized) quantifiers to linguistics. Quantifier meanings are defined set-theoretically by means of categorical evaluation functions yielding either truth or falsity of a sentence in which a quantifier is present. The core idea is that a quantifier like 'some' or 'many' expresses a relation between two sets. To illustrate:

1. some(A, B) is true iff $\|A\| \cap \|B\| \neq \varnothing$
2. many(A, B) is true iff $\|A\| \cap \|B\| > n$, where n is some large number

That is, the sentence 'some donkeys fly' is true if and only if the intersection of the donkeys and the flying things is not empty. That implies that the sentence always holds truth except in the case when no donkeys fly. In other words, it is true either when only one donkey out of all donkeys in the world can fly or when all of them do. In the case of 'many donkeys fly', the sentence is true if the cardinality of the flying donkeys is larger than some contextual norm n. In formal semantics, there exists an extensive literature on quantifiers whose meanings depend on such a contextual norm, like 'few' and 'many' [Partee, 1989, Solt, 2009]. Partee [1989], for example, proposes that 'few' and 'many' are ambiguous because of the nature of n, which can stand for either a *cardinal* or a *proportion*. The idea is further extended and formalized by Solt [2009] in terms of 'scale' structures. In a nutshell, the cardinal reading would arise when the involved scale does not display a clear upper bound (hence, the scale is numerical). In contrast, the presence of an upper bound would license the proportional reading (hence, the scale is made of proportions).

Motivation & Goal In this work we aim at exploring whether different types of quantifiers are more/less dependent on contextual information, as claimed in previous theoretical work. To test this hypothesis, we investigate whether quantifiers can be predicted on the basis of linguistic context, based on the assumption that quantifiers might express relations between co-occurrences or expectancies. Since such relations might be well described by models of Distibutional Semantics, we hypothesize that if some quantifiers depend more on context than others do, we should observe a different performance of computational models in predicting the correct quantifier given a sentence where the originally embedded quantifier has been removed. To illustrate, predicting *few* in the sentence "_____ of us have the discipline to exercise alone" should be more challenging compared to when the context is "_____ of us have the discipline to exercise alone, so joining a health club or gymnasium has obvious advantages". Moreover, the use of a quantifier like *some* should be affected by the surrounding context to a lesser extent.

2.2 Implementation

Dataset The first step was select a set of quantifiers to be explored. Based both on theoretical considerations and corpus frequency, the following 9 quantifiers were chosen: *none*, *a few*, *few*, *some*, *many*, *most*, *more than half*, *almost all*, *all*. We then built a dataset of sentences as extracted from a big corpus of written English, i.e. the concatenation of ukWac, BNC and a

2009-dump of Wikipedia [Baroni et al., 2009]. To avoid undesired effects due to morpho-syntactic factors (e.g. the intensifier "very" preceding only some quantifiers, but not others), we decided to select only sentences starting with the given quantifiers, e.g. "All of the members attended the meeting". Moreover, all quantifiers were taken into account in their partitive form, namely quantifier + *of* (e.g. *some of*, *many of*, etc.). This choice was aimed at discarding cases where the selected expressions do not actually act as quantifiers, or they precede uncountable/mass nouns. After these filtering steps, we randomly selected a balanced number of sentences for each of the 9 quantifiers. The target quantifier in each sentence was removed, so to obtain sentences of the type: "the members attended the meeting". In total, $12.5K$ sentences were included in the dataset. The dataset was further split into training ($10K$ datapoints), validation (1250) and testing (1250) set.

Models The problem of predicting the correct quantifier was addressed as a classification task. Given a sentence and 9 classes, the models have to provide the class that stands for the correct quantifier. Several models were implemented and tested by the applicant during the STSM, namely (1) bag-of-word (BoW) MLP classifier with summed vectors; (2) bag-of-word (BoW) MLP classifier with concatenated vectors; (3) Long Short Term Memory (LSTM) network; (4) bilinear LSTM (bi-LSTM); (5) fastText classification model by Joulin et al. [2016]. The most important parameter we investigated was the size of the linguistic context, i.e. the number of words given to the network to perform the task. We experimented with (a) the median number of words in the sentences, i.e. up to 20 words, (b) up to 10 (small context) and up to 50 words (big context). A fairly intensive ablation study was performed to explore different optimizers (e.g. SGD, adam, nadam, etc.), parameters (e.g. dropout), hidden layer size, etc. For each model, the best weights for the validation set were saved and used to obtain results in the test set. The best weights were chosen based on the lowest validation loss.

2.3 Results and Discussion

Table 1 reports the accuracy for each model in the 3 settings. As can be seen, the LSTM model turns out to be overall the best one, which makes sense given that the feed the network only with linguistic context following the target quantifier. We could have probably expected a higher accuracy for the bi-LSTM in case we had included context from the previous sentence, that is something that we definitely want to test in the next steps of the project. In general, the task of predicting the correct quantifier is quite hard for any model. By zooming into the performance obtained for each quantifier, we notice that some quantifiers are indeed easier to be guessed than others (e.g. *more than half*). However, we believe that (a) experimenting with context either preceding and following the quantifier would be necessary to better understand the impact of

accuracy (test)			
	10-word	20-word	50-word
chance	0,1111	0,1111	0,1111
MLP-sum	0,3141	0,3029	0,3309
MLP-conc	0,3173	0,2813	0,2693
LSTM	**0,3549**	**0,3581**	**0,3629**
biLSTM	0,3461	**0,3581**	0,3605
fastText	0,352	0,352	0,349

Table 1: Accuracies of the various models on the task of predicting the correct quantifier. Values in **bold** are the highest per each column.

the context size. Also, we believe that (b) setting a minimum number of words to be fed into the model would be beneficial to avoid including sentences made up by just 2 or 3 words.

2.4 After the STMS

The applicant and the hosts planned to continue working on the project during the next months. They planned to have weekly updates and agreed on a few possible conference deadlines where to seek to submit the work. The next steps also include starting a crowdsourcing experiment aimed at collecting human judgements on the same task.

References

Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. Vqa: Visual question answering. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2425–2433, 2015.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3):209–226, 2009.

Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics and philosophy*, 4(2):159–219, 1981.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*, 2016.

Edward L Keenan and Jonathan Stavi. A semantic characterization of natural language determiners. *Linguistics and philosophy*, 9(3):253–326, 1986.

Barbara Partee. Many quantifiers. In *Proceedings of the 5th Eastern States Conference on Linguistics*, volume 5, pages 383–402, 1989.

Stephanie Solt. *The semantics of adjectives of quantity*. PhD thesis, City University of New York, 2009.

Johan Van Benthem. *Essays in logical semantics*. Springer, 1986.