PM1/2, SoSe 19, Semantics with Pictures, Uni Potsdam, David Schlangen

Some Intermediate Projects (Week 3–4)

These are just some suggestions, you can chose other topics if you have a good idea.

Note that many of these tasks may involve some further processing steps, like running a parser over the expressions to identify head nouns, etc. Make use of the many great toolkits that are available! (spacy, nltk, etc.)

Analysis Tasks

- evaluate canonicalisation in visual genome. How good is the mapping to WordNet?
- make use of having category information (through coco categories or vg WordNet normalisations) and expressions. What is the range of expressions. Head nouns. For which categories is it large, for which is it small?
- with the perspective on generation, you can ask what constrained the choices that the speakers made.
    - Why did they decide to use a relational construction (getting at the target object via a landmark), or decide not to do that?
    - Why did they call this an animal and not a dog?
    - How do you perceive that someone is someone else's friend?

  Here you could first think about what such choice points could be, then search for instances of the different choices having been made, think about what contextual factors may explain these choices, and then perhaps even try to learn a model that predicts these choices.
- analyse visual genome relations. Are there ever real "part of" relations? (Like, "leg of the giraffe".) Or are these mostly spatial prepositions?
- investigate spatial language using Visual Genome.
- Investigate the "closed world assumption" that I have mentioned. That is, can be conclude that if an object is not associated with a certain expression, than that expression is not applicable to that object. This is not something that was enforced when collecting these corpora, but maybe it is something that naturally happened to be the case, at least for some types of expressions?
- How reliable is the plural annotation in Visual Genome? (One phrase resolved to multiple entities.)
- analyse vocabulary (e.g., in referring expressions) for abstractness / concreteness, using Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods, 46(3), 904–911. http://doi.org/10.3758/s13428–013–0403–5
- analyse vocabulary (e.g. in referring expressions) for age-of-acquisition, using Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. Behavior Research Methods. http://doi.org/10.3758/s13428–012–0210–4
- analyse for modality exclusivity: Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. Behavior Research Methods. http://doi.org/10.3758/s13428–012–0267–0
- Analyse "complexity" of the language, using some established metric. Or some measure of comparison with written language. What is difference between referring expressions for example and captions?
- quick calculation: How many years of interaction does our data represent? If you assume that one referential interaction (A refers to something for B, B recognises it) takes n minutes, and that you can do m of those per day, how many days / months / years worth of interaction do our corpora represent? Try to come up with a reasonable estimate. (Or an argument for why that doesn't make sense.)

Processing tasks

- Add NLVR2: Suhr, A., Zhou, S., Zhang, I., Bai, H., & Artzi, Y. (2018). A Corpus for Reasoning About Natural Language Grounded in Photographs. ArXiv.

- Add the CHILDES data: the dialogue annotated by [1] and augmented with (somewhat dubious) images by [2] (which I can give you). [1] Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. Psychological Science, 20(5), 578–585. http://doi.org/10.1111/j.1467–9280.2009.02335.x [2] Lazaridou, A., Chrupała, G., Fernández, R., & Baroni, M. (2016). Multimodal Semantic Learning from Child-Directed Input. In NAACL 2016. – NLTK already has a reader for CHILDES, maybe use that (instead of our format)? http://www.nltk.org/howto/childes.html
- Unify objects, based on overlap of bounding boxes (intersection over union). E.g., visual genome seems occasionally to have more than one bounding box (and hence, more than one image ID) for the same image object, where the bounding boxes may even be slightly differently.
- You can also try to unify object bounding boxes across corpora. Relate the VG bounding boxes to those in COCO (for the images in the intersection of the two datasets).