Frequencies of apparent Hurford constraint obviations track implication rates·

Emmanuel Chemla, CNRS

According to Hurford (1974), a disjunction is not felicitous if one of its disjuncts entails the other: (1). There are systematic obviations to this constraint (Gazdar 1979). Effectively, it has been proposed that a sentence such as (2) is felicitous, despite the fact that *X*="reading some books" is entailed by *Y*="reading all books", because *X* can be strengthened and interpreted as *X and not-Y*="reading some but not all books" (even locally through scalar implicatures, as Chierchia, Fox & Spector 2013 have it), see (3)/(3').

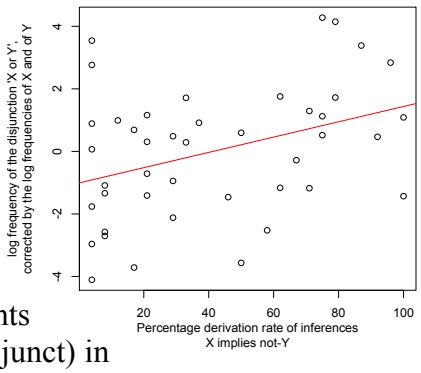
(1) John is in France or in Paris.
(2) John read some or all of the books. (2') X or Y
(3) John read some of the books. → John did not read all the books. (3') X → not-Y

If the obviation indeed is due to the possibility for X to imply not-Y, then **we expect the frequency of (2') to co-vary with the derivation rate of the inference in (3').**

Propitiously, van Tiel et al. (2013) present data revealing important variability in the derivation rate of inferences such as (3)/(3') for different *(X,Y)* pairs. For each *(X,Y)* investigated in van Tiel et al., we want to compare X→not-Y inference rates with disjunction frequencies. We therefore collected a noisy estimate of the frequencies of *X*, of *Y* and of the disjunction *X or Y*, as the number of hits obtained from a google search of these elements (between quotation marks).

Technically, we first ran a linear model by which the log-frequency of the disjunction *X or Y* is predicted by the log-frequencies of *X* and of *Y*. We used the residuals obtained from this model as a corrected frequency of the disjunction. Crucially, we ran a second model to see whether the derivation rate of the inference (3'), as reported by van Tiel et al. accounts for some of the remaining variability in this corrected frequency of the disjunction.

We obtain a significant correlation (r^2*=.14, F(1,41)=6.5, p=.015*, see plot): the more participants are willing to derive an inference *X→not-Y* (as measured in van Tiel et al.'s task), the more the corresponding *X or Y* disjunction generates google hits.[1]

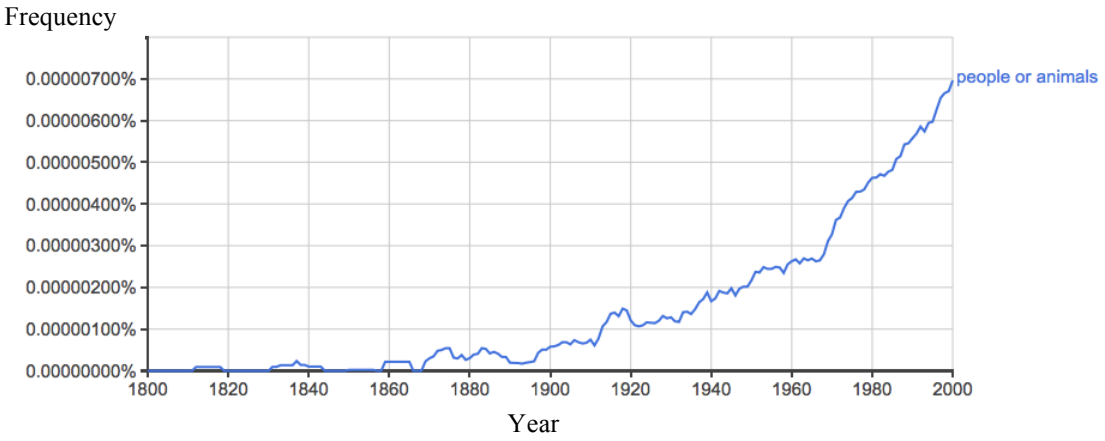


This result provides further evidence for Hurford's constraint, such that disjuncts should not entail each other (or maybe even: should be incompatible). Apparent obviations appear as undercover reinforcements of *X* into *X and not-Y,* which may occur within a sentence (the first disjunct) in a comparable way as they occur at the sentential level.

Methodological coda. We may use corpora to do semantics/pragmatic in a new way: **the frequency of *X or Y* is indicative of the implication relation between X and not-Y**. This result delivers new tools to study historical semantics and pragmatics. Google ngram (Michel et al 2009) is available online and allows the extraction of the frequency of phrases at different point in time. The graph below seems to demonstrate that (although the frequencies of these words are flat during that same period) "people" stopped thinking they are "animals" in the course of the XXth century.

· Thanks to Danny Fox, **Roger Levy**, Raj Singh, Philippe Schlenker, Benjamin Spector and **Kristen Syrett**. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610 and was supported by ANR-10-IDEX-0001-02 and ANR-10-LABX-0087.

[1] Worrying about the use of Google hits as frequency estimates, Roger Levy generously shared similar analyses of two other corpora (the Google Web 1T ngram dataset and the ngram datasets associated with the Google Books project) yielding *p*-values of .15 and .14 (dropping "X or Y" disjunctions that did not occur). The direction of the effect is replicated, although below significance level. More tormented meta-analyses yield *p*-values between .006 and .07 (methods: (i) successive applications of Bayesian regressions to the different datasets, assuming independence, using as priors the posteriors of the previous regression, (i') excluding Google hits from this stepwise analysis, (ii) analyzing one of Levy's corpora using Google hits to set the priors for the slope at .025, although sticking to a conservative standard deviation twice or ten times as large of .05 or .25).



Chierchia, G., Fox, D. & Spector, B. (2013). The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics. In P. Portner, C. Maienborn, and K. von Heusinger (eds), Handbook of Semantics. Mouton de Gruyter, New York, NY.

Gazdar, G. (1979). Pragmatics: Implicature, Presupposition and Logical Form. Academic Press, New York, NY.

Hurford, J. R. (1974). "Exclusive or Inclusive Disjunction" *Foundation of Language*,11: 409-411.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176-182.

van Tiel, B., van Miltenburg, E., Zevakhina, N. & Bart Geurts, B. (2013). Scalar diversity. Ms. University of Nijmegen.