The similarity approach fights back.

Katrin Schulz, ILLC, Amsterdam

Introduction. The central question in the semantics of counterfactual conditionals is the question of the selection function: how to single out the set of relevant antecedent worlds on which the consequent has to be true to make the counterfactual true. The similarity approach ([6], [7]) is one way to answer this question. According to this approach the relevant antecedent worlds are those worlds that are most similar to the actual world. This approach has established itself firmly as paradigm of how to approach conditional sentences in the semantic and philosophical literature.

The challenge. Recently a serious challenge to this approach has been brought forward. [4] and [3] report on an experiment involving a scenario with two switches $S1$, $S2$ and a lamp L. The lamp is on iff the switches are in the same position, up or down. In a situation where both switches are up ($S1$ and $S2$ are true) and the lamp is on (L is true) the majority of the participants take the counterfactuals $\neg S1 > \neg L$, $\neg S2 > \neg L$ and $(\neg S1 \vee \neg S2) > \neg L$ to be true, while they take the counterfactual $\neg(S1 \wedge S2) > \neg L$ to be false or undetermined. A simple similarity approach cannot account for such an observation.[1] The problem seems to be that among the selected antecedent worlds for the counterfactual with antecedent $\neg(S1 \wedge S2)$ are possible worlds where both switches are down (that explains why the participants consider this counterfactual not to be true), while for the counterfactuals with antecedent $\neg S1$ or $\neg S2$ or $\neg S1 \vee \neg S2$ such worlds are not selected (that's why in these cases the counterfactuals are judged as true). Such a selection function cannot be modelled based on an antecedent-independent similarity-relation.

[4] and [3] conclude that we need to give up the paradigm of the similarity approach. [3] proposes a new selection function f that they combine with Inquisitive Semantics ([2]) to account for the data. In Inquisitive Semantics the meaning of a sentence ϕ is not given in terms of truth conditions with respect to possible worlds, but in terms of support conditions with respect to information states. An information state is a subset of the set W of possible worlds. The maximal information states supporting a sentence ϕ are called the alternatives for ϕ, and the set of alternatives is denoted by $Alt(\phi)$. The union of the alternatives of ϕ, $|\phi|$, give the truth conditions of the sentence. While negation and conjunction work in standard ways in Inquisitive Semantics, disjunction introduces a set of alternatives, one for each disjunction. The counterfactual is then proposed to quantify over the alternatives its antecedent gives rise to, see (1).[2] Both ingredients, the new selection function and Inquisitive Semantics work together in [3]'s account of the observations.

$$A > C \text{ iff } \forall \phi \in Alt(A) : f(\phi) \subseteq |C|. \quad (1)$$

We will argue that the critical observation needs to be analysed differently. The problem is not the similarity approach but the semantic meaning assigned to the antecedent. Not only disjunction, but also negation leads to the introduction of alternatives that the counterfactual operator quantifies over. With this modification the selection function can still be similarity-based. We can account for the data, but also overcome certain mis-predictions the new approach of [3] still inherits from standard similarity approaches.

[1] The logic of the similarity approach validates $\neg A > C, \neg B > C \models \neg(A \wedge B) > C$.

[2] For similar proposals see [1] and [5].

The proposal. Pick your favourite similarity-based selection function f and insert it into the interpretation rule for counterfactuals given in (1). We only change the definition of the alternatives Alt by changing the interpretation rule for negation in Inquisitive Semantics.[3] The solution we propose is inspired by standard approaches to truth makers of negations. A truth maker of a formula $\neg\phi$ is a formula φ that contradicts the formula ϕ in question ($\varphi \bot \phi$). We additionally restrict truth makers of negations to relevant sentences/propositions that contradict ϕ. As we are concerned with semantics here, we use a notion of relevance that relies on the sentence itself. Each utterance defines a bottom line for what is relevant by the language it uses. We restrict ourselves here to a propositional language and define $\mathcal{L}(\phi)$ as the set of atomic formula occurring in ϕ. Relevant according to ϕ is to know the truth value of all elements in $\mathcal{L}(\phi)$. In other words, the issue raised by the utterance is $Q(\phi)$, the partition introduced by $\mathcal{L}(\phi)$ (i.e. the set of sets of possible worlds that assign the same truth value to all elements in $\mathcal{L}(\phi)$). For example, if $\phi = p \wedge q$, then $\mathcal{L}(\phi) = \{p, q\}$ and $Q(\phi) = \{pq, p\bar{q}, \bar{p}q, \bar{p}\bar{q}\}$.[4] Any formula using the same vocabulary gives rise to the same issue. We extend support to issues in the standard way: an information state s supports an issue I ($s \models I$) in case s completely answers I, i.e. $\exists i \in I : s \subseteq i$. The new interpretation rule for negation is given in (2). It states that a situation supports $\neg\phi$ in case it's a complete answer to the issue raised by ϕ and contradicts ϕ.

$$s \models \neg\phi \text{ iff } s \models Q(\phi) \text{ and } s \bot \phi. \tag{2}$$

Predictions and Outlook. According to this interpretation, the semantic value of the sentences $\neg p \vee \neg q$ and $\neg(p \wedge q)$ differ: $Alt(\neg p \vee \neg q) = \{\bar{p}, \bar{q}\}$, but $Alt(\neg(p \wedge q)) = \{\bar{p}q, p\bar{q}, \bar{p}\bar{q}\}$. Crucially, the sentence $\neg(p \wedge q)$ contains an additional alternative, $\bar{p}\bar{q}$. When this sentence occurs as antecedent of a counterfactual, also this alternative needs to counterfactually entail the consequent. This accounts for the critical observations in [4] and [3]. We will show that this also helps to account for additional observations concerning counterfactuals with redundant material in the antecedent, which standard similarity approaches, but also [3] cannot account for.

From a more general perspective, this talk is another case-study arguing for more fine grained semantics for natural language. But while so far the focus has been nearly exclusive on disjunction, we argue that such a step is also needed to account for negation. The antecedent of counterfactuals is not the only context in which this becomes evident. As we will discuss in the last part of the talk, similar observations can be made for negation in scalar implicatures and answers to questions. For instance, it has been noted that answers containing negation (like "Not John" to the question "Who came to the party?") receive a non-exhaustive or restricted exhaustive interpretation. The proposed alternatives semantics for negation can account for this observation.

References: [1] Alonso-Ovalle, 2009. Counterfactuals, correlatives, and disjunction. *Linguistics and Philosophy*, 32(2). *** [2] Ciardelli et al., 2013. Inquisitive semantics: A new notion of meaning. *Language and Linguistics Compass*, 7(9). *** [3] Ciardelli et al., ms. *Two switches in the theory of counterfactuals.* *** [4] Champollion et al. 2016. Breaking DeMorgan's law in counterfactual antecedents. *SALT 26.* *** [5] Fine, 2012. Counterfactuals without possible worlds. *The Journal of Philosophy*, 109(3). *** [6] Lewis, 1973. *Counterfactuals.* Blackwell. *** [7] Stalnaker, 1968. A theory of conditionals. *Studies in logical theory: Essays.* Cornman et al. (eds.). Blackwell.

[3] We could as well have used truth makers semantics.

[4] To simplify notation we write $p\bar{q}$ to refer to the set of worlds where p is true and q is false.