

## Presuppositions of the gender features of anaphoric pronouns

**Cooper's treatment of (natural) gender features.** Since [Cooper, 1983], it has become common wisdom to treat gender features (in natural, as opposed to formal, gender languages) as contributing presuppositions about the referent of the pronoun; the entry for a pronoun usually looks like the following (using [Heim and Kratzer, 1998]'s notation for presupposition):

$$(1) \llbracket she_i \rrbracket^g = ( g(i) \text{ is female} . g(i) )$$

However, the original proposal by Cooper was more complex. For free pronouns, Cooper noticed that their gender presupposition cannot be accommodated: in 2, the speaker must believe that the referent of *she* is female; this is not the case for normal presupposition-inducing expressions. To account for that, Cooper uses what he calls indexical presuppositions — presuppositions which must be satisfied in the actual world, and builds them into free pronouns. Bound pronouns are exempt from this special requirement, and are analyzed as inducing regular presuppositions. Thus bound and free pronouns are treated in essentially different ways.

$$(2) \text{ John suspects that } she \text{ left.}$$

**New data.** In this paper I show that the actual data pattern is even more complex. While the judgements are hard, it can be shown that the gender features on pronouns behave differently both from the usual presuppositions and descriptive adjectives or nouns referring to the gender property.

The emerging empirical generalizations are as follows: 1) the gender feature of the pronoun must match the actual gender property of the individual in question in the actual world — if this individual exists in the actual world. (Thus all pronominal gender presuppositions are indexical in Cooper's sense.) 2) if the individual does not exist in the actual world, there is no way to determine its actual-world gender. Yet the choice is constrained — 2A) if there are two embedding intensional predicates introducing two sets of worlds where gender of some individual is different, the gender feature must match the gender in the set of worlds introduced by the least embedded operator, and 2B) If there are two operators which are at the same level of embedding, then one does not interfere with the gender features of another. 3) there are some cases when it appears that those constraints are violated. I argue that it happens when two individuals from different sets of worlds are not viewed as having the same essence; if they are different objects, the closer-to-the-actual-world gender is not imposed on the less-actual individual, as is generally the case when it comes to two distinct individuals. Some illustrations (incomplete for the reasons of space) follow.

Alice mistakenly thinks that the current MIT's President is male. Still, in this situation, knowing that the current President is a woman, I must use a feminine pronoun even when talking about Alice's attitudes, inside of which the President is male (which shows generalization 1 at work):

$$(3) \text{ Alice hopes that the President of MIT will give back to the Institute a fifth of } {}^{OK} \text{ her/*his compensation package this year.}$$

My friend Richard and I go to a bar and see there a person we both find beautiful. Richard thinks that this person is a man, and says to me: "Hey, this man in the corner is gorgeous! I wish I could have his looks"; but I think that this person is a woman, and answer to Richard: "I guess you would have to change your gender identity then, because it is a woman, Richard. But she is beautiful, that's for sure." After that, we argue for about two minutes, but neither Richard nor I can persuade the other that his view of the matter is correct. Now, suppose a day later I tell about that to my other friend Sarah. Even after I have told Sarah about our argument, I will not be able to say the following:

- (4) *I*: \*[That person]<sub>2</sub> looked so gorgeous that Richard<sub>1</sub> wanted to look like him<sub>2</sub>.

However, if I explicitly make clear that the he-person, according to me, existed only in Richard's thoughts (and thus that I talk not about the person I consider actual, but about a different imaginary person), it becomes possible to use a masculine pronoun (which illustrates the generalization 3.)

- (5) *I*: <sup>OK</sup>[That person who Richard invented looking at the girl sitting in the corner]<sub>2</sub> looked so gorgeous that Richard<sub>1</sub> wanted to look like him<sub>2</sub>.

The acceptability of the following example depends on the speaker's beliefs about Carl's beliefs. Suppose that I know that Carl's neighbors are Suzan and Helen, both women. In Scenario 1, Carl also believes that, but mistakenly thinks that Suzan and Helen are male. In Scenario 2, Carl does not believe in the existence of his real neighbors, but instead has a set of imaginary ones, all of which are male. Suppose also that in both scenarios I know about those beliefs of Carl. In both cases, the set of individuals which Carl believes to be his neighbors is all-male. But 6 is not admissible in Scenario 1 and is OK in Scenario 2 (which vindicates generalizations 1 and 2.)

- (6) Carl<sub>1</sub> said [every neighbour of his<sub>1</sub>]<sub>2</sub> told him<sub>1</sub> that he<sub>2</sub> saw a cow yesterday.

**New burden of the semantic system.** The facts outlined above cannot be straightforwardly captured by current semantic systems. There are two main problems: we need to be able to talk about two individuals as being the same one in one sense, and different ones in another (for which I argue there is independent evidence from the ship of Theseus-style natural language discourses); we also need to be able to access the privileged actual-world, or closest-to-the-actual-world, gender property of some individual from inside of the scope of intensional operators.

**Sketch of a new system.** In the paper I sketch a system that possibly can cope with those two challenges. With Lewis's Counterpart Theory as the important predecessor, I build a theory which, unlike Lewis's, has a counterpart relation sensitive to an implicit argument, corresponding to the body of knowledge available for making identity judgements (cf. Kratzer's 2009 treatment of modal bases). Choosing different implicit bases, we get different counterpart relations; thus two objects can be judged counterparts on one base, and not counterparts on another. Furthermore, from the chain of identity statements and intensional embeddings it is possible to deduce, for a given formula, what is the counterpart of a given individual closest to the actual world. Since this information is available in the formula itself (which will never be the case for a theory without identity statements freely present in formulas), such operation is relatively well-behaved.

Of course, it is crucial to show that such a new system can handle the basic semantics for natural language in the first place. While more work is needed to show that in detail, I will outline possible treatments of binding, quantification, de re/de dicto ambiguity, and predicate evaluation-world ambiguity in such a framework.

If my system is viable in the long run, then the gender features data presents a new empirical argument in favor of (a specific version) of counterpart theory for natural language translation.

Another important philosophical consequence bears on the actualist vs. possibilist debate. The result is mixed in a sense: it turns out that the linguistic semantics assigns privileges to the actual world when possible (good for actualists), but when it is impossible, it singles out the closest-to-the-actual non-actual worlds (surprisingly for actualists). The empirical observations presented in the paper may lead to a kind of synthesis: 1) we accept there is an actual world; 2) we hope that we know what it is; 3) yet we give credit to the others' conceptions of what the actual world may be, as opposed to their conceptions about what other possibilities may be (while still preferring ours view on what is actual).