# Synonymy, Common Knowledge, and the Social Construction of Meaning*

Reinhard Muskens

In this paper it is shown how a formal theory of interpretation in Montague's style can be reconciled with a view on meaning as a social construct. We sketch a formal theory in which agents can have their own theory of interpretation and in which groups can have common theories of interpretation. Frege solved the problem how different persons can have access to the same proposition by placing the proposition in a Platonic realm, independent from all language users but accessible to all of them. Here we explore the alternative of letting meaning be socially constructed. The meaning of a sentence is accessible to each member of a linguistic community because the way the sentence is to be interpreted is common knowledge among the members of that community. Misunderstandings can arise when the semantic knowledge of two or more individuals is not completely in sync.

## 1 Introduction

In formal theories of semantics the notion of meaning often seems to be an inherently static descendant from the Platonic world of Forms. As a consequence, semantic relations are predicted to hold once and for all, while divergencies between agents are disallowed. Once, for example, such a theory has established the synonymy of *eye doctor* and *ophtalmologist*, perhaps with the help of a meaning postulate, these expressions must from then on be the same in all contexts and all agents are predicted to believe that John is an ophtalmologist if they believe that he is an eye doctor. It is well-known that such predictions are wrong.

This Platonic view on meaning (inherited from Frege) contrasts with ordinary life where it seems that meaning is something that gets constructed in language communities and between language participants. Is this more pedestrian and earthly perspective compatible with the logical work that has been done so far? And can it contribute to a solution of the well-known problems that the Platonic perspective runs into? In this paper we provide a logical theory of meaning as a social construct that dovetails well with formal semantic

*From: Paul Dekker and Michael Franke (eds.), *Proceedings of the Fifteenth Amsterdam Colloquium*, ILLC, University of Amsterdam, 2005, pp. 161–166.

theories such as Montague's. We will show how a single agent's *knowledge of meaning* can be formalized and how this leads to a formalization of the *common knowledge about meaning relations* of a set of agents or a linguistic community. This common knowledge is then held to constitute the social construction of meaning. If it is common knowledge between language participants that *woodchuck* and *groundhog* are synonymous then these words are treated as such and if any of the participants commits himself to the sentence (say) *woodchucks are fertile* it can be concluded that he is committed to *groundhogs are fertile* as well. It can also be common knowledge within a certain group that (say) *woodchuck* denotes the kind of animal that is normally called a woodchuck and this will enable the members of this group to communicate meaningfully about that animal. But the overall theory that we will propose will in itself make no predictions about such form–meaning relations at all and the theory will not exclude the possibility of agents having misconceptions about denotations or relations of synonymy. If your theory of interpretation diverges from mine, I will consider some of your semantic assumptions to be misconceptions and in our conversations miscommunications may arise.

## 2 A Theory of Propositions

Our point of departure will be the Montague-like theory of propositions proposed in Thomason [4], as streamlined in Muskens [3]. In this theory, each sentence of a given fragment of English is sent to a logical term of a primitive type $p$ (propositions). These logical terms are very close to the syntactic objects they translate. The sentence *Mary is aware that no man talks if a woman walks*, for example, has a translation $((\mathsf{a\ woman})\lambda x(\mathsf{mary}(\mathsf{aware}((\mathsf{if}(\mathsf{walk}\,x))((\mathsf{no\ man})\mathsf{talk})))))$, corresponding to the form of the sentence in which *a woman* has obtained wide scope in some way. Such terms of type $p$ are systematically related to the domain $st$ (sets of possible worlds) with the help of meaning postulates such as the following.[1]

(1) a. $\forall\pi\pi'\tau\tau'(d^0(\pi,\tau)\wedge d^0(\pi',\tau')\rightarrow d^0(\mathsf{if}\ \pi\pi',\lambda i.\tau i\rightarrow\tau' i))$

b. $d^1(\mathsf{man}_{ep}, man_{e(st)})$

Here the $d^n$ ($d$ is for 'determines') are relations connecting objects of type $e^n p$ with objects of type $e^n(st)$. With the help of these and other meaning postulates facts like the following are readily established.

(2) $d^0(((\mathsf{a\ woman})\lambda x(\mathsf{mary}(\mathsf{aware}((\mathsf{if}(\mathsf{walk}\ x))((\mathsf{no\ man})\mathsf{talk}))))),$
$\lambda i.\exists x(woman\ xi\wedge aware((\mathsf{if}(\mathsf{walk}\,x))((\mathsf{no\ man})\mathsf{talk}))\ mary\ i))$

In this particular case the proposition under consideration is associated with a set of possible worlds, namely those in which it is true that there is a woman such that Mary is aware of the proposition that if that woman walks no man

[1] We simplify here for the sake of exposition, but in (3) below similar meaning postulates are given with the generality that is required.

$$\leftarrow \underline{d((\mathsf{if}((\mathsf{a\ woman})\mathsf{walk}))((\mathsf{no\ man})\mathsf{talk}), \tau, j)}$$
$$(3d) \downarrow \tau := \lambda i.\tau_1 i \rightarrow \tau_2 i$$
$$\leftarrow d(((\mathsf{a\ woman})\mathsf{walk}), \tau_1, j),\ \underline{d(((\mathsf{no\ man})\mathsf{talk}), \tau_2, j)}$$
$$(3g) \downarrow \tau_2 := \lambda i.\neg\exists x[P_1 xi \wedge P_2 xi]$$
$$\leftarrow d(((\mathsf{a\ woman})\mathsf{walk}), \tau_1, j),\ \underline{d^1(\mathsf{man}, P_1, j)},\ d^1(\mathsf{talk}, P_2, j)$$
$$(3m) \downarrow P_1 := man$$
$$\leftarrow \underline{d(((\mathsf{a\ woman})\mathsf{walk}), \tau_1, j)},\ d^1(\mathsf{talk}, P_2, j)$$
$$(3f) \downarrow \tau_1 := \lambda i.\exists y[P_3 yi \wedge P_4 yi]$$
$$\leftarrow \underline{d^1(\mathsf{woman}, P_3, j)},\ d^1(\mathsf{walk}, P_4, j),\ d^1(\mathsf{talk}, P_2, j)$$
$$(3m) \downarrow P_3 := woman$$
$$\leftarrow d^1(\mathsf{walk}, P_4, j),\ \underline{d^1(\mathsf{talk}, P_2, j)}$$
$$(3m) \downarrow P_2 := talk$$
$$\leftarrow \underline{d^1(\mathsf{walk}, P_4, j)}$$
$$(3m) \downarrow P_4 := walk$$
$$\leftarrow$$

Figure 1: A refutation of $\leftarrow d((\mathsf{if}((\mathsf{a\ woman})\mathsf{walk}))((\mathsf{no\ man})\mathsf{talk}), \tau, j)$. Selected atoms are underlined. Composition of the substitutions that are found gives the value $\tau = \lambda i.\exists x[woman\ xi \wedge walk\ xi] \rightarrow \neg\exists x[man\ xi \wedge talk\ xi]$.

talks. The treatment is hyperfine-grained, for Mary could well be aware of this proposition but not, for example, of its contraposed form. In Muskens [3] it is shown how terms of type $p$ in fact can function as small programs for computing the truth-conditions of the sentences associated with them. This is in line with the senses-as-algorithms view of Moschovakis [2]. The paper also explains that some of these programs may *diverge*. For example, the programs connected with the Liar and the Truth-teller never halt and no truth-conditions are therefore associated with these sentences. They have a sense but no reference.

## 3 Relativizing to Agents and their Common Beliefs

What is important for present purposes is that in Muskens [3] most of the real work of the interpretation process takes place on the *object* level of the interpreting logic. It is the meaning postulates that do the work. This allows for the possibility to make the interpretation process dependent upon agents in the following way. First, we make the $d$ relations *world dependent* by providing them with an extra argument for a possible world. E.g. $d^1(\mathsf{woodchuck}_{ep}, woodchuck_{e(st)}, j)$ will now mean that the predicate *woodchuck* is determined by the propositional function woodchuck *in world* $j$. In (3) the set of meaning postulates considered in Muskens [3] is repeated in a slightly generalized form that takes care of the extra world argument that is now added to the $d$ relations. For all notational conventions and for more general explanation the reader is referred to Muskens [3].

The meaning postulates in (3), in which all free variables are understood to have a universal interpretation, form a *logic program* and therefore combine a declarative interpretation with a procedural one. In Figure 1 a refutation of the *query* $\leftarrow d((\mathsf{if}((\mathsf{a\ woman})\mathsf{walk}))((\mathsf{no\ man})\mathsf{talk}), \tau, j)$ is given that simultaneously computes a certain value for $\tau$ $(\lambda i.\exists x[woman\ xi \wedge walk\ xi] \rightarrow \neg\exists x[man\ xi \wedge talk\ xi])$. The computation establishes that the proposition (if((a woman)walk))((no man)talk) determines that value in all possible worlds $j$.

(3) a. $d^n(\mathcal{R}, R, j) \rightarrow d^n(\lambda\vec{z}.\mathsf{not}\ \mathcal{R}\vec{z}, \lambda\vec{z}\lambda i.\neg R\vec{z}i, j)$

b. $d^n(\mathcal{R}, R, j) \wedge d^n(\mathcal{R}', R', j) \rightarrow d^n(\lambda\vec{z}.\mathsf{and}(\mathcal{R}\vec{z})(\mathcal{R}'\vec{z}), \lambda\vec{z}\lambda i.R\vec{z}i \wedge R'\vec{z}i, j)$

c. $d^n(\mathcal{R}, R, j) \wedge d^n(\mathcal{R}', R', j) \rightarrow d^n(\lambda\vec{z}.\mathsf{or}(\mathcal{R}\vec{z})(\mathcal{R}'\vec{z}), \lambda\vec{z}\lambda i.R\vec{z}i \vee R'\vec{z}i, j)$

d. $d^n(\mathcal{R}, R, j) \wedge d^n(\mathcal{R}', R', j) \rightarrow d^n(\lambda\vec{z}.\mathsf{if}(\mathcal{R}\vec{z})(\mathcal{R}'\vec{z}), \lambda\vec{z}\lambda i.R\vec{z}i \rightarrow R'\vec{z}i, j)$

e. $d^{n+1}(\mathcal{R}, R, j) \wedge d^{n+1}(\mathcal{R}', R', j) \rightarrow$
$d^n(\lambda\vec{z}.\mathsf{every}(\mathcal{R}'\vec{z})(\mathcal{R}\vec{z}), \lambda\vec{z}\lambda i\forall x[R'\vec{z}xi \rightarrow R\vec{z}xi], j)$

f. $d^{n+1}(\mathcal{R}, R, j) \wedge d^{n+1}(\mathcal{R}', R', j) \rightarrow$
$d^n(\lambda\vec{z}.\mathsf{a}(\mathcal{R}'\vec{z})(\mathcal{R}\vec{z}), \lambda\vec{z}\lambda i.\exists x[R'\vec{z}xi \wedge R\vec{z}xi], j)$

g. $d^{n+1}(\mathcal{R}, R, j) \wedge d^{n+1}(\mathcal{R}', R', j) \rightarrow$
$d^n(\lambda\vec{z}.\mathsf{no}(\mathcal{R}'\vec{z})(\mathcal{R}\vec{z}), \lambda\vec{z}\lambda i.\neg\exists x[R'\vec{z}xi \wedge R\vec{z}xi], j)$

h. $d^{n+1}(\mathcal{R}, R, j) \rightarrow d^n(\lambda\vec{z}.\mathsf{mary}(\mathcal{R}\vec{z}), \lambda\vec{z}\lambda i.\exists x[x = mary \wedge R\vec{z}xi], j)$

i. $d^n(\mathcal{R}, R, j) \rightarrow d^n(\lambda\vec{z}.\mathsf{necessarily}(\mathcal{R}\vec{z}), \lambda\vec{z}\lambda i.\forall k[acc\ ik \rightarrow R\vec{z}k], j)$

j. $d^n(\mathcal{R}, R, j) \rightarrow d^n(\lambda\vec{z}.\mathsf{possibly}\ (\mathcal{R}\vec{z}), \lambda\vec{z}\lambda i.\exists k[acc\ ik \wedge R\vec{z}k], j)$

k. $d^{n+2}(\lambda\vec{u}.\mathsf{is}\ xy, \lambda\vec{u}\lambda i.x = y, j)$, where $\vec{u}$ contains $x$ and $y$

l. $d^{n+2}(\lambda\vec{u}.\mathsf{love}\ xy, \lambda\vec{u}.love\ xy, j)$, where $\vec{u}$ contains $x$ and $y$

m. $d^{n+1}(\lambda\vec{v}.\mathsf{planet}\ x, \lambda\vec{v}.planet\ x, j)$, where $x$ is among the $\vec{v}$

n. $d^{n+1}(\lambda\vec{z}.\mathsf{believe}\ (\mathcal{R}\vec{z}), \lambda\vec{z}.believe\ (\mathcal{R}\vec{z}), j)$

But there is a second interpretation of these meaning postulates in which the variable $j$ does not range over *all* possible worlds but only over a subset of them, the subset of worlds that is consistent with the semantic assumptions of a certain agent, for example, or the subset of worlds that are in accordance with the common semantic knowledge of a certain community. Let $B$, of type $e(s(st))$, be the *doxastic alternative* relation, so that $B\ john\ ij$ (or $B(john, i, j)$ for readability) formalizes that in world $i$ world $j$ is a doxastic alternative for John.[2] [3] The postulates in (3) can be interpreted with the variable $j$ ranging over John's doxastic alternatives. Technically this can be done by adding $B(john, w_0, j)$ (where $w_0$ is a constant denoting the actual world) as an extra conjunct to the antecedent of all postulates in (3), so that, for example, (3a) becomes $B(john, w_0, j) \wedge d^n(\mathcal{R}, R, j) \rightarrow d^n(\lambda\vec{z}.\mathsf{not}\ \mathcal{R}\vec{z}, \lambda\vec{z}\lambda i.\neg R\vec{z}i, j)$. In a computation such as the one in Figure 1 $B(john, w_0, j)$ will now be added to all lines except the first and will act as a constraint on worlds $j$. In fact, the computation in Figure 1 can now be interpreted as *John's* reasoning about the sense-reference relation of a certain sentence, just as the meaning postulates in

[2]This doxastic alternative relation can be used to render John's *implicit* beliefs; postulate (3n) talks about *explicit* belief.

[3]In the following I will make no distinction between belief and knowledge. While all alternative relations under consideration will be constructed out of agents' *doxastic* alternatives, I will, in accordance with common usage, nevertheless speak of "everyone's *knowledge*" and "common *knowledge*".

(3) have been made contingent upon John's implicit beliefs, encoded by John's doxastic alternatives.

Other agents will have their own sets of beliefs and these sets will lead to theories of interpretation that are possibly different from that of John's. If we take any set of possible worlds, there will be a certain set of $d$ relations that hold in every element of that set and a set of possible worlds therefore determines a *theory of interpretation*. If an agent bears the doxastic alternative relation to certain possible worlds, then the theory determined by the set of those worlds may be said to be *that agent's* theory of interpretation. It is also possible to associate a theory of interpretation with *a group of agents*. Let $G$ be such a group ($G$ is supposed to be a constant of type $e(st)$). The relation denoted by $\lambda ij.\exists x(Gxi \wedge Bxij)$ is the alternative relation underlying the modality "everyone in $G$ knows/believes that" (see Fagin et al. [1]). If we take its transitive closure (easily definable within our logic) we arrive at an alternative relation which we will abbreviate as $\mathcal{C}_G$ and which underlies the modality of "common knowledge". The statement $\lambda i.\forall j(\mathcal{C}_G ij \rightarrow \varphi j)$ will be true in all worlds $i$ such that $\varphi$ (of type $st$) holds in all worlds $j$ that are $\mathcal{C}_G$ alternatives to $i$. We abbreviate it as $C_G\varphi$, "it is common knowledge in group $G$ that $\varphi$". For a wealth of information about the common knowledge operator and its logic, see Fagin et al. [1].

When above we sketched how the meaning postulates in (3) could be relativized to an agent's set of doxastic alternatives, we seemed to be heading for a rather solipsistic notion of meaning, with each agent entitled to his own theory of interpretation and no communication being possible between agents. While this picture may strike some as realistic we take the perhaps overly optimistic view that communication sometimes is possible and this is where the notion of common knowledge comes in. Suppose that the postulates in (3) do not only belong to the meaning postulates that you and I accept but are in fact common knowledge between us. Then I can signal to you that $\lambda i.\exists x[\mathit{woman}\ xi \wedge \mathit{walk}\ xi] \rightarrow \neg\exists x[\mathit{man}\ xi \wedge \mathit{talk}\ xi]$ holds in the actual world by getting the proposition $(\mathsf{if}((\mathsf{a\ woman})\mathsf{walk}))((\mathsf{no\ man})\mathsf{talk})$ across. The Fregean assumption of a mysterious realm where propositions reside and where we can grasp them is unnecessary for explaining the possibility of communication. Common knowledge provides a more earthly explanation.[4]

Much of what was said about the sense-reference relation above can also be said about the relation of *synonymy*. A completely fine-grained theory of meaning, such as the ones in Thomason [4], Moschovakis [2], or Muskens [3], will not allow any pair of non-identical expressions to be synonymous. This will evade problems of non-substitutivity but fails to explain in what sense say *woodchuck* and *groundhog* or *ophtalmologist* and *eye doctor* are synonymous. A solution seems to lie in a relativization to the common knowledge of linguistic communities. For each $n$, let $\mathit{syn}^n$ be a relation of type $(e^n p)((e^n p)(st))$ with the intended meaning of expressing synonymity between expressions of type $e^n p$. For example, $\mathit{syn}^1(\mathsf{woodchuck}, \mathsf{groundhog}, j)$ is intended to express that

[4]Of course, the question how common knowledge can come about or how it can be approximated is a non-trivial one (see Fagin et al. [1], Vanderschraaf [5]), but in principle it seems to be amenable to rational investigation.

*woodchuck* and *groundhog* are synonymous in $j$. It is reasonable to stipulate that $\lambda\mathcal{R}\mathcal{R}'.syn^n(\mathcal{R},\mathcal{R},j)$ is an equivalence relation for each $n$ and each world $j$ and, moreover, the following interdependency with the $d^n$ relations should hold.

(4) $syn^n(\mathcal{R},\mathcal{R},j) \wedge d^n(\mathcal{R},R,j) \rightarrow d^n(\mathcal{R},R,j)$

If $syn^1(\mathsf{woodchuck},\mathsf{groundhog},j)$ now holds of all $j$ such that $\mathcal{C}_G ij$ for some group $G$, the members of that group will have common knowledge that *woodchuck* and *groundhog* are synonymous and denote the same animal. The notion of synonymy has thus been relativized to groups as well and now has a social interpretation. In a future longer paper we hope to investigate some of the consequences of this perspective on synonymy with regard to some classic foundational puzzles of semantics.

## 4 Conclusion

We have sketched a theory in which central notions of semantics are relativised to a group interpretation. This brings formal semantics more in line with certain standard linguistic insights than it was before. The Saussurean insight that the form–meaning relation is *arbitrary* dovetails well with the present set-up. That there may be individual *divergencies* from the form–meaning relation accepted by a certain group is also easily explained, as is the possibility for the form–meaning relation of a certain group to shift over time. The model also strongly suggests that it is advantageous for a group to have a stable and large common theory of interpretation and that it may be advantageous to an individual to adopt that common theory.

## References

[1] Fagin, R., Halpern, J., Moses, Y., and Vardi, M.: 1995, *Reasoning about Knowledge*, MIT Press, Cambridge, MA

[2] Moschovakis, Y.: 1994, Sense and Denotation as Algorithm and Value, in *Logic Colloquium '90 (Helsinki 1990)*, Vol. 2 of *Lecture Notes in Logic*, pp 210–249, Springer, Berlin

[3] Muskens, R.: 2005, Sense and the Computation of Reference, *Linguistics and Philosophy* 28(4), 473–504

[4] Thomason, R.: 1980, A Model Theory for Propositional Attitudes, *Linguistics and Philosophy* 4, 47–70

[5] Vanderschraaf, P.: Summer 2002, Common Knowledge, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, `http://plato.stanford.edu/archives/sum2002/entries/common-knowledge/`