Proceedings of the

Amsterdam Graduate Philosophy Conference

—Meaning and Truth—

Amsterdam, October 1–3, 2009

Proceedings of the Amsterdam Graduate Philosophy Conference —Meaning and Truth—

Amsterdam, October 1–3, 2009

Theodora Achourioti, Edgar Andrade,
and Marc Staudacher (eds.)

Department of Philosophy/ILLC
Universiteit van Amsterdam

Preface

The Amsterdam Graduate Philosophy Conference—Meaning and Truth—was organised by the Department of Philosophy and the Institute for Logic, Language and Computation (ILLC) of the Universiteit van Amsterdam. The organisers were Dora Achourioti, Edgar Andrade, and Marc Staudacher.

The conference was dedicated to exploring new ideas on what has been and remains a fundamental theme in the philosophy of language, namely, the relation between meaning and truth. The papers that appear in these proceedings come from researchers who have an original contribution to make regarding the role of truth in a theory of meaning, the role of meaning in a theory of truth, or even the question of whether meaning and truth are actually related in an interesting way.

The conference was motivated by the ongoing debates and discussions that pose new challenges on how to conceive of meaning and of truth, and the relation between them. Some areas of interest here included: truth-functional vs. proof-theoretic semantics; semantic theories of truth; the role of context in interpretation; semantic normativity; deflationism; meaning as use; inferentialism; compositionality; vagueness; the semantics-pragmatics interface; language evolution.

The conference organisers have consulted: Dr. Maria Aloni, Dr. Paul Dekker, Dr. Catarina Dutilh Novaes, Prof. Dr. Jeroen Groenendijk, Prof. Dr. Wolfram Hinzen, Prof. Dr. Michiel van Lambalgen, Dr. Benedikt Löwe, Dr. Robert van Rooij, Prof. Dr. Martin Stokhof, and Prof. Dr. Frank Veltman.

Our Programme Committee included: Kathrin Glüer-Pagin, Jeroen Groenendijk, Stephan Hartmann, Wolfram Hinzen, David Hunter, Andreas Kemmerling, Michiel van Lambalgen, Alex Miller, Thomas Müller, Peter Pagin, Klaus Petrus, François Recanati, Martin Stokhof, Frank Veltman, Albert Visser.

The organizers are most grateful to the following sponsors for their support: Institute for Logic, Language, and Computation—ILLC, Leerstoelgroep Logica en Taalfilosofie, NWO-Project "The Origins of Truth, and the Origins of the Sentence", NWO-Project "Indefinites and beyond: Evolutionary pragmatics and typological semantics", Research Training Site—GLoRiClass, Gemeente Amsterdam, Gesellschaft für Analytische Philosophie—GAP.

The conference organisers appreciate and acknowledge local support on finances and administration from the ILLC buro, in particular from Peter van Ormondt, Ingrid van Loon, Tanja Kassenaar and Marco Vervoort.

Finally, the organizers want to thank the speakers for their contributions.

The Editors,
Amsterdam, March 2010

Contents

Achourioti, T., Andrade, E.J., Staudacher, M. (eds.), Proceedings of the Amsterdam Graduate Philosophy Conference — Meaning and Truth, ILLC Publications X-2010-1, pages 1–8

Varieties of Contextualism

François Recanati (Institut Jean Nicod/Arché)

recanati@ens.fr

I Introduction

What defines contextualism is the generalization of context-sensitivity. Context-sensitivity (or context-dependence) is the property in virtue of which a given expression, with its conventional meaning fixed by the rules of the language, may nevertheless carry different contents in different contexts. (This formulation rules out a trivial form of context-dependence characteristic of sounds or inscriptions as opposed to expressions, viz. language-dependence: the dependence of the content of an inscription upon the language it is contextually taken to belong to. I take 'expressions' to be individuated in terms of the language they belong to.)

In this talk, I will distinguish three varieties of contextualism: methodological contextualism, modulation-based contextualism, and radical contextualism.

II Methodological Contextualism (MC)

Methodological Contextualism is what we get when we reject a certain presumption which was still prevalent twenty years ago and which we may call the 'Literalist presumption'. Let us start from the following (uncontroversial) premiss:

- There is a 'basic set' of expressions whose content is known to depend upon the context in a systematic manner: the indexicals.

The presumption which MC rejects can now be stated as follows:

LITERALIST PRESUMPTION: Expressions *not* in the basic set are (by default) assumed to be context-*in*sensitive.

The Literalist presumption is explicit in the writings of so-called 'semantic minimalists', but it is also implicitly at work in a number of fallacious arguments using Grice's 'Modified Occam's Razor', or an equivalent principle of parsimony, to demonstrate that a semantic analysis in terms of conversational implicature is preferable to an account in terms of truth-conditional content proper (Recanati, 1994, 2004, pp. 155-58). Classic examples involve the use of Modified Occam's Razor against Strawson's (1952) view of the contextually varying truth-conditional contributions of 'and' (in *Introduction to Logical Theory*), or against Donnellan's (1966) view of the contextually varying truth-conditional contributions of definite descriptions. In each case, the possibility that the relevant expression (which seems to carry different contents in different contexts) might be context-sensitive even though it does not belong to the basic set is ignored, in virtue of the Literalist presumption, and the argument proceeds as if the only options available to account for the data were lexical ambiguity, on the one hand, and conversational implicature, on the other hand (with Modified Occam's Razor being used to rule out the former option).

Methodological Contextualism rejects the Literalist presumption. According to MC, we don't know in advance which expressions are context-sensitive and which aren't. For all we know, *every expression might be context-sensitive*. Here the universal quantifier takes scope over the epistemic modal, so what generalizes is the *possibility* of context-sensitivity. For every expression e—including 'and' or definite descriptions—it may be that e is context-sensitive and contributes different contents in different contexts (even though e is not ambiguous). As a result, we need to draw a *general* distinction between linguistic meaning and propositional content, while allowing for special cases in which they will be identical, instead of doing the opposite (i.e., equating conventional meaning and propositional content, while allowing for exceptions—the expressions in the 'basic set').

III Substantial contextualism (in general)

It also considers the possibility that every expression might be context-sensitive, but the universal quantifier now takes narrow scope. The possibility that is being considered is the possibility that: for every expression e, e *is* context-sensitive. Here what tentatively generalizes is (actual) context-sensitivity, not the possibility of context-sensitivity.

IV Two types of (substantial) contextualism

In the case of indexicals the propositional contribution made by an expression depends upon the context and is not fully determined by the (context-independent) meaning which the expression (type) possesses in virtue of the semantic rules of the language. Contextualism generalizes that feature to 'ordinary' expressions. It holds that, in general, the propositional contribution of an expression is not fully determined by the invariant meaning conventionally associated with the expression type.

Two versions, more or less radical:

(MBC): The linguistic meaning of an (ordinary) expression *need not be* what the expression contributes to propositional content. [Example: 'There is a lion in the middle of the courtyard'—the meaning of 'lion' is pragmatically modulated]

(RC): It *cannot be* what the expression contributes to propositional content. ['wrong format' view]

MBC and RC can be characterized in terms of a certain assumption I call the 'Fregean assumption'. Both MBC and RC reject it, though for different reasons.

V The Fregean assumption

Let us assume that utterances express 'propositions' or 'thoughts', and that these propositions/thoughts are made out of, or can be analysed into, certain building blocks or constituents, to be called *senses*. The standard assumption in the literature stemming from Frege is that:

(FA): what an expression contributes, when it is used (together with other expressions) in making a complete utterance, is the sense which it independently possesses in virtue of the conventions of the language.

This presupposes that *the conventions of the language associate expressions with senses*. Call this the 'Fregean presupposition' (FP). Is FP tenable? RC, the more radical of the two forms of contextualism, rejects it. MBC accepts it but still rejects the Fregean assumption.

VI Is FP tenable?

Consider the most frequently used expressions, those which exhibit a high degree of polysemy. What do they mean? A number of scholars believe their meaning is schematic and has to be fleshed out on any particular use. This suggests that, perhaps, their conventional meaning is not a fully-fledged sense. Can we argue that they are *ambiguous* between a number of distinct senses? That is not obvious because it does not seem that there is a *discrete* list of such senses available but, rather, a continuum of possible senses to which one can creatively add in an open-ended manner. That is not to say that the meaning of such an expression reduces to an abstract schema: the expression is undoubtedly *also* associated in memory with conventional ways of using it in collocations with (more or less) determinate senses. All this—the abstract schema or schemata, the collocations, the senses—arguably goes into the linguistic meaning of the expression, which starts looking rather messy. On such a view, the meaning of an expression does not have the right 'format' to be what the expression contributes to propositional content. In other words, linguistic meanings are not senses (though they may involve, inter alia, senses): the Fregean presupposition must therefore be rejected. That is also the conclusion one can draw from Putnam's (1975) considerations regarding another class of expressions, namely natural kind terms and, more generally, lexical items with a referential dimension (those which cannot be defined but serve as labels for aspects of the environment). According to Putnam, the context-independent meaning of such terms is a 'vector' consisting of a bunch of elements, including a 'semantic marker' and a 'stereotype' which

itself consists of various pieces of encyclopedic information. This does not look like a sense at all. So, in these two types of (arguably central) cases, the conventional meaning of an expression (type) is either too rich or too poor to be what the expression (token) contributes to content.

The views I have just mentioned may well be wrong, of course. When it comes to lexical semantics, nearly everything is up for grabs. That, however, is precisely my point. As theorists, we have an idea what senses are, i.e., what words contribute when we speak. We know, more or less, how to model that. But we know very little about what words themselves mean and what relation there is between word meaning and contributed sense. In view of the limits of our knowledge, it is reasonable to give up the simplifying assumption FP we inherited from Frege, in order at least to start making serious enquiries in that area.

VII Does RC threaten systematic semantics?

As we have just seen, the Fregean assumption presupposes that linguistic meanings are senses, i.e., propositional constituents. If we give up that presupposition, we are left with the idea that lexical meanings plays some role in determining the sense which is an expression's contribution to the thought expressed. This idea can be expressed as follows:

(CA): Sense contributed by expression e=f(lexical meaning of e, x)

where 'x' is whatever, in addition to lexical meaning, is needed to determine sense. If, as seems very likely, 'x' includes the *context* in which the expression is tokened (and in particular the most important among contextual factors: what one is talking about), then we get a form of contextualism that 'generalizes indexicality'. That is RC.

How detrimental is acceptance of RC to the project of systematic semantics? Very little, I think. Indexicality does not prevent us from systematically mapping semantic values for the parts to semantic values for the wholes in which they occur. The only qualification induced by indexicality is that the semantic values for the parts may not be given in a context-independent manner; they may not

be determined solely by the conventions of the language. Thus, in a truth-theoretic framework, the theorems of one's semantic theory have to be *conditional* T-sentences whenever the sentence at issue contains indexicals (Burge, 1974; Higginbotham, 1988). Generalizing indexicality, as RC does, simply means that the theorems will *always* be conditional T-sentences.

VIII Modulation-based Contextualism (MBC)

In contrast to RC, MBC accepts that the conventions of the language associate expressions with senses. MBC nevertheless rejects the Fregean assumption that the senses which are the meanings of expressions are also what these expressions contribute when they are used (together with other expressions) in making a complete utterance. MBC holds that an expression may, but need not, contribute its sense—i.e., the sense it independently possesses in virtue of the conventions of the language; it may also contribute an indefinite number of *other* senses resulting from modulation operations (e.g., free enrichment, predicate transfer, sense-extension, etc.) applied to the proprietary sense. This is a form of contextualism, because modulation itself is context-sensitive: whether or not modulation comes into play, and if it does, which modulation operation takes place, is a matter of context. It follows that what an expression actually contributes to the thought expressed by the utterance in which it occurs is always a matter of context.

IX Objection to MBC

In contrast to the contextual assignment of values to indexicals, modulation is not driven by the linguistic meaning of words. Nothing in the linguistic meaning of the words whose sense is modulated tells us that modulation ought to take place. Modulation takes place purely as a matter of context, of 'pragmatics'; what drives it is the urge to make sense of what the speaker is saying. So MBC raises the following objection: as soon as one accepts 'free' pragmatic processes, interpretation (content recovery) is no longer driven by the grammar: it is pragmatic through and through and does not significantly differ from "the kind [of interpretation] involved in interpreting kicks under the table and taps on the shoulder" (Stanley, 2000, p. 396).

X Response to the objection

Semantic interpretation remains grammar-driven even if, in the course of semantic interpretation, pragmatics is appealed to not only to assign contextual values to indexical and free variables but also to freely modulate the senses of the constituents in a top-down manner.

Let us distinguish between the *literal* sense of a simple expression, namely its semantic interpretation, and its *modulated* sense. The modulated sense of an expression (in context) results from applying the contextually appropriate modulation function to its semantic interpretation. So far, so good. Now MBC differs from standard frameworks in letting the semantic interpretation of a *complex* expression be a function of the *modulated senses* of its parts and the way they are put together. The semantic process of composition and the pragmatic process of sense modulation are therefore irreducibly intertwined. For simple expressions, their semantic interpretation is their literal sense, but for complex expressions pragmatic modulation is allowed to enter into the determination of semantic content. In this framework, even though free pragmatic processes are allowed to enter into the determination of truth-conditional content, they come into the picture as part of the compositional machinery. On this view semantic interpretation is *still* a matter of determining the sense of the whole as a function of the (possiby modulated) senses of the parts and the way they are put together.

References

Burge, T. (1974). Demonstrative constructions, reference, and truth. *Journal of Philosophy*, 71:205–23.

Donnellan, K. (1966). Reference and definite descriptions. *Philosophical Review*, 75:281–304.

Higginbotham, J. (1988). Contexts, models, and meanings. In Kempson, R., editor, *Mental Representations: The Interface between Language and Reality*, pages 29–48. Cambridge University Press, Cambridge.

Putnam, H. (1975). *Language, Mind, and Reality: Philosophical*

Papers, volume 2, chapter The Meaning of Meaning, pages 215–271. Cambridge University Press, Cambridge.

Recanati, F. (1994). Contextualism and anti-contextualism in the philosophy of language. In Tsohatzidis, S., editor, *Foundations of Speech Act Theory*, pages 156–166. Routledge, London.

Recanati, F. (2004). *Literal Meaning*. Cambridge University Press, Cambridge.

Stanley, J. (2000). Context and logical form. *Linguistics and Philosophy*, 23:391–434.

Strawson, P. (1952). *Introduction to Logical Theory*. Methuen, London.

Achourioti, T., Andrade, E.J., Staudacher, M. (eds.), Proceedings of
the Amsterdam Graduate Philosophy Conference — Meaning and Truth,
ILLC Publications X-2010-1, pages 9–38

Abstractions and Idealisations: The Construction of Modern Linguistics

Martin Stokhof and Michiel van Lambalgen
(ILLC/Universiteit van Amsterdam)[*]

m.j.b.stokhof@uva.nl and M.vanLambalgen@uva.nl

I Introduction

In many ways, modern linguistics is one of the most remarkable and successful scientific innovations of the twentieth century. The rise of generative grammar in the fifties and sixties produced an atmosphere of intellectual excitement that seemed to be reserved for fundamental developments in the natural sciences. And the excitement was not restricted to linguistics as such, it stretched out to other disciplines, such as philosophy, the emerging disciplines of computer science and cognitive psychology, anthropology and literary studies. And to the present day modern linguistics is held up as a model of scientific innovation to other disciplines in the humanities.

A satisfactory account of this remarkable development will have to factor in a number of things. The role of the natural sciences and the formal sciences as a 'standard model' of scientific inquiry is one of them. Another is the way in which modern linguistics appears to tie in with internal, disciplinary developments in other fields. Sociological factors, such as the way in which the discipline organises

[*]A Dutch version of this paper is under submission with *Tijdschrift voor Filosofie*.

itself, are also relevant.[1] And then there is the way in which linguistics appears to have succeeded to conceptualise its central objects of study so as to fit a particular methodology.

In this paper we deal with this last issue, i.e., with the question how modern linguistics has constructed its objects of study, such as 'language', 'grammar', 'competence', 'meaning', 'rule'. Apparently, a major factor that explains the success and prestige of modern linguistics is that it has succeeded to come up with scientific characterisations of its core concepts that have allowed linguists to develop theories that are both descriptively and explanatory adequate. In what follows we focus on a particular aspect of this complicated process that, we feel, has not received adequate attention in the literature to date, viz., the nature of the kind of constructions that modern linguistics employs.

There are two things we would like to mention at the outset. First of all, in what follows we use the phrase 'modern linguistics' mainly as an indication of what is still a dominant approach, viz., the generative tradition. And secondly, our considerations primarily have a 'meta'-character, i.e., the observations that follow are not intended as arguments pro or con particular positions, although they could have such repercussions. But the spelling out of such consequences is beyond the scope of this paper.

II The state of the art

As we noted above, the rise of modern linguistics, its success and influence, and its enormous intellectual prestige, as such are intriguing phenomena, that call for an explanation. But also from an internal perspective, i.e., from the perspective of linguistics itself, its present state is one that raises a number of questions.

One of these is, that despite the solid reputation that linguistics has as a successful discipline, many of the expectations have not (or not yet?) been realised. If we look at the description of individual languages, we can note that complete and explicit grammars are still

[1] A thorough, empirical sociological study of the development of modern linguistics does not exist, as far as we know. For studies that are more of the nature of a 'history of ideas', cf., e.g., Newmeyer (1986); Harris (1993).

far off. In the area of typology many studies have been done, but it remains to be seen how much of that work actually depends on the methodology of modern linguistics. Little or no explanations of properties of natural languages exist that are accepted generally, i.e., across theoretical boundaries. When it comes to applications, especially computational ones, we can observe that the theoretical models of modern linguistics, based as they are on the concept of a grammar as a rule system, in general are less successful than stochastic approaches. And with regard to psycholinguistic investigations and research into the neurophysiological processes that underly language and language use, it appears that modern linguistics in general is unable to come up with leading questions and hypotheses.[2]

Another observation regarding the present state of modern linguistics, and one that definitely calls for further study, is the substantial diversity in approaches and models, and even in definitions of central concepts, that has become a distinctive feature of linguistics to date. With the rise of generative grammar, as proposed and developed by Chomsky and others, modern linguistics seemed to be heading towards a remarkable uniformity vis à vis its goals, methodology, and central concepts. At least this appeared to hold for core disciplines such as syntax, morphology, and phonology. In semantics a similar development occurred at the end of the sixties when formal semantics appeared on the scene. 'Montague grammar' apparently developed into a generally accepted model for semantic description and explanation. But the uniformity and consensus that at some point seemed almost natural have disappeared: there is an enormous variety of approaches, theoretical models, methodologies, and even with regard to the goals of linguistics and its very object of investigation there are fundamental differences of opinion.[3]

These observations give rise to a fundamental question with regard to linguistics as such: Could modern linguistics perhaps be an

[2]To be sure, this is not just a problem for modern linguistics. Quite generally, it is difficult to derive from theories concerning macroscopic phenomena predictions regarding the underlying neurophysiology due to the absence of clear bridging principles that link the often disjoint conceptual systems.

[3]Cf., Kamp and Stokhof (2008) for a description of this development, and an attempt to explain what drove it, for the case of formal semantics.

example of a 'failed discipline'? As was already noticed above, the adoption of the models and methodologies of the natural sciences and the formal sciences was one of the keys to the success of modern linguistics. Moreover, especially in Chomsky's views a clearly naturalistic goal can be discerned: according to him linguistics studies what in the end is an aspect of human biology. Is this naturalism perhaps one of the causes of the present, confusing situation? Is it that modern linguistics, knowingly or unknowingly, follows a naturalistic approach to phenomena, —language and linguistic competence—, that are of a fundamentally different nature?

This last question is too complex to be even properly articulated in the context of this paper, let alone that it can be answered here. However, we do feel that the observations about abstraction and idealisation as constructive processes that are the subject of what follows do present reasons to think that the question just formulated touches on a central problem with regard to the status of modern linguistics as a scientific discipline. And if we are correct in thinking so, then it is also the case that, precisely because modern linguistics has functioned as a model for other disciplines in the humanities for more than four decades, the relevance of this question extends beyond linguistics as such.

III Examples of constructions

To give the reader some idea of the kind of constructions[4] we have in mind, here are a few examples.

At first sight, 'language' appears to be the most central concept of linguistics. Be it specific natural languages, such as English or Quechua or Rennellese, or natural (human) language in general, language seems to be the core phenomenon that linguists want to describe and explain. Now, from an observational point of view language is first and foremost language use: spoken or written utterances.[5] For the child that acquires its mother tongue, language use

[4]What follows will make clear that the term 'construction' is used here not in its linguistic sense, but as a term that belongs to the vocabulary of philosophy of science.

[5]Obviously, spoken language is primary vis à vis written language, not just historically but also ontogenetically. Yet in linguistics, as is the case in most

is what it encounters in its environment, for adult language users language is what they use to communicate with each other.

In modern linguistics[6] the intuitive concept of language, viz., that what is encountered in everyday use, has been replaced by the logical, mathematical (algebraic) concept of a language, viz., that of a potentially infinite set of well-formed expressions generated by a finite, or finitely characterisable, set of rules (i.e., a grammar).[7] Not only does this concept emphasise the formal aspect of language, and hence the focus on written language, it also introduces a notion of 'structure' that can be tested against actual linguistic material only indirectly, and partially.[8] Another immediate consequence of the shift towards a formal construction of the concept of language is that expressions are being studied at the level of types, not tokens, with regard to both their form as well as their meaning. Obviously, the historically contingent availability of writing is instrumental in this change.

A related move is that linguistic competence, i.e., the ability of humans to use language, actively in production and passively in interpretation and understanding, is being studied in terms of a comparable construction. Here the well-known distinction between 'competence' and 'performance' plays a key role. Knowledge of a language is conceived as the availability of a grammar, and competence as the ability to use that grammar to distinguish well-formed expressions from non well-formed ones, to assign the former an interpretation,

philosophical treatises on language, the focus is mainly on written language, not on speech. Cf., Kraak (2008) for a recent study of the effects of this shift.

[6]What follows applies not just to the generative tradition, but also the many approaches it has helped shape in this respect. But there are other approaches in which the construction described here does not play a role, or at least not in the same way.

[7]Cf., Tomalin (2006) for an extensive study of the role that the developments in logic in the first half of the twentieth century have played in Chomsky's early work.

[8]In the light of this, one particular development in modern linguistics becomes more easy to understand, viz., the fact that one of the most central notions, that of 'syntactic structure', has been subjected to many, and radical, changes. This constant re-conceptualisation and re-modelling of a core notion makes sense only if we keep in mind its mainly theoretical nature (and that of related notions, such as 'rule', 'constituent', and so on). Cf., Stokhof (2002) for a discussion of similar observations with regard to the central notion of semantics, viz., 'meaning'.

and then to use them both actively and passively. This linguistic competence, though an individual capacity in the sense of being ascribable to an individual as such, is not introspectively accessible to the individual that has it.[9]

Another phenomenon, that is closely related to the idea of competence as an individual ability and that has strongly influenced contemporary thought about language, and hence also the goals of modern linguistics, is the so-called 'problem of creativity' (or 'compositionality'). It is the 'observation'[10] that a language consists of a potentially infinite number of wellformed expressions that somehow has to be represented in a finite manner in the finite individual human brain. In a certain sense this 'problem' is generated directly by the shift towards the logical, mathematical characterisation of the core concept of a language. Closely related is what Kraak in his aforementioned book calls 'the myth of representation', viz., the idea that language, and in particular written language, serves as a medium of representation of internal, mental contents. If we assume that humans are capable of a potentially infinite number of thoughts (and desires, and conjectures, and questions, and so on), then the myth of representation inevitably leads to the conclusion that the language we use to express such contents also has to have a unlimited character.

These constructions, and others like them, lead to a relative neglect of both the actual use of language as well as the context in which that actual use appears: the physical, social and cultural environment, both synchronically as well as diachronically. Whenever attention is being paid to language use, it is always as complementary to the idea of language as characterised by the form and (lit-

[9]This creates what Jackendoff (Jackendoff, 1987, p.20) calls 'the mind – mind problem'. On the one hand, we can be clearly and consciously aware of what we do with language (we may consciously opt for a certain interpretation of an utterance, or for a certain formulation of what we want to say, we may be at a loss as to the meaning of what is being said, or object to a certain choice of words for a number of reasons), but, on the other hand, the mechanisms that are postulated to constitute the essence of our competence are in principle shielded from direct inspection.

[10]In scare quotes because in fact of any natural language only a finite number of utterances will ever be observed. Cf., Groenendijk and Stokhof (2005) for further discussion.

eral) meaning of its expressions. Almost all theories about what it is that people do with language start from these very assumptions about what language and linguistic competence are. The result is very much an abstract and individualistic picture: linguistic competence is an individual ability, and language use is a process in which autonomous and competent individuals exercise their linguistic competence. That language use has a social nature, in which communication plays a central role, is, of course, not something that many linguists would like to deny. But, so the leading idea proclaims, the language that is being used and the competence that is being applied in that social process, can be described, characterised and explained as such, and quite independently from language use.[11] Behind this is the fundamental assumption that in the end language and linguistic competence can be understood as phenomena that are anchored in human biology, and that it is only via the methodology of the natural sciences that we may acquire insight into their nature and function.

This, admittedly concise, sketch of some core moves in the construction of the central concepts and goals of linguistics gives reason to believe that modern linguistics has been decisively influenced by ideas and developments in other disciplines, notably the formal and the natural sciences, but also philosophy. As for the influence of the latter, Chomsky's rationalism is an obvious and explicit example, but at other points it is more subtle and therefore perhaps less often noticed.[12] In what follows we will not so much be concerned with the actual details of such constructions, but rather focus on the nature of the process as such. In doing so, our central question is the following: Are these constructions like the abstractions we are familiar with from the natural sciences, or are they of a different nature? And if the latter turns out to be the case, what are the consequences for the status of linguistics?

[11] This is very much the dominant view, one that can be found explicitly in the work of Chomsky, and one that has gone unchallenged for a long time. Recently other views have started to emerge, in the concluding section we will briefly mention some of them.

[12] Cf., Stokhof (2002, 2007, 2008) for an analysis of various philosophical distinctions and goals that have shaped and continue to guide formal semantics.

IV Abstractions as constructions

Abstraction is a well-known tool for turning a natural phenomenon into a 'suitable' object of scientific investigation. Standard examples are the frictionless plane in classical mechanics, the perfect vacuum, pure chemical substances, and so on. Whereas in reality moving objects always are subject to friction, a perfect vacuum does not exist and can not be created, and chemical substances almost always contain contaminations from other substances, these facts, when considered from the point of view of studying certain central natural phenomena, are complications which are either deemed irrelevant or too complex or intractable to be captured in a theory, at least for the time being. The latter phenomena in particular are interesting if we want to determine what exactly is that an abstraction is, and does.

The physical theory of tides provides another illustrative example. Newton's theory of 1687 gave an explanation of the frequency and amplitude of tidal waves based on his theory of gravitation, in terms of the combined gravitational pull on the earth exercised by the sun and the moon. His calculations assumed that the entire surface of the earth is covered by one ocean and that this ocean has no inertia of its own. These two assumptions meant that, first of all, local circumstances on the earth could not play a role, and, second, that the earth's rotation was not taken into account. Also, the effect of other celestial bodies, such as the planet Venus, was disregarded.

Of course the reality of the phenomena that did not fit into this model was not denied. In fact, further work on the theory produced a model in which these phenomena can be accounted for, using both physical calculations as well as observations of the local circumstances at locations where the actual tidal heights needed to be calculated. (Relevant factors include the depth of the ocean, the form of coast lines, the presence of pack ice, and so on.) The more accurate model is analogous to that of a vibrating violin string: the timbre of the sound it produces is determined by the many frequencies, each with its own amplitude, that co-occur with the basic tone. Analogously, the periodic process of tidal waves is determined by many frequencies, some of which are determined by astronomic

laws, others by local circumstances.[13] But even in this more complex model one is forced to abstract, since some frequencies, such as the disturbances caused by moving sand banks, are too difficult to predict. However, the reality of the factors from which one abstracts, is never denied, and in principle the model is capable of incorporating them.

This is a crucial feature of the way in which abstraction in the natural sciences works: the phenomenon from which we abstract is a real one, and its reality is acknowledged in the theory or in the model that is based on the abstraction. After all, in factual observations and experiments these phenomena inevitably occur. One of the main reasons for nevertheless abstracting from them is that by doing so one is able to come up with a better explanation of the underlying causal mechanisms while keeping the predictions of the theory based on the abstraction within certain acceptable limits of accuracy.

This means that there is a real and acknowledged interaction between the theory, i.e., the explanation it provides of a certain phenomenon together with the predictions it delivers, and reality as it occurs in observations and experiments. Another example of this is provided by the concept of a perfect vacuum. In physics so-called 'free space constants', such as the speed of light and the magnetic constant, play a key role. The quantitative values of these constants is theoretically determined with reference to a perfect vacuum. In reality, in which a perfect vacuum does not occur, these constants always have slightly different values, but the differences can be approximated with sufficient precision to make the predictions of the theory practically useful. (And in many cases the differences are so small that they can be safely ignored.) So what we see is that theory based on abstraction and observation and experiment without abstraction remain intimately connected, both conceptually as well as practically.

And the reason that this is a crucial feature of the way in which abstraction in the natural sciences works is that it explains why

[13] In fact, in the case of tidal waves the model is somewhat more complex, since, unlike in the case of the vibrating string, there is more than one basic tone at work at the same time.

theories that make use of abstractions still work: they do not 're-conceptualise' the phenomena.

V Abstraction in linguistics?

In modern linguistics, too, we often find appeals to abstraction when it comes to explaining how a linguistic theory is related to observable reality. The following quote from Chomsky (1980, p. 219) illustrates what is at stake:

> Any serious study will [...] abstract away from variation tentatively regarded as insignificant and from external interference dismissed as irrelevant at a given stage of inquiry. [...] It should come as no surprise, then, that a significant notion of 'language' as an object of rational inquiry can be developed only on the basis of rather far-reaching abstraction.

What Chomsky is suggesting here is that abstraction in linguistics is the same process as in the natural sciences. It allows us, he claims, to concentrate on the core of the phenomenon, disregarding those aspects that are deemed 'insignificant' or 'irrelevant'. As such this is a remarkable statement, because as we have seen above, in the natural sciences abstraction usually does not concern irrelevant or unimportant aspects of phenomena, but features that for one reason or another can not (yet) be incorporated into the theory because they are too complex or intractable. Note also that in this passage Chomsky does not provide any argument why for example the phenomenon of language can be studied only via abstraction. What is it that he means by a 'serious study' or a 'rational inquiry' that it can only be done on the basis of far-reaching abstractions?[14]

For Chomsky, then, it is apparently obvious that the fact that language and linguistic competence, certainly at first sight, are different kinds of phenomena than movement of physical bodies or chemical reactions, constitutes no reason to think that abstraction could not, and should not, play the same role as it does in the natural sciences. Thus he writes in Chomsky (1995, p. 7):

> ... it is a rare philosopher who would scoff at its [i.e., physics'] weird and counterintuitive principles as contrary to right thinking

[14]For an incisive criticism of Chomsky's often heavily rhetorical writing, cf., Paul Postal's essay 'Junk Ethics' in Postal (2004, Part 2).

> and therefore untenable. But this standpoint is commonly regarded as inapplicable to cognitive science, linguistics in particular. Somewhere between, there is a boundary. Within that boundary, science is self-justifying; the critical analyst seeks to learn about the criteria for rationality and justification of scientific success. Beyond that boundary, everything changes; the critic applies independent criteria to sit in judgment over the theories advanced and the entities they postulate.

But this really rest on a misrepresentation of how things are done in the natural sciences. No physicist, for example, would be of the opinion that any aspect of a physical theory is 'self-justifying', including the abstractions on which the theory is based. The final judgement always resides with observational and experimental verification and explanatory adequacy. In other words, the last word is spoken, not by the physicist (and, of course, also not by the philosopher), but by reality itself.[15]

Apart from this misrepresentation, what is intriguing about this passage is that Chomsky apparently thinks that criticism of the constructions that define modern linguistics is not justified because the mechanism employed there does not differ from that in the natural sciences. To put it differently, Chomsky does not differentiate criticism of the process from criticism of the result. But the question is whether that is justified in this particular case. In order to see whether it is, we take a bit more systematic look at the essential features of abstraction in the next section.

VI Features of abstraction

There is some discussion in the literature about the role of abstraction in the natural science,[16] but that by and large concentrates on the modelling of this mechanism (in terms of formal models of

[15] No doubt there are concrete instances in the development of the natural sciences where one might observe a difference between ideology and practice, e.g., when empirical observations are neglected in favour of a theoretically motivated judgement. But that is not what is at stake here. What counts is that in the end one is willing to let the facts, such as they are to the best of one's knowledge, to have the final say. And that principle stands also in the case of theories that are founded on abstractions.

[16] Cf., e.g., Jones (2005).

theories, theoretical vocabularies, and so on). Though interesting and important these are not the aspects we are concerned with here. Our primary interest concerns those features of abstraction that may settle the question whether abstraction plays, or should play, a role in linguistics.

From the examples we have briefly discussed in section III the following features of abstraction emerge:

- Object: a quantitative parameter of a phenomenon that is subject to abstraction, is assigned a specific value (zero, infinite, ...)
- Result: a model of a phenomenon in which the parameter that is being abstracted over is still present
- Motivation: primarily methodological and practical

The quantitative nature of the object of abstraction does not come as a surprise: most theories in the natural sciences aim for a description and explanation of phenomena in terms of interactions and causal connections between quantitative features (speed, mass, spin, magnetic force, and so on). Relevant candidates for abstraction then are those quantitative features of which the exact actual values are irrelevant or too complex to determine. Examples of the former are the exact values of the afore-mentioned physical constants, keeping in mind that the question of '(ir)relevance' ultimately depends on the application of the theory. Examples of the latter we may find for example in the theory of tidal waves, in fluid dynamics and in the study of other semi-chaotic physical systems.

As for the result of abstraction, what is crucial there is that abstraction is not the same as negation. What is being neglected is the actual value of a parameter in a concrete situation, but *not the parameter itself*. For example, if we employ the concept of a perfect vacuum we assume that there are no particles with mass, but not that mass is not a relevant concept.[17] In this sense abstraction is

[17] As another example, cf., how negation functions in the law of inertia: 'If the vector sum of all forces (that is, the net force) acting on an object is zero, then the acceleration of the object is zero and its velocity is constant'.

conservative: in the resulting model the features that we abstract over, are still present. In other words, abstraction does not change the ontology of the phenomena, and that make it possible, at least in principle if not always in practice, to 'undo' an abstraction. This is also evident from the fact that the predictions we derive from a theory based on an abstraction can actually be compared with observations and the outcomes of experiments.

And in the end, that is what we actually want, since it is only through observation and experiment of the phenomena as they actually present themselves that we can evaluate our theories and gauge their explanatory power. In other words, abstraction first and foremost is a means to an end, it is there to enable us to start theorising by lifting some of the epistemological burden. In sum: abstraction is methodologically and practically motivated, not ontologically or ideologically.

VII Features of idealisation

As we will illustrate in this section, the type of construction that is used in linguistics and that is often taken for abstraction as it is used in the natural sciences, differs from the latter on a number of fundamental points. In particular, in linguistics the objects lack the quantitative nature that is so characteristic for objects of abstraction in the natural sciences. What we are dealing with in linguistics are rather qualitative features of phenomena that are being ignored. In order to terminologically distinguish the two types of construction we will reserve the term 'abstraction' for the process that we know from the natural sciences, and use the term 'idealisation' to refer to the kind of construction that occurs in linguistics.[18]
Distinctive features of what we call idealisations are the following:

- Object: a qualitative feature of a phenomenon that is being ignored
- Result: a model of a phenomenon in which the feature that is being idealised is missing
- Motivation: primarily ideological and theoretical

[18]Do note that both terms, 'abstraction' and 'idealisation', are used in the literature also in other ways. Cf., the afore-mentioned Jones (2005).

One of the reasons that idealisation differs from abstraction is that whereas the objects of study in the natural sciences are defined (mainly) quantitatively, those in the humanities are (primarily) characterised in qualitative terms. A definition of, say, 'epic poetry', or 'the western christian tradition', but also of such objects as 'meaning' or 'subject', determines an object (almost) completely in terms of qualitative properties. Consequently, a scientific study of such objects focusses on those properties and their relationships with other, similarly qualitative features. Quantitative features (such as determinations of time, location, and so on) may play a role also, of course, but usually they are not really essential, neither for the definition of the object of study as such, nor for the explanations that one is after. What is important to note is that leaving one or more of such qualitative features out of consideration, is not abstraction in the sense in which we discussed that in the previous section. It does not concern a quantitative parameter the value of which is fixed, but a qualitative feature that is left out.

One consequence of this fundamental difference is that the result of an idealisation is likewise fundamentally different from that of an abstraction: in the resulting model the phenomenon in question has turned into something essentially different from the original one. In other words, in the case of idealisation we are dealing with an ontological change, rather than with an epistemological one, as is the case with abstraction. Obviously, this has repercussions for the relation between the idealisation and the original phenomenon: that relation is not longer 'symmetrical'. A simple example may serve to illustrate the point. If in a study of the western christian tradition one limits the object of study to the church, and leaves out aspects that are related to lay people, lay communities and the like, then one actually studies a different (in this case, more restricted) phenomenon, and one can not expect that explanations and connections that are uncovered in the limited model extend to the broader phenomenon. In fact, the limited model will simply not make any predictions 'beyond its scope' whatsoever.

The motivation for a particular idealisation may very well be practical in nature (as in the simple example just given), and as long as one remains aware of the implied restrictions it may be an

unobjectionable move. However, quite often the motivation is not so much practical as ideological. Then certain features of a phenomenon are left out because one wants to apply a specific methodology to the idealised result. That is a move that is based on ideological reasons having to do with the conviction that only certain methods lead to scientifically reputable results. As we will illustrate below, idealisations in linguistics are often motivation by such ideological concerns.

It is worth noting that methodological considerations may play two, essentially different roles. In some cases the choice for a particular methodology is justified by an assessment that the use of a particular method increases the chance of a successful investigation, where what count as 'successful' is determined *independently* of the methodology as such. But one may also choose a particular methodology on ideological grounds, in which case what counts as 'success' *is changed by the methodological choice* (partly because it changes the nature of the object of study). Abstraction is a methodological choice of the former kind, idealisation, in so far as it is (also) motivated by ideological concerns, one of the latter.

VIII Idealisation in linguistics: an example

One of the most prominent and well-known examples of construction in modern linguistics is the 'competence – performance' distinction. in his ground-breaking book *Aspects of the Theory of Syntax* (Chomsky, 1965, p. 3) Chomsky introduces the distinction in the following way:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest and errors (random or characteristic) in applying his knowledge of the language in actual performance.

What happens here is that competence, regarded as the proper object of study of linguistics, is constructed from what we can observe, i.e., everyday use of language, by stripping it from a number of features, such as memory limitations, mistakes, (communicative) goals, attention shifts, and so on. In other words, Chomsky con-

structs from observable language use a concept of linguistic competence by simply ignoring a number of its actual, real properties. In that way a new object of study is created, i.e., an object that has an ontological status that differs from that of the original one.

The reasons for this construction are not given in the passage quoted, it is just being asserted that the features that are left out by the idealisation are 'grammatically irrelevant'. In other words, it is claimed that in the study of language, grammar, and linguistic competence, no attention needs to be paid to such factors as memory, attention, goals, and the like. But note that this claim does not rest on a comparison of (the study of) two independently given objects, viz., idealised competence and actual language use. Rather, one of the two, competence, is being constructed on the basis of this claim, and hence whatever results studying it provides cannot give independent evidence that justifies the construction in the first place. This is a strong indication that we are dealing with an ideologically motivated claim.

That this idealisation actually creates a new object is also evident from the fact that the relation between the original phenomenon, of observable language use, and the new idealised object, competence, creates *new* issues:

> To study actual linguistic performance, we must consider the interaction of a variety of factors, of which the underlying competence of the speaker-hearer is only one.

This passage, also from *Aspects* (p. 4), illustrates that an idealisation raises additional epistemological questions, viz., how the idealised object and the original observable phenomenon can be related to each other. This is quite different in the case of an abstraction, where the relation between the abstraction and the phenomenon really boils down to a specification of actual values of quantitative parameters, a procedure that, though sometimes hard to carry out in practice, does not introduce any new epistemological problems.

This complication is a real one. For example, if we construct a competent language user by idealisation as an individual with implicit knowledge of the grammar of his/her I-language,[19] we leave out

[19]Cf., Chomsky (1986) for the introduction of the concept of 'I-language'. For a thorough criticism from a broadly Wittgensteinian perspective, cf., Stein (1997,

many of the features that are characteristic of actual language users: the already mentioned memory limitations, the fact that language is used in order to reach certain goals (most of the time non-linguistic ones), the social environment in which language is used, but also for example the fact that language users are embodied subjects. Such factors, precisely because they are 'idealised away', are no longer present in the model of the competent language user (a competent user is 'disembodied'), and *the resulting model by itself does not contain any suggestion or clue as to how it could be related in the end to what we can in fact observe.* In that sense, idealisations don't make life any easier, on the contrary, they create a lot of extra work.[20]

IX Abstraction versus idealisation: characteristics and backgrounds

Abstraction and idealisation, then, are two radically different ways in which objects of scientific investigation can be constructed. In the table 1 we summarise their various characteristics:

Abstraction	Idealisation
Methodological	Ontological
Symmetric	Asymmetric
No ontological consequences	Additional epistemological task
Quantitative	Qualitative

Table 1: Abstraction versus Idealisation

An obvious question is why abstraction works in the natural sciences, but not in linguistics. It appears that this is no coincidence but something that is intimately related with the nature of the respective enterprises and with the nature of their respective domains of inquiry.

chapter 3).

[20] Another illustration of this effect, connected with the construction of an ideal competent user, has to do with the 'knowledge' such a user is supposed to have of language. The postulated mental state lacks several characteristic features of what knowledge is, and hence requires the introduction of yet another idealised concept: the competent user 'cognises' language.

By way of illustration table 2 lists some differences between research in the natural science and research in linguistics that are pertinent to this issue:

'Natural science'	Linguistics
Experimental design	Hardly any experiments
Natural ontology	Hybrid ontology
Quantitative differences between theory and application	Qualitative differences between theory and application
Deterministic explanation, causal laws	Interpretative explanation, no strict laws

Table 2: 'Natural science' versus Linguistics

That the natural sciences[21] are intrinsically based on an experimental design is closely related to the symmetric nature of the relation between a theory based on abstractions and the natural phenomenon we investigate via observation and experiment. It is due to the experimental design that there are the necessary 'checks and balances' on the relation between theory and practice, and due to their quantitative nature abstractions respect those constraints. In its turn, this relates to the primarily methodological nature of abstractions: they do not change the nature of the object of study. This means that a theoretical prediction can be tested by means of an application on the original, natural phenomenon, precisely because the parameter which the abstraction fixes at a certain value, has been preserved in the theory. Linguistics lacks an experimental design, and hence everything that come with it.[22]

[21] 'Natural science' here represents a number of central characteristics of various disciplines, such as physics, biology, chemistry, and so on. Of course we are well aware that a characterisation of the differences between various fields of science is an enormously complicated and, at points, questionable enterprise, and that what is listed in table 2 needs to be extended and nuanced in many ways. However, for our present purposes this rough indication suffices.

[22] To be sure, in psycho(patho)linguistics experiments are being conducted, but these are (almost) never experiments that attempt to test two alternative linguistic theories.

Unlike abstractions, idealisations are not methodological but ontological in nature. They change the object of study, and one of the consequences of this is that there no longer is an immediate relation between the idealised object and the original, natural phenomenon. And that means that predictions derived from the theory can not, at least not as such, be tested by means of an application to the phenomenon. We always need an additional 'bridging' theory that connects the idealised object and the natural phenomenon. Not only is creating such a bridging theory an additional epistemological task, because of the theoretical nature of the idealised object it is very hard to base such a bridging theory on empirical data. And that compromises the empirical nature of the theory based on the idealisation as such.

A possible, and we think plausible, explanation of this difference between natural science and linguistics comes from the nature of their respective ontologies. The natural sciences deal with ontologies consisting of natural phenomena that are subject to strictly deterministic[23] causal laws that can be formulated in quantitative terms. Linguistics, on the other hand,[24] are concerned with an ontology that is not purely natural in the same sense. The phenomena that linguistics studies admittedly have physical, biological, and psychological features, but at the same time they are also historical, social and cultural phenomena. It is the hybrid nature of the ontology that explains why abstractions as we know them from the natural sciences do not occur in linguistics. It also explains why attempts at abstraction result in idealisations, with all the consequences we have outlined above.

If this is right, or at least in the right direction, as we believe it is, it has important consequences for the nature and the goals of theories in linguistics. More about that in section XI.

X Some more examples of idealisations in linguistics

The competence – performance distinction is by far not the only example of an idealisation by means of which modern linguistics has

[23]We disregard the indeterministic nature of quantum physical phenomena, because that is not relevant for the issue at hand.

[24]And more generally, many of the humanities and social sciences.

defined itself as a scientific discipline. Many of the consequences indicated in section III appear to have characteristic features of idealisation, and not those of abstractions.

The idea of language as an infinite object, for example, is closely related to the competence – performance distinction. Modelled after concepts from the formal sciences (mathematical logic, mathematics, computer science)[25] this idea is based on the assumption that actual limitations on the use of, for example, embedding constructions (in terms of memory limitations, finite computational resources, and so on) are not intrinsically part and parcel of what language is. So what can be observed in reality, viz., that such limitations exist, is not considered to be an actual feature of the object 'language', but is taken to be 'merely' the result of intervening factors that as such are not intrinsically tied to the object. Language as we can observe it in actual use (in production and in interpretation) is a phenomenon in which unlimited recursion does not occur. Yet, it is being transformed into an ontologically different kind of object, for which there is no limit to recursive processes.[26]

Another example that was already mentioned is the characteristic, and almost exclusive, emphasis on written language. This also relates to the modelling of the object of modern linguistics on concepts from the formal sciences.[27] From a certain perspective the emphasis on written language seems quite justified: from a practical point of view written language is an object that is much easier to deal with than spoken language. Before the advent of sound registration equipment writing was the only tool that could be used to collect speech and to share observations and analyses of it. In that light, traditional grammars can be considered as compact, codified reports on what could be observed in the field: speech. This is clearly a non-ideological, practical use of a methodological constraint. However, in modern linguistics such practical considerations are clearly

[25] Cf., the afore-mentioned book by Tomalin (Tomalin, 2006).

[26] Cf., Fitz (2009) for extensive discussion, and a neural net model that is able to learn limitations of embedding constructions without an appeal to recursion. Cf., also Pullum and Scholz (2005).

[27] Cf., e.g., Harris (2000) and the already mentioned Kraak (2008) for extensive discussion.

not the only, or even the most important ones. The emphasis on written language also serves to treat 'language' as a well-defined, clearly delineated object. Speech is momentary, context-dependent, and seldom comes alone: prosody, gestures, facial expression, simultaneous interactions with elements of the non-linguistic context, it all occurs and happens at the moment of speaking, and that makes it difficult[28] to isolate as an object of study. Of course we can distinguish between sound and other components, but in particular when questions of meaning and interpretation are at stake that is in many cases not the relevant distinction: all components may contribute to the determination of what is being said. Hence, in so far as written language simply ignores these components, the transition from language as speech to language as writing is a clear example of an idealisation.

Along with this idealisation come yet others. One of them is the idea that competent language users can be considered as 'disembodied' individuals. Of course, embodiment is an essential property of human subjects, and moreover one that is in many respects connected with their being linguistic creatures. The body not only is an important intermediary with our physical environment,[29] it also plays a crucial role in determining the contents of large parts of our mental vocabulary, and it is a reservoir of all kinds of knowledge and abilities that are an integral part both of our linguistic competence, and of the way in which, and the ends to which, we use language. But neglecting embodiment has more effects than leaving out these essential features. The idealised competent language user whose linguistic competence is the central object of study of modern linguistics, is not just accidentally a disembodied subject, it is principally without a body. As was already noticed above, nothing in a theory about the resulting entity contains any lead as to how embodiment might be 'added' to it: the theory about the competent, disembodied language user is supposed to be a complete theory of

[28] And according to some even impossible; cf., Wittgenstein's concept of a 'language game' that explicitly united both verbal and nonverbal elements (Wittgenstein, 1967, section 7).

[29] Something that is revealed in language in many ways, for example in spatial indexicality.

human linguistic competence. From that perspective embodiment is not some real phenomenon from which we abstract, but an irrelevant property of human subjects.

A last example concerns semantics and pragmatics as branches of linguistic theory, and the central role played by the concept of 'propositional content'. The dominant paradigm here relies on a principled distinction between propositional content as semantic meaning, and the use of expressions with such contents that results in pragmatic meaning. With the distinction comes a hierarchical relation: propositional content is independent from pragmatic meaning, whereas the latter needs the former as the base from which it is derived. This is the Gricean model and certainly within linguistics it is still the most used one.[30] In the philosophical literature the distinction as such has been subject of some debate.[31] However, what is relevant to note here is that 'radical contextualism', the view that rejects the distinction, does not seem compatible with the goals of modern linguistics. And that indicates that the concept of propositional meaning as such is yet another example of a construction that is not so much an abstraction as an idealisation.

XI Consequence of idealisation

We hope that the foregoing discussion has made clear that the relationship between, on the one hand, the objects of study that modern linguistics has constructed via idealisations, and, on the other hand, language and linguistic competence as everyday, observable phenomena, is a complicated one, to say the least. This is something that Chomsky seems to recognise as well, as the following passage from Chomsky (1995, p. 20) shows:

> At the conceptual-intentional interface [between sound-meaning pairs of *I*-language and actual language use] the problems are even more obscure, and may well fall beyond human naturalistic inquiry in crucial respects.

[30] Of course there are different views on what exactly the propositional content of an expression is, on how it is to be determined, and, consequently, where exactly the dividing line between semantics and pragmatics is to be drawn. But those discussions still operate within the assumption that the distinction, and the hierarchical relation between the two concepts of meaning, make sense.

[31] Cf., various contributions in Preyer and Peter (2007).

The construction of competence and the accompanying concept of an I-language (roughly, the 'internal language' which the idealised competent language user 'cognises') has distanced the object that according to Chomsky is the proper object of study so far from everyday language and its users that, as he himself acknowledges in this passage, it is not even clear which problems need to be solved in order for us to be able to relate them again. To put it differently, not only is there no bridging theory, it is not even clear what that theory is supposed to do. No doubt this aporetic situation is a direct consequence of the fact that it is not clear at all whether the idealised object puts any empirical constraints on such a theory, and if it does, what these might be.

To those who are primarily interested in language as an empirical phenomenon Chomsky's conclusion will no doubt sound quite defeatist. But Chomsky sees things differently. That, too, is clear from the passage just quoted: it is shown by his use of the qualification 'naturalistic'.[32] The use of this term reveals both a background ideology and an escape from this apparent impasse that Chomsky deems possible. What Chomsky aims at is not just some theory of language and linguistic competence, but one that is naturalistic through-and-through. Language and linguistic competence, as Chomsky sees it, are purely natural phenomena, of the same stature and nature as other human biological capacities and phenomena. For Chomsky the notorious claim 'Language is an organ' is not a metaphor (useful or not), but a factual statement. In the same way, and for the same reasons, that we study the human perceptual apparatus, the human motor system, and other biological capacities, with the means of the natural sciences, we can not but study human linguistic competence, and hence human language in the same manner.

So what motivates the idealisations Chomsky defends, is, as we have indicated before, a ideological position with regard to science and scientific method. It is scientistic naturalism, plain and simple.[33] That such a choice for a strictly naturalistic methodology actually brings about a fundamental shift in ontology, is a consequence that

[32] As for the qualification 'human' in 'human naturalistic inquiry': we don't need to take that too seriously, we think.

[33] Cf., also Lappin et al. (2000) on this issue.

Chomsky is apparently willing to accept, as the following passage from Hauser et al. (2002, p. 1570) shows:

> The word 'language' has highly divergent meaning in different contexts and disciplines. In informal usage, a language is understood as a culturally specific communication system [...] In the varieties of modern linguistics that concern us here, the term 'language' is used quite differently to refer to *an internal component of the mind/brain* [...] We assume that *this* is the primary object of interest for the study of the evolution and function of the language faculty. [emphasis added]

But we do well to note that in this passage more is at stake than accepting the consequence that a naturalistic approach of language and linguistic competence studies a different object than another, more humanities-based approach. Apparently, the point is not to state that there are two (or more) alternative methodologies that we can choose from (and that we perhaps may provide arguments for a particular choice). Rather, what is claimed is that there is only one scientific approach possible in the first place, viz., the naturalistic one. Language and linguistic competence as they present themselves to us in real life, in observations about actual language use, simply are not phenomena that qualify for a scientific investigation.[34]

A last observation concerning the position that is defended here by Chomsky and his associates concerns the scope of the resulting theory. That the linguistic competence of human is rooted also in aspects of their biology is something no-one would doubt. That is a minimal rejection of an ontological dualism that seems quite generally accepted. The real question whether a theory that *reduces* the relevant core concept to biological entities and that accepts *only* a naturalistic methodology, will be able to come up with insightful explanation of properties of the original object of study. As the passage just quoted also illustrates, that seems to be a goal that Chomsky c.s. apparently are not willing to give up on. Their concern is 'the

[34]It is also interesting to note that in this passage the authors speak of 'the varieties of modern linguistics that concern us here'. Apparently, the present-day diversity of approaches (cf., section II) is something that the authors do acknowledge, if only by stating that alternative approaches do not 'concern' them. Cf., also footnote 14.

study of the evolution and *function* of the language faculty' [emphasis added]. Despite the pessimism that Chomsky displayed in the earlier cited passage from Chomsky (1995), the ambition to account for the function of language has not been abandoned, it seems. But in view of the ontological rift that the idealisations that are used have created, it certainly appears doubtful that this ambition can be realised.

XII Consequences of these consequences

What are the consequences for linguistics when its object of study is constructed via idealisation? Of course, it is not possible to answer this question fully and definitively. But what is clear is that the approach that modern linguistics has pursued over the last decades runs into a number of serious difficulties, difficulties that by the way also provide a partial explanation for the curiously diversified state in which linguistics finds itself today (cf., section II).

As was argued in the above in some detail, idealisation results in an ontological shift and creates an additional epistemological task, viz., the formulation of an adequate bridging theory. This leads to a number of problems. First of all, empirically motivated adequacy criteria for the bridging theory are very hard to come by: the idealised object itself does not deliver them, and observations with regard to the original phenomenon can not function as such without further ado. This is a characteristic feature of idealisation, since, as we have seen, abstraction does not run into this problem. The second problem, which is an immediate consequence of the first one, is that there is a serious lack of empirical validation of the theory about the idealised object. Apart from the fact in the case of linguistics the original phenomenon is hard to fit into an experimental design, there is the problem that, without an independently verified bridging theory, no theory about the idealised object will lead to predictions that can be tested on the original phenomenon (via observation or by other means). And thirdly, as a result of that, the intuitive plausibility of the theory is seriously hampered.

Looking at the state of the art in applied linguistics, we see the consequences of this problematic situation clearly emerging. As the theoretical models of the generative tradition, based as they are on

the notion of a grammar as a system of explicit rules, failed to deliver in applications such as machine translation, question-answer systems, and the like, people started to use other constructions of central concept such as 'language', 'meaning', and so on. Often these new constructions were based on stochastic properties and patterns derived from large corpora of actual text (and, later, speech). These constructions were based on other, often less far-reaching idealisations, i.e., they stayed closer to the original phenomenon and hence were more amenable to empirical testing. This development, however, is clearly motivated and steered by practical, pragmatic considerations, rather than by theoretical and explanatory ones. Theory, so it appears, lags behind practical application, which is also why we can observe a certain proliferation of theoretical models that are strongly influenced by very concrete, often also quite limited practical applications. To that extent, we might say that theory has become ad hoc.

In other contexts where linguistics touches on empirical research, another trend is visible. Language and linguistic competence are also important objects of study in the rapidly developing cognitive neurosciences. Inspired by a long tradition of psycholinguistic research, in particular research on language disorders, linguists have taken up the challenge provided by new, non-invasive techniques of studying the brain. The problems that occur here are partly related to the strongly naturalistic and reductionistic nature of a lot of neurophysiological and brain research, partly they are due to the inherent limitations of the kind of experiments that the new techniques allow. One of the consequences is a reinforcement of the kind of idealisation that we have discussed in the foregoing, in particular the individualistic nature of the competent language user, and an accompanying diminishing possibility of linguistic theory to come up with leading hypotheses and testable predictions.

XIII Conclusions

What conclusions can be drawn from these observations? Obviously, more research into the way in which linguistics, especially in its present-day diversity, copes with its central concepts, is needed. But

one question will be central: Is naturalism in linguistics a methodology that is forced upon us by the nature of the phenomena it studies? Or is it a choice? The observations and considerations put forward in this paper strongly suggest that the latter answer is the correct one: the naturalism that is so characteristic for modern linguistics, in particular, but not exclusively, for the generative tradition, is based on a scientistic ideology. Note that as such, that does not imply that the resulting methodology is necessarily the wrong one. (It could be the right choice made for the wrong reasons.) But it does show along which lines further research in this area should be conducted: it is the consequences of this choice that need to be thoroughly scrutinised.

Should it turn out, as we strongly suspect it will, that the ideologically motivated choice for naturalism severely hampers the explanatory power of the resulting linguistic theory, then that by itself provides a clear pointer to the direction in which one may look for alternatives. For that a naturalistic approach that is not ideologically motivated may lead to interesting and, to some extent, testable results is shown by various alternative theoretical frameworks that, partly as a response to the deficiencies of work done in the generative tradition, have been developed over the last decade or so. Examples are cognitive linguistics[35], stochastic linguistics,[36] and approaches in which neuronal models of language acquisition and language use are studied.[37]

References

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Mass.

Chomsky, N. (1980). *Rules and Representations*. Blackwell, Oxford.

Chomsky, N. (1986). *Knowledge of Language*. Praeger, New York.

Chomsky, N. (1995). Language and nature. *Mind*, 104(413):1–61.

[35] Cf., the recent Tomasello (2003).
[36] For an interesting recent example, cf., Daelemans and Van den Bosch (2005).
[37] Cf., e.g., MacWhinney and Chang (1995) and the afore-mentioned Fitz (2009)

Daelemans, W. and Van den Bosch, A. (2005). *Memory-Based Language Processing. Studies in Natural Language Processing*. Cambridge University Press, Cambridge.

Fitz, H. (2009). *Neural Syntax*. PhD thesis, ILLC/Department of Philosophy, Amsterdam.

Groenendijk, J. and Stokhof, M. (2005). Why compositionality? In Carlson, G. and Pelletier, J., editors, *Reference and Quantification: The Partee Effect*, pages 83–106. CSLI, Stanford.

Harris, R. (2000). *Rethinking Writing*. Athlone, London.

Harris, R. A. (1993). *The Linguistics Wars*. Oxford University Press, Oxford.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598):1569–1579.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. MIT Press, Cambridge, Mass.

Jones, M. R. (2005). Idealization and abstraction: A framework. In Jones, M. R. and Cartwright, N., editors, *Idealization XII: Correcting the Model. Idealization and Abstraction in the Sciences*, pages 173–217. Rodopi, Amsterdam. *Poznań Studies in the Philosophy of the Sciences and the Humanities*, vol. 86.

Kamp, H. and Stokhof, M. (2008). Information in natural language. In van Benthem, J. and Adriaans, P., editors, *Handbook of Philosophy of Information*, pages 49–112. Elsevier, Amsterdam.

Kraak, A. (2008). *Homo Loquens en Homo Scribens*. Amsterdam University Press, Amsterdam.

Lappin, S., Levine, R. D., and Johnson, D. E. (2000). The structure of unscientific revolutions. *Natural Language and Linguistic Theory*, 18:665–671.

MacWhinney, B. and Chang, F. (1995). Connectionism and language learning. In Nelson, C. A., editor, *Basic and Applied Perspectives on Learning, Cognition, and Development*, pages 33–58. Erlbaum.

Newmeyer, F. J. (1986). *The Politics of Linguistics*. The University of Chicago Press, Chicago.

Postal, P. M. (2004). *Skeptical Linguistic Essays*. Oxford University Press, Oxford.

Preyer, G. and Peter, G., editors (2007). *Context-sensitivity and Semantic Minimalism: New Essays on Semantics and Pragmatics*. Oxford University Press, Oxford.

Pullum, G. K. and Scholz, B. C. (2005). Contrasting applications of logic in natural language syntactic description. In Hájek, P., Valdés-Villanueva, L., and Westerståhl, D., editors, *Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress*, pages 481–503. King's College Publications, London.

Stein, H. (1997). *The Fibre and the Fabric*. PhD thesis, ILLC/Faculteit der Wijsbegeerte, Universiteit van Amsterdam, Amsterdam.

Stokhof, M. (2002). Meaning, interpretation and semantics. In Barker-Plummer, D., Beaver, D. I., van Benthem, J., and di Luzio, P. S., editors, *Words, Proofs and Diagrams*, pages 217–240. CSLI, Stanford.

Stokhof, M. (2007). Hand or hammer? On formal and natural languages in semantics. *Journal of Indian Philosophy*, 35(5):597–626.

Stokhof, M. (2008). The architecture of meaning: Wittgenstein's *Tractatus* and formal semantics. In Levy, D. and Zamuner, E., editors, *Wittgenstein's Enduring Arguments*, pages 211–244. Routledge, London.

Tomalin, M. (2006). *Linguistics and the Formal Sciences: The Origins of Generative Grammar*. Cambridge University Press, Cambridge.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, Cambridge, Mass.

Wittgenstein, L. (1967). *Philosophische Untersuchungen*. Suhrkamp, Frankfurt a/M.

Achourioti, T., Andrade, E.J., Staudacher, M. (eds.), Proceedings of the Amsterdam Graduate Philosophy Conference — Meaning and Truth, ILLC Publications X-2010-1, pages 39–51

Against Crude Semantic Realism?

Florian Demont (King's College London and University of Zurich)

florian.demont@web.de

I Introduction

A common way of spelling out a factualist conception of meaning, what appears to be the canonical option, is Semantic Realism. I shall not discuss Semantic Realism in general, but a crude version of it, one that holds that there is 1) exactly one relation between proper names and objects (the word-referent relation), 2) that this relation is paradigmatic for semantics, and 3) that the relation is not normative, i.e., does not entail what one should say if it holds. According to this view, the prime fact about meaning is the word-referent relation which Frege had introduced for complete expressions minus a specific conception of linguistic normativity that often goes together with it. In the first section, Semantic Realism and linguistic normativity will be introduced in George Wilson's (1994) terms. Crude Semantic Realism will then be explained as a departure from these specific notions. The second section will present linguistic considerations against it. The third section will present philosophical considerations which attempt to show that Crude Semantic Realism cannot fully explain the fundamental character of meaning. The fourth, and last, section draws a moral for the agenda of Semantic Realism.

II A Guide to Descending into Crude Semantic Realism

In 1982 Saul Kripke presented a sceptical challenge for any possible conception of meaning (Kripke, 1982). He argued that the challenge can be gathered from various writings of Ludwig Wittgenstein and that there is also a solution for the challenge. The 'sceptical' solution that Kripke presents us with has usually been read as a form of non-factualism about meaning. Colin McGinn (1984), Paul Boghossian (1989) and Crispin Wright (1984; 2001, ch 4) have championed this reading and Alexander Miller (2009) has resourcefully defended its general gist against George Wilson's dissenting view by presenting an amended non-factualist reading. The present discussion does not depend on details about Kripke's challenge, his solution, or the debate it has produced. Rather, it will take the intuition behind the dissent and independently assess a specific elaboration of it.

George Wilson (1994) has argued that Kripke's solution features a factualism about meaning. The reason why the standard reading is wrong is because it fails to identify the exact target of the sceptical challenge. At the heart of Wilson's alternative reading lies the view that Kripke's sceptic assumes Classical Realism about meaning-facts, whereas Kripke's Wittgenstein does not. In other words, Kripke's Wittgenstein uses the sceptical challenge to reject Classical Realism, which Wilson (1994, p. 244) exemplifies as follows—call this instance 'Semantic Realism':[1]

> Let 'a' denote an object o, 'ϕ' a property ϕ and let F be the (possible) state of affairs of o's being ϕ. Then:
>
> (1) 'a is ϕ' is true if and only if F obtains

The reason why, according to Wilson (1994, p. 239), Kripke's Wittgenstein wants to reject Semantic Realism is because it leads to a substantial scepticism together with the intuition that meaning is normative. The Classical Realist has the following conception of linguistic normativity—call it 'Ns':

[1] Note that I have changed Wilson's notation. I am indebted to Kai Büttel for suggesting the alternative, he also helped improving the notation of the other definitions.

> If X means something by a term 'ϕ' or 'a', then there is a set of properties $P_1 \dots P_n$ of at least one o in the domain that govern the correct application of 'ϕ' or 'a' for X.

The reason why linguistic normativity and Semantic Realism go together well is that the normativity thesis explains what is special about the relation between word and referent: the relation governs correctness and, ultimately, explains meaning because meaning is inherently normative, i.e., the properties governing correct applications entail prescriptive rules for applications of terms. Furthermore, only because of the fact that meaning is inherently normative does the word-world relation turn out to be primitive, for prescriptions are commonly assumed not to be reducible to descriptions.

It does not matter for present purposes how Ns and Semantic Realism give raise to a sceptical challenge and whether this is an accurate reading of Kripke (or Wittgenstein). What is important is that one might want to preserve some form of Semantic Realism. After all, 'The banana is green' is true iff the banana is green and false if, for example, it is already yellow (or brown). One might find this intuition very powerful and, therefore, accept that Ns does not hold in order to save it from scepticism. This amounts to Crude Semantic Realism—the position that Classical Semantic Realism is true and that meaning is not normative, that Ns is false if it is read as entailing any prescriptive rules. The resulting position is crude, because it dispenses with the idea that the word-referent relation is primitive due to the normativity of meaning. The question then is whether Crude Semantic Realism can, without significant alterations, make sense of the word-referent relation in a way that is substantial enough to fully explain meaning. Before we come to that assessment, let me introduce a proposal that classifies as Crude Semantic Realism.

Paul Boghossian (2005) and Anandi Hattiangadi (2006, 2007) have, amongst others, forcefully argued against the normativity of meaning and have, thus, adopted Crude Semantic Realism. Only Hattiangadi's proposal will play a role here in order to keep the argument concise and illustrate the general strategy against Crude Semantic Realism.

Hattiangadi (2007, pp. 179–180), after having presented an insightful discussion of most of the debate on the fundamental issues

that Kripke had raised, introduces a distinction which, if applied to Wilson's definition of linguistic normativity, reveals an ambiguity. The word 'govern' in Ns can mean two things: 1) X applies 'ϕ' or 'a' correctly, 2) X should apply 'ϕ' or 'a' correctly. On the first reading it implies, at best, a constitutive rule, while it clearly implies a prescription on the second reading. On Hattiangadi's view, the first reading is plausible, but the second one has a bogus result: scepticism about meaning. To get rid of prescriptive rules in semantics, she introduces a new notion, one which only allows constitutive rules: 'Linguistic Norm-Relativity':[2, 3]

(2) X means ϕ by 'ϕ' $\rightarrow \forall o(X$ applies 'ϕ' correctly to $o \leftrightarrow o$ is $\phi)$

As far as the sceptical challenge is concerned, Hattiangadi seems, for reasons not to be rehearsed here, warranted in assuming that Crude Semantic Realism is a plausible conception of meaning.

So far, Crude Semantic Realism has been introduced and the motivation for it has been sketched. An assessment on independent grounds follows in the next two sections. The leading question is: can Crude Semantic Realism make sense of the word-referent relation in a way that is substantial enough to fully explain meaning?

III Spanish DOM and Japanese Occurrence

In order to formulate the linguistic considerations against Crude Semantic Realism, more details on Hattiangadi's position are needed. It seems useful to start with her general reasons for rejecting semantic non-factualism. The worry here is that it is hard to make sense of a substantial conception of meaning without basing the idea that somebody means something on the assumption that 'the judgements or statements of our semantic discourse [i.e., the practice of ascribing meaning] are legitimate in some sense—either true in a weakly deflationary sense or justified' (Hattiangadi, 2007, p. 102). After all, if it was not usually the case that people are right in meaning something

[2]From here on I shall only talk about the application of predicates like 'ϕ' to properties, I assume that the lessons for naming can easily be draw from this.

[3]Note that I have altered Hattiangadi's notation as well.

rather than nothing, one could hardly claim that the underlying conception of meaning was substantial enough, because meaning anything, then, would be arbitrary. In other words, without anything like Linguistic Norm-Relativity or another conception of correctness, a notion of meaning can hardly be said to be interesting at all.

Semantic non-factualism must, thus, appear incoherent, as it seemingly requires a notion of correctness without presupposing that there is anything that can be correct or incorrect beforehand. There must be a content that can be true or false behind every statement or judgement in semantic discourse if semantic discourse is not to be arbitrary. Hattiangadi is therefore committed to the idea that judgements and statements always have a semantic content, one that can be true or false. So far so good.

Consider what she says about judgements. In her discussion of answers to the sceptical challenge she assesses Crispin Wright's judgement-dependent account of meaning and intention (cf. Wright, 2004), according to which the extension of the truth-predicate for claims about what we mean or intend depends on the form of judgement we base these claims on. Such judgements are therefore called 'extension-determining'. Her objection to this proposal is as follows:

> The problem, quite simply, is that Wright appeals to my judgements, and the sceptic can always question the contents of those judgements. The sceptic can accept that John's intention to [θ] is constituted by his judgement that he intends to [θ], she can still ask what makes it true that John judges *that he intends to [θ]* (Hattiangadi, 2007, p. 159).

Hattiangadi supposes here a difference between the constitution of semantic content and its actual interpretation, the truth (or falsehood) of any 'θ' that has semantic content. This amounts to the claim, if seen together with the definition of Semantic Realism, that semantic content must be constituted independently of what is actually true.[4] In other words, there is a semantic content which refers to states of affairs in virtue of containing names and predicates which, considered by themselves, refer to objects and properties—reference

[4]See also Hattiangadi (2007, pp. 12–13) where she explicitly makes the independence claim.

(the word-referent relation) is the bridge between the two unconnected shores: content and states of affairs. On that view, an *a priori* covariation of some judgement that one intends to θ (or any other candidate content of judgement) and some states of affairs, as Wright proposed, is never licit.

Now, Wright claims that there is a class of judgements for which there is no difference between constitution of content and interpretation of content. Such a content must, under ideal circumstances, be actual if it is possible at all—which is to say that it must be necessary. If there are necessarily true judgements, the constitution of the semantic content and its interpretation happen at once. Compelling examples, ideally not restricted to any idiosyncratic philosophical account of judgement (such as Wright's), are needed to suggest that there is more about semantic content and, thus, meaning, than Crude Semantic Realism can handle. Finding such examples and arguing for their relevance will require a few linguistic examples:

(3) *Jorge besó a Lucía*
George kissed to(DOM) Lucia
"George kissed Lucia"

(4) *Pedro besó el retrato*
Peter kissed the picture
"Peter kissed the picture"

(5) *Kasa ga arimasu*
Umbrella (subject-marker) there-is
"There is an umbrella"

(6) *Neko ga imasu*
Cat (subject-marker) there-is
"There is a cat"

The linguist George Bossong has introduced the concept of 'differential object marking' (DOM), a linguistic feature of more than 300 languages (Bossong, 1984). Examples (3) and (4) are well-formed sentences in Spanish. (3) contains the DOM 'a' which marks 'Lucía' as a living being, (4) does not require such a marker, because pictures are not living beings. If Lucia was a sports car, there would

be no marker in (3) either. Of relevance here is that every semantic content has a syntactic structure, it must be (at least) well-formed to be true or false. In both examples, the well-formed sentence conveys semantic information to anybody who understands Spanish by using one grammatical form rather than another. In the Japanese sentences (5) and (6) we find a similar grammatical feature. The occurrence of living beings (such as cats) requires the predicate 'imasu', while the occurrence of an object requires 'arimasu'.[5]

Based on such examples, philosophers of language and theoretical linguists rooted in recent developments of Chomskian grammar have argued that such grammatical findings do tell us something about conceptual capacities.[6] They argue that the conceptual capacity to distinguish, for example, between living and non-living beings is a feature of our grammatical capacities. More specifically, Wolfram Hinzen (2007) has argued that certain lexical items, amongst which we may want to include the predicate 'living being', have an internal structure which determines how the item is to be combined with others when well-formed expressions in a language are built, but that structure itself is said to have no compositional semantics—i.e., it must be taken as primitive. I shall explore this line of thought with as few Chomskian assumptions as possible in order to come up with a counterexample to Hattiangadi.

Since Noam Chomsky's first mention of language acquisition in relation to grammatical form (Chomsky, 1965), it has become increasingly well corroborated that grammatical features are, by and large, innate. But what does 'innate' mean here? The notion does not require much work if we follow Gabriel Segal:

> The work is done by the negative clause: 'not acquired by a psychological process'. Thus empiricists propose that their chosen explanandum is acquired by: induction, abduction, deduction, perception, statistical inference, conditioning etc... Nativists, say: no,

[5]Note that trees and plants in general do not count as living being when it comes to the grammatical distinction between animate and inanimate objects. One might, thus, prefer to say that there is a grammatical distinction between sentient and non-sentient beings. I shall, however, not make use of those terms.

[6]I write from a minimalist point of view as introduced in Lasnik and Uriagereka (2005), but I hope that other views associated with Noam Chomsky find that my arguments pass muster.

> it was acquired some other way. It has to be considered primitive from the psychological point of view—something that we assume that the learner has before the process of learning begins. If alien scientists programmed us with UG [Universal Grammar] when we were foetuses, then I take it that nativists come out as more nearly right than empiricists (Segal, 2008, p. 91).

If this is true, DOM and any other grammatically realised animate-inanimate distinction like the 'imasu'/'arimasu' distinction in Japanese are innate as well, because they are grammatical features. The argument for this is standard (this is an adapted version of the argument in Segal, 2008):

1. All normal first-language speakers of Spanish and Japanese acquire DOM or the 'imasu'/'arimasu' distinction during their childhood.

2. Children acquire DOM or the 'imasu'/'arimasu' distinction via general purpose learning mechanisms (viz. induction, abduction, deduction, perception, statistical inference, conditioning) or those features are, at least partially, innate.

3. The stimuli are too poor to provide children with the empirical basis sufficient for acquiring DOM or the 'imasu'/'arimasu' distinction.

4. Children do not learn DOM or the 'imasu'/'arimasu' distinction via general purpose learning mechanisms.

5. DOM or the 'imasu'/'arimasu' distinction are, at least partially innate.

The reason why premise 3 holds is that the evidence for syntactic structures in general is strong enough that it seems safe to assume it true for DOM or the imasu/arimasu distinction.[7] Now suppose that 1) syntax constrains referring expressions, because only well-formed expressions can refer at all and that 2) a syntactic distinction between living and non-living beings will inform every explicit reference to

[7] See (Boeckx, 2006, §2.3) for a general wording of the argument and a bibliography.

living and non-living being. Under these assumptions, we can hold, tentatively approximating Hinzen's position, that the conceptual capacity to distinguish between living and non-living is somehow inscribed into the innate capacity to build grammatically well-formed sentences. It does not really matter if the conceptual distinction has, maybe even after a genuine learning process, informed the lexicon from which the syntactic operations construct grammatical sentences containing DOM or the 'imasu'/'arimasu' distinction. What matters is that the conceptual distinction informs the generating process of linguistic content without any influence from experience during the generating.

This amounts to the commitment that there are two categorically different ways of referring to objects: one relates a name and a living being (e.g., 'neko' and cats), the other a name and an object (e.g., 'kasa' and umbrella). It also entails that the syntactic distinction is innate in native speakers of English, even though it does not show up in the structure of actual sentences in English. How does that help in finding a counterexample for the conception of judgement that is implied by Crude Semantic Realism?

If the linguistic consideration is correct, expressing the thought 'The cat is a living being' in any language is, under the assumption that 'The cat' actually refers, the sort of counterexample we need. Grammar tells us that forming a sentence about any cat must involve that cats are living beings; if the predicate is 'is a living being', the sentence is true in virtue of its grammatical form whenever its subject 'The cat' actually refers to a cat. Meaning that the cat is a living being by saying so is a case which Crude Semantic Realism cannot handle, because it poses an *a priori* independence between constitution of content and its interpretation. It is the sort of counterexample needed against Hattiangadi.

The linguistic examples and their interpretation suggest that sentences are not only true because they successfully track states of affairs (and simply reflect them). It seems that truth does also depend, at least partially, on a structural affinity between sentences and states of affairs: of actual cats being living things in every state of affairs and in every linguistic content we form in order to refer to actual cats.

The problem for Crude Semantic Realism is clear: by stripping off all normative force from the word-referent relation it became possible to relocate the source of correctness in the syntactic constitution of semantic content and states of affairs rather than in their relation. Doing so showed us that there are meaningful sentences, like 'The cat is a living being' (when referring to an actual cat), where we cannot uphold the sort of independence between content and states of affairs upon which any Semantic Realism is built. We can now argue that Crude Semantic Realism cannot explain the meaning of certain grammatical sentences which obviously do have a meaning. I shall elaborate this point in the next section.

IV Explaining the Possibility of Meaning

Meaning ϕ by saying 'ϕ' can be explained in different ways. It was assumed here that trying to find the right meaning-facts is a good strategy. Crude Semantic Realism adduces one sort of fact: meaning ϕ by 'ϕ' amounts to whether 'ϕ' refers to a property which is part of a possible state of affairs that actually obtains or not. The crucial word here is 'refers'. Any spawn of Classical Realism gives reference a prime role.

How does Crude Semantic Realism explain meaning? The explanation it proposes seems to be complete once one has a conception of reference which does not entail any normative properties or relations and according to which reference is supposed to be primitive and also the right meaning fact. Does this suffice? The crucial point is that it can be questioned whether this view can account for the constitution of semantic content and the state of affairs it corresponds to: reference presupposes that there is a 'ϕ' which can be related to some property which partakes in a (possible) state of affairs. In other words, by assuming that reference is primitive, Crude Semantic Realism cannot explain how reference is possible at all: it cannot fend of calls for reducing reference to some further fact, as claiming a primitive notion of reference is not based on the claim that meaning is normative and, hence, irreducible. Now, under the assumption that there is no such thing as semantic normativity, any explanation of how 'reference' is be understood here would contain an account of

what can be related and how such a relation works—but it turned out (in section 2) that different things are related in different ways. Suggesting, as the definition presented in section 1 does, that some paradigmatic sort of word referent relation is the right prototype does, hence, not suffice.

Crude Semantic Realism is no full-fledged factualism about meaning, because it does not offer us a story about how the (allegedly) relevant meaning-fact constitutes meaning. The argument that denying this leads to a contradiction is rather simple (note, premise 4 results from the interpretation of the linguistic consideration):

1. Crude Semantic Realism is a factualism about meaning.
2. Every factualism about meaning assumes some basic fact.
3. Classical Semantic Realism assumes reference as the basic fact.
4. Reference in Crude Semantic Realism is not a basic fact, because it must partially rely on how linguistic content is syntactically constrained.
5. Therefore, Crude Semantic Realism is not a Classical Semantic Realism.
6. Therefore, Crude Semantic Realism is not a factualism about meaning.

The way out is to provide a suitable basic fact. I have, following Hinzen, suggested that working out the syntactic constraints on the correct application of terms can turn out to close the gap. A positive account of how linguistic content is generated can explain the structural affinities between content and states of affairs that the realistic doctrine of independent constitution has problems with. Meaning-theories based on an understanding of generative syntax are readily available and it is unclear why the Crude Semantic Realist should not want to join that camp.[8] If such an assimilation is

[8] Cf. Larson and Segal (1995) for how to assimilate truth-conditional semantics to Universal Grammar. See also Pietroski (2005) for a more recent proposal.

accepted and worked out in detail, some refined version of Semantic Realism might become possible—but choosing a crude version as one's starting point misses the real issue.

References

Boeckx, C. (2006). *Linguistic Minimalism. Origins, Concepts, Methods, and Aims*. Oxford University Press, Oxford.

Boghossian, P. A. (1989). The rule-following considerations. *Mind*, 98:507–49. Quoted from the edition in Miller and Wright (2004).

Boghossian, P. A. (2005). Is meaning normative? In Beckermann, A. and Nimtz, C., editors, *Philosophy—Science—Scientific Philosophy*. Mentis, Paderborn.

Bossong, G. (1983/1984). Animacy and markedness in universal grammar. *Glossologia*, 2/3:7–20.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Mass.

Hattiangadi, A. (2006). Is meaning normative? *Mind and Language*, 21:220–40.

Hattiangadi, A. (2007). *Oughts and Thoughts. Scepticism and the Normativity of Meaning*. Oxford University Press, Oxford.

Hinzen, W. (2007). *An Essay on Names and Truth*. Oxford University Press, Oxford.

Kripke, S. A. (1982). *Wittgenstein On Rules and Private Language*. Blackwell, Oxford.

Larson, R. and Segal, G. (1995). *Knowledge of Meaning*. MIT Press, Cambridge (Mass.).

Lasnik, H. and Uriagereka, J. (2005). *A Course in Minimalist Syntax. Foundations and Prospects*. Blackwell, Oxford.

McGinn, C. (1984). *Wittgenstein on Meaning*. Blackwell, Oxford.

Miller, A. (2009). Kripke's wittgenstein, factualism and meaning. In Whiting, D., editor, *The Later Wittgenstein on Meaning*. Palgrave.

Miller, A. and Wright, C., editors (2004). *Rule-Following and Meaning*. Acumen, Chesham.

Pietroski, P. M. (2005). *Events and Semantic Architecture*. Oxford University Press, Oxford.

Segal, G. (2008). Poverty of stimulus arguments concerning language and folk psychology'. In Carruthers, P., Laurence, S., and Stich, S., editors, *The Innate Mind. Foundations and the Future*. Oxford University Press, Oxford. Volume 3.

Wilson, G. (1994). Kripke on wittgenstein on normativity. *Midwest Studies in Philosophy*, 19:366–390. Quoted from the edition in Miller and Wright (2004).

Wright, C. (1984). Kripke's account of the argument against private language. *Journal of Philosophy*, 81:759–778.

Wright, C. (2001). *Rails to Infinity*. Harvard University Press, Cambridge (Mass.).

Wright, C. (2004). Meaning and intention as judgement dependent. In Miller, A. and Wright, C., editors, *Rule-following and Meaning*, pages 129–40. Acumen, Chesham.

Achourioti, T., Andrade, E.J., Staudacher, M. (eds.), Proceedings of
the Amsterdam Graduate Philosophy Conference — Meaning and Truth,
ILLC Publications X-2010-1, pages 52–66

Anaphoric Deflationism and Theories of Meaning

David Löwenstein (Freie Universität Berlin)

david.loewenstein@fu-berlin.de

I Introduction

It is widely held that truth and reference play an indispensable explanatory role in theories of meaning. By contrast, so-called deflationists argue that the functions of these concepts are merely expressive and never explanatory. Robert Brandom has proposed both a variety of deflationism—the anaphoric theory—, and a theory of meaning—inferentialism—which doesn't rely on truth or reference. He argues that the anaphoric theory counts against his (chiefly referentialist) rivals in the debate on meaning and thereby paves the way for inferentialism.

In this paper, I give a friendly reconstruction of anaphoric deflationism (section II) and point to a distinguishing feature of the theory with respect to other deflationist proposals. While Brandom simply assumes, but doesn't earn this feature, I propose a natural argument to justify it (section III). Then, however, I point out a subtle but clear sense in which truth and reference can play a role in explaining meaning, even if the anaphoric theory holds. Thus, anaphoric deflationism will turn out to be neutral in the debate on meaning (section IV).

II Anaphoric deflationism

While there is some disagreement over how best to define deflationism, I take it that the most fruitful proposal relies on distinguishing two kinds of questions (compare Armour-Garb and Beall 2005).

THE SUBSTANTIAL QUESTIONS What is truth? What is reference?

THE FUNCTIONAL QUESTIONS What is the conceptual role of truth and the linguistic role of 'true' and related expressions? What is the conceptual role of reference and the linguistic role of 'refers to' and related expressions?

Traditionally, philosophers take both kinds of questions seriously. They give an account of what truth and reference are, and then employ this account to explain their role in thought and talk. I will call these proposals 'substantial theories of truth and reference' or 'substantivism'. By contrast, deflationists think that asking what truth and reference *are* is at least uninteresting and maybe even deeply misguided. Rather, answering the functional questions is the only interesting thing to do. In shrugging their shoulders at the substantial questions, deflationists are local metaphysical quietists: As for the alleged *nature* of truth and reference, they don't undertake any ontological commitment whatsoever.

Deflationary theories differ in how to make sense of the functional roles of truth and reference. As its name already reveals, the anaphoric theory relies on anaphoric mechanisms. [1] That is, it relies on relations between linguistic expressions that allow some expression tokenings, anaphoric dependents, to anaphorically inherit the meanings of other expression tokenings, anaphoric antecedents. Pronouns are paradigmatic examples for anaphoric expressions. Consider the following example.

(1) Even though *a friend of mine* strongly believes in *him*, *she* could never convince me to believe in *God* myself.

[1] For the original presentation and development of the theory, compare Grover et al. 1975, Brandom 1984, 1988, 1994 and 2002.

Here, the tokening of 'she' is anaphorically dependent on its antecedent 'a friend of mine'. It inherits its meaning by *referring back* to that antecedent. Also, the tokening of 'him' is dependent on its antecedent 'God'. In this case, the anaphoric dependent refers *forth* to that antecedent and thereby inherits its meaning.[2]

II.1 Truth

The main idea of the anaphoric account of truth is to treat complex expressions formed with expressions like 'is true' as *prosentences*. When concerned merely with truth and not with reference, anaphoric deflationism is therefore also often referred to as 'prosententialism'.

Prosentences are anaphorically dependent sentences that inherit their meanings from other sentence tokenings. Truth talk, the claim goes, does not involve attributions of a property to, say, sentences or propositions. Rather, truth talk relies on a unique anaphoric prosentence-forming operator.

Consider the following example of ordinary truth talk.

(2) Sarah's self-description is true.

(2) is understood as a prosentence which inherits its meaning from a claim by Sarah, a sentence tokening picked out by the expression 'Sarah's self-description'. For example, in a context where Sarah has said 'I am a maverick', (2) means

(3) Sarah is a maverick.

Other instances of truth talk add a quantificational dimension to the story. For example, (4) is understood as (5) and (6) as (7).

(4) Some of these provocative remarks are actually true.

(5) There are some sentence tokenings 't' such that (i) 't' is among these provocative remarks, and (ii) actually, t.

(6) Everything John just said is true.

[2]Linguists call the latter phenomenon 'cataphora' and reserve the term 'anaphora' for 'forwards' cases like the former. Cataphora and anaphora are, in turn, subclasses of endophora. Thus, strictly speaking, the theory is endophoric rather than anaphoric.

(7) For all sentence tokenings 't', if John just said 't', then t.

Given the occurrence of the variables both inside and outside of quotation marks, the quantifiers have to be understood substitutionally.

Prosentences are generic. That is, any sentence tokening can be the antecedent of a prosentence, given that it is specified uniquely, e.g., by description or quotation. Prosentences inherit their meanings from the set of their anaphoric antecedents—a singleton in 'lazy' cases like (2) as opposed to quantificational cases like (4) and (6). Prosentences can occur free-standing or embedded in logically complex sentences, e.g., as antecedents of conditionals.[3]

But what are prosentences good for? The answer is that they enable speakers to say things they couldn't say otherwise. For example, one can blindly endorse or reject claims, as in (8), one can reject one from a set of claims without specifying which, like in (9), and one can endorse large or even infinite sets of claims in one breath, as in (10).

(8) What the editorial of tomorrow's *Times* will say is true.

(9) At least one of Brandom's claims is false.

(10) All theorems of first-order logic are true.

Finally, the anaphoric theory also implies an important scheme:

TRUTH SCHEME 'ϕ' is true just in case ϕ.

The left hand side of this biconditional is interpreted as a prosentence, and if the sentence quoted on the left is the one used on the right, they will share their meaning by anaphoric inheritance and thereby validate the biconditional.

II.2 Reference

The main idea of the anaphoric account of reference is to treat complex expressions formed with expressions like 'refers to' as *indirect definite descriptions*. These are anaphorically dependent singular terms which inherit their meanings from other singular term

[3]This implies that the Liar is a prima facie admissible prosentence. Grover (1977) has attempted a prosententialist dissolution of the Liar, but I cannot discuss these issues here.

tokenings. Reference talk, the claim goes, does not involve attributions of a relation between linguistic expressions and objects. Rather, reference talk relies on a unique anaphoric pronoun-forming operator.

Take this example to help clarify the idea:

(11) What you referred to as 'that cat' was actually a fox.

Here, the indirect definite description 'what you referred to as 'that cat" inherits its meaning from the addressee's latest tokening of 'that cat'. Assuming a suitable context, it is therefore a way of saying

(12) What we saw in the garden was actually a fox.

Like prosentences, indirect definite descriptions can be used quantificationally:

(13) All of these names refer to friends of mine.

(14) I referred to Plato earlier.

Like quantified truth talk, (13) and (14) are elaborated as the following substitutionally quantified statements:

(15) For all singular term tokenings 't', if 't' is among these names, then the one referred to by 't' is a friend of mine.

(16) There are some singular term tokenings 't' with the property that (i) I uttered 't' earlier and (ii) the one referred to by 't' is Plato.

What are indirect definite descriptions good for? Just like prosentences, they enrich the expressive resources of a language. Chiefly, they allow for speakers to continue any anaphoric chain of singular term tokenings, as long as the antecedent tokening is specified uniquely. This includes both tokenings of a type-substitution invariant expressions, such as proper names, and expressions not cotypically substitutable, such as pronouns. Furthermore, in the quantificational cases, one may use, say, (13) in order to avoid giving

all the names in question. And one may employ (14) regardless of whether one has used the expression 'Plato' or 'Aristotle's teacher' or anything like that.

Finally, the anaphoric theory also implies an important scheme:

REFERENCE SCHEME 'ϕ' refers to ψ just in case ϕ is ψ.

Changing merely what is focused on rather than its meaning, the left hand side of this scheme is reformulated as an identity statement:

REFERENCE SCHEME* The one referred to by 'ϕ' is ψ just in case ϕ is ψ.

Now, the expression 'the one referred to by 'ϕ'' is interpreted as an indirect definite description which inherits its content from some previous tokening of 'ϕ'. The scheme is thereby validated whenever the 'ϕ' quoted on its left hand side is the one used on its right. Obviously, this includes all type-substitution invariant expressions 'ϕ', e.g., proper names.

III Anaphoric operators

On the anaphoric theory, prosentences are defined as the unique expression type formed by the truth operator, and indirect definite descriptions as the one formed by the reference operator. How should these anaphoric operators be understood?

III.1 The need for demarcation

Let me elaborate why this question is crucial for proponents of the anaphoric theory.

A standard way to support the idea, against deflationists, that substantive accounts of truth and reference are called for, is to say that *since* 'is true' and 'refers to' are predicates, there have to be underlying properties which substantive accounts of truth and reference need to analyze. Other varieties of deflationism react to this argument by saying that the properties in question are merely deflationary or purely logical. Or they deny that there are properties at all underlying the predicates in question. By contrast, the anaphoric theory is more radical. It denies that truth and reference talk are

predicative at all and holds that it involves unique anaphoric operators instead. This strategy blocks the move from language use to properties from the outset.

Moreover, it also comes in handy elsewhere: Michael Devitt (2002) has argued that deflationism has hidden metaphysical committments. He holds that metaphysics is "explanatory prior" to language (Devitt, 2002, pp. 61, 63), and argues that

> an anti-realist metaphysics is needed to *motivate* the revisionist view of language [...]. If there were not something problematic about the area of reality that 'true' [..] appear[s] to concern why suppose that [it] does not have the standard semantics of a descriptive predicate? (Devitt, 2002, p. 65)

But this complaint is based on the premise that 'is true' is a predicate. Thus, if the anaphoric theory can account for truth talk in terms of an anaphoric operator rather than a predicate, Devitt's criticism misfires. Analogous arguments apply to 'refers to'.

III.2 Defining the operators

I have argued that the anaphoric theory relies heavily on the distinction between the analysis of truth and reference talk as predicative and their analysis as anaphoric. But what accounts for this distinction? Brandom explicitly defines the reference operator and thus distinguishes reference talk from predication. Surprisingly, however, he remains silent about the truth operator. In what follows, I will argue that a suitable definition of the latter can be modeled on the former.[4]

Brandom claims that an expression qualifies as an instance of the reference operator ϱ just in case it satisfies the following criterion solely in virtue of its meaning.

ITERATION CONDITION (REFERENCE) Let $[\phi]$ be the type of ϕ and $([\phi])$ a token of the type $[\phi]$. Then $\varrho([\phi]) = \varrho([\varrho([\phi])])$.

[4] In a response to a paper by Mark Lance, Brandom can be read as already hinting at this idea. But even if so, what he explicitly says about this is far from clear (compare Brandom 1997).

Using 'refers' as the paradigm of the reference operator ϱ and stipulating that this alone suffices to pick out some previous utterance uniquely, the criterion states:

(17) the one referred to as 'ϕ' = the one referred to as 'the one referred to as 'ϕ"

According to this definition, expressions like 'the one denoted by 'ϕ" also qualify as instances of the reference operator. But expressions like 'the one praised as 'ϕ" or 'the one insulted by saying 'ϕ" are excluded. For example, one may pick up the meaning of a tokening of the expression 'this important philosopher' by saying 'the one praised as 'this important philosopher". But saying 'the one praised as 'the one praised as 'this important philosopher"' might well fail to continue this anaphoric chain. Reporting and expressing praise don't coincide.

My proposal for an amendment of prosententialism, the anaphoric account of truth, is strictly analogous. I propose to treat an expression as an instance of the truth operator ϑ just in case it satisfies the following criterion solely in virtue of its meaning.

ITERATION CONDITION (TRUTH) Let $[\phi]$ be the type of ϕ and $([\phi])$ a token of the type $[\phi]$. Then $\vartheta([\phi])$ if and only if $\vartheta([\vartheta([\phi])])$.

Using 'true' as the paradigm of the truth operator ϑ, and stipulating that this alone suffices to pick out some previous utterance uniquely, this second criterion states:

(18) 'ϕ' is true if and only if "ϕ' is true' is true.

On this definition, expressions such as 'It is the case that ϕ' or, arguably, "ϕ' is a fact' will turn out to be instances of the truth operator. Other expressions, by contrast, are excluded, e.g., 'It is a pity that ϕ'. For it might well be a pity that my favorite team lost since victory would have been a great reward for a season of hard work. But that doesn't mean that it is a pity that it is a pity that they lost, since their defeat's being a pity doesn't itself have any bad consequences.

III.3 The opposition to predicates

How do these criteria underwrite the claim that truth and reference talk are not predicative?

For a start, the iteration conditions would have to be reinterpreted to fit predicative cases. A truth predicate T, for example, would be defined as holding of any x just in case it also holds of $T(x)$. And a relational reference predicate R would be defined as holding between any x and y just in case it also holds between x and the unique z which stands in R to y. However, even if this can be somehow made to work, the anaphoric theory rules out such an interpretation. Let me show this for the truth operator and keep an analogous argument about the reference operator implicit.

Take three sentences like the following.

(19) Dogs bark.

(20) 'Dogs bark' is true.

(21) "Dogs bark' is true' is true.

On the anaphoric theory, (20) inherits its meaning from (19) and (21) from (20). Thus, they all have the very same meaning. By contrast, if truth talk is understood as predicative, (19) is about dogs, (20) about a sentence about dogs and (21) about a sentence about a sentence about dogs. Thus, there is a difference in meaning after all, a difference one may describe in terms of three levels of semantic ascent, from non-semantic in (19) to semantic in (20) to meta-semantic in (21). On the anaphoric theory, this ascent is an illusion. The semantic 'ladder' is horizontal. Thus, the iteration condition implies that truth talk cannot be predicative.

This last point brings out that the above iteration conditions, together with the rest of the anaphoric theory, don't merely have to be fulfilled solely in virtue of the meanings of the candidate expressions. After all, the expression 'is a quotable expression' arguably also fulfills both of them in virtue of its meaning.[5] Rather, the *resultant* expressions, too, have to share their meanings, as their respective roles in the anaphoric chain ensures.

[5] I am indebted to Miguel Hoeltje for confronting me with this example.

This feature is often criticized: Couldn't somebody accept that 'Dogs bark' is true without knowing the meaning of 'Dogs bark'? How, then, are these supposed to share their meanings? But it is entirely possible to accept a prosentence without being able to identify its antecedent. This is even one of the expressive advantages of the truth operator: blind endorsements like (8) on page 55.

IV Substantivism and theories of meaning

I have argued that proponents of the anaphoric theory can coherently shrug their shoulders at the substantive questions what truth and reference consist in. However, what does this mean for the *answers* actually proposed by substantivists and for explanations of meaning in terms of truth and reference?

IV.1 Substantivism

I take it that substantial accounts in general can be characterized by the claim that there is some substantial property— T, say— had by everything and only what is true, and that whatever is referred to by some expression stands is some substantial relation to it— say, R. Thus, the position can be expressed as follows:

SUBSTANTIVISM 'ϕ' is true just in case 'ϕ' has the substantial property T.

'ϕ' refers to ψ just in case 'ϕ' stands in the substantial relation R to ψ.

The first and crucial thing to note about Substantivism is its logical independence from the anaphoric theory. Recall that the anaphoric theory validates the following schemes:

SEMANTIC SCHEMES 'ϕ' is true just in case ϕ.

'ϕ' refers to ψ just in case ϕ is ψ.

On the anaphoric theory, these schemes are true in virtue of the meanings of the expressions formed with 'true' and 'refers'. For example, 'ϕ' is true just in case ϕ because "ϕ' is true' means that ϕ. Thus, the anaphoric theory implies that Substantivism means Substantivism*:

SUBSTANTIVISM* ϕ just in case 'ϕ' has the substantial property T.

ϕ is ψ just in case 'ϕ' stands in the substantial relation R to ψ.

However, Substantivism* is a claim without any use of truth and reference talk. Thus, it should be clear that it is logically independent from the anaphoric theory.

Mark Lance (1997) has already discovered half of this result, the part concerning truth talk only. He concluded that

> the anaphoric theory is not incompatible with any currently popular account of truth, charitably understood ... [N]o one has yet succeeded even in offering a theory incompatible with the anaphoric account, much less a refutation (Lance, 1997, pp. 283, 297).

However, Lance is too quick with the expression 'charitably understood'. If substantive accounts should be understood as maximally plausible from the point of view of anaphoric deflationism, he is certainly right. But we should also seek a charitable reading of the substantivists' ideas which takes their self-description as actually contradicting deflationism seriously.

What Lance fails to see is that Substantivism turned out to be compatible with anaphoric deflationism because it was stated as a purely extensional claim about expressions like "ϕ' is true' and "ϕ' refers to ψ' rather than about their meanings, or intensions. Full-blown substantivists hold that truth and T on the one hand and reference and R on the other don't merely *coincide*, but that truth *is* T and reference *is* R. In contrast with the above characterization of Substantivism, this idea can be expressed as follows:

FULL-BLOWN SUBSTANTIVISM "ϕ' is true' means that 'ϕ' has the substantial property T.

"ϕ' refers to ψ' means that 'ϕ' stands in the substantial relation R to ψ.

This, however, collides with anaphoric deflationism. The iteration conditions employed to define the truth and reference operators clearly show that "ϕ' has the substantial property T' is not an instance of the truth operator, and that 'the one standing in the substantial relation R to ψ' is not an instance of the reference operator. Even if they

were to fulfill thir respective iteration conditions, the resultant expressions would fail to share their meaning.[6] Thus, these expressions are predicates, the anaphoric operators are not.

Thus, there is a clear disagreement between anaphoric deflationism and substantivism, after all: However the property T and the relation R are understood exactly, substantivists claim and anaphoric deflationists deny that 'true', 'refers' and cognates mean T and R, respectively.

However, this is indeed all there is to the disagreement. In order to reject Full-blown Substantivism, the anaphoric deflationist needs additional arguments against T and R. And in order to reject anaphoric deflationism after Full-blown Substantivism has already fallen, the substantivist needs additional arguments against its functional anaphoric analysis of truth and reference talk. This is where their debate should continue.

IV.2 Meaning

Let me now discuss the implications of these insights for theories of meaning. Brandom thinks that the anaphoric theory makes it impossible to explain meaning in terms of truth or reference. He tries to establish this claim with what I call his Argument from Anaphora:[7]

1. If the anaphoric theory is true, then truth and reference are explained in terms of anaphora.
2. Anaphora can only be explained in terms of meaning, namely in terms of inheritance of meaning.
3. For all concepts, if α is explained in terms of β, and β can only be explained in terms of γ,then γ cannot be explained in terms of α.

[6] Colin McGinn (2002) takes up this dialectical position about truth when he claims that the disquotational nature of truth fails to support deflationism, but reveals its nature as "a device of ontological leapfrog". He even *defines* the substantial property of truth by something like the iteration condition.

[7] Compare Brandom 1994 and 2002. The literature on deflationism and meaning is full of cognate arguments. Their shared structure can be made explicit by substituting, say, 'disquotation' for 'anaphora' in the argument presented here.

4. Therefore, if the anaphoric theory is true, then meaning cannot be explained in terms of truth or reference.

However, this argument is far from a refutation of the idea that substantive accounts of truth and reference can be employed to explain meaning. True, the anaphoric theory precludes truth and reference from playing any explanatory role, but it is entirely neutral on what substantivists claim truth and reference *consist in*. If the anaphoric theory is correct, the substantivist has to sacrifice Full-blown Substantivism, but nothing more. When claiming that truth and reference explain meaning, one can simply *give away the words* 'truth' and 'reference', but maintain that *the substantial properties* T *and* R explain meaning.

Let me spell this out for the referentialist family of substantivist proposals, which Brandom is at most pains to reject. Referentialism, I take it, is the conjunction of a correspondence theory of truth and a causal-historical theory of reference. Thus, referentialists replace 'T' in the above schemes with, say, 'corresponds to a truth-maker' and 'R' with 'is causal-historically connected to'.

Crucially, whenever a referentialist appeals to what she calls 'truth' and 'reference', there will be a clear sense in which she actually appeals to correspondence and causal-historical connectedness.

Traditionally, the situation is this: Truth and reference are supposed to explain meaning, truth is correspondence, and reference is causal-historical connectedness. But explanatory relations are transitive. Thus, at the ground level, it is correspondence and causal-historical connectedness which explain meaning.

By contrast, if the referentialist buys into the anaphoric theory, drops Full-blown Substantivism and contents herself with Substantivism*, this argument becomes superfluous. Rather, the claim that truth and reference explain meaning becomes a potentially misleading shorthand for saying that, *from the outset*, the actual explanans is correspondence and causal-historical connectedness.

Thus, Brandom cannot invoke the anaphoric theory in order to argue against referentialist or other substantivist explanations of meaning. And likewise, one cannot invoke such an explanation of meaning in order to reject the anaphoric theory (like Schantz 2002 and many

others). The anaphoric theory is logically independent from theories of meaning such as referentialism and inferentialism.

V Conclusion

I have argued, first, that the anaphoric theory can account for its most radically deflationary claim that 'truth' and 'reference' are operators as opposed to predicates. Second, I have shown how substantivists can accept the anaphoric theory if they are willing to give away the words 'truth' and 'reference' and content themselves with the claim that truth and reference coincide with certain substantial properties. Thus, the explanatory value of those substantial properties is untouched by anaphoric deflationism and the debate over the explanation of meaning remains undecided.

However, these results shouldn't be taken to suggest that the debate between deflationism and substantivism is over. Instead, we should continute to discuss the functional analyses of truth and reference talk proposed by deflationists. As for anaphoric deflationism, we are now in a position to assess its distinctive and controversial characteristic, the claim that 'truth' and 'reference' are operators as opposed to predicates. Further, we should discuss the plausibility and explanatory power of Full-blown Substantivism in comparison with its more modest cousin Substantivism*. But, one way or the other, theories of meaning are independent from the outcome of this debate.[8]

References

Armour-Garb, B. and Beall, J. C. (2005). Deflationism: The Basics. In *Deflationary Truth*, pages 1–29.

Brandom, R. (1984). Reference Explained Away. Anaphoric Reference and Indirect Description. In *Deflationary Truth*, pages 258–281. Open Court. Originally published in: The Journal of Philosophy 81, 469–492.

[8]For insightful commentaries and helpful criticism, I am grateful to Christian Voigt, Eugen Pissarskoi, Sebastian Bänker, and to the participants of the conference in Amsterdam—especially to David Hunter, who commented on my talk.

Brandom, R. (1988). Pragmatism, Phenomenalism, and Truth Talk. *Realism and Antirealism. Midwest Studies in Philosophy*, 12:75–93.

Brandom, R. (1994). *Making It Explicit. Reason, Representing, and Discursive Commitment*. Harvard University Press, Cambridge.

Brandom, R. (1997). Replies to Commentators. *Truth. Philosophical Issues*, 8:199–214.

Brandom, R. (2002). Expressive versus Explanatory Deflationism about Truth. In *Deflationary Truth*, pages 237–257. Originally published in: Schantz (2002b), 103–119.

Devitt, M. (2002). The Metaphysics of Deflationary Truth. In *What is Truth?*, pages 60–78. Walter de Gruyter.

Grover, D. (1977). Inheritors and Paradox. *The Journal of Philosophy*, 74:590–604.

Grover, D., Camp, J., and Belnap, N. (1975). A Prosentential Theory of Truth. *Philosophical Studies*, 27:73–125.

Lance, M. (1997). The Significance of Anaphoric Theories of Truth and Reference. In *Deflationary Truth*, pages 282–297. Originally published in: Truth. Philosophical Issues, 8, 180–198.

McGinn, C. (2002). The truth about truth. In *What is Truth?*, pages 194–204. Walter de Gruyter.

Schantz, R. (2002). Truth, Meaning, and Reference. In *What is Truth?*, pages 79–99. Walter de Gruyter.

Achourioti, T., Andrade, E.J., Staudacher, M. (eds.), Proceedings of
the Amsterdam Graduate Philosophy Conference — Meaning and Truth,
ILLC Publications X-2010-1, pages 67–77

Do We Still Need Compositionality in a Contextualist Framework? Some Remarks on Recanati's Contextualism

Carlos Márquez and David Rey
(Universidad Nacional de Colombia)
cmmarquezs@gmail.com and dareys@gmail.com

I Compositionality and Contextualism

Forty years have passed since the publication of "Truth and Meaning" (1967). In this classic essay Davidson argued that a recursive theory of truth is the proper format for a compositional theory of meaning. The requirement of compositionality led Davidson to see the concept of truth as the foundation of meaning theories. But why is a theory of truth an appropriate instrument to reveal the compositional structure of a language? The main virtue of such a theory lies in its capacity to exploit the logical forms of sentences in order to derive truth conditions, and truth-values, from the interpretations of sentence constituents. Davidson was captivated by this trait of Tarski-style theories of truth. It is then not surprising that the study of logical form in natural language has played a key role in the development of the so-called Davidsonian program in semantics. Beginning with Frege, the semanticist tradition in the philosophy of language has established a deep connection between the notions of truth and meaning by means of compositionality. Philosophers of this tradition have regarded compositionality as a basic test for any satisfactory theory of meaning; a test so demanding that truth-conditional theories are the only ones able to pass it.

However, in the last decades the compositional picture depicted in "Truth and Meaning" has been challenged by a group of pragmatic oriented philosophers interested in the context-sensitivity of language. Roughly, members of this tradition hold that truth conditions cannot be expressed without the contribution of certain pragmatic factors that go beyond those recognized by semanticist theories. A sentence can be used to express different truth conditions in different contexts of use, and this is not merely due to the fact that it contains indexicals, or other linguistically governed slots, but rather to the fact that various pragmatic processes, besides those required for semantic completion, are involved in the specification of the truth conditions. For brevity, let us call this view 'contextualism'. *Prima facie* it seems that contextualism is a threat to the compositionality of truth-conditional content. If some pragmatically relevant contextual factors are required to express truth conditions in uttering a sentence, those truth conditions cannot be determined just by the logical form (or the syntactic structure) of the sentence plus the conventional meanings of its constituents. The indispensability of such contextual factors suggests that we need more than lexical meanings and linguistic structure in order to ascribe truth conditions to a truth-vehicle. Some prominent contextualist philosophers have said things that point in this direction. Consider what Charles Travis says about his well-known example of Pia's Japanese maple:

> A story. Pia's Japanese maple is full of russet leaves. Believing that green is the colour of leaves, she paints them. Returning, she reports, 'That's better. The leaves are green now.' She speaks truth. A botanist friend then phones, seeking green leaves for a study of green-leaf chemistry. 'The leaves (on my tree) are green,' Pia says. 'You can have those.' But now Pia speaks falsehood.
>
> If the story is right, then there are two distinguishable things to be said in speaking ['the leaves are green'] with the stipulated semantics. One is true; one false; so each would be true under different conditions. That semantics is, then, compatible with semantic variety, and with variety in truth-involving properties. So what the words of ['the leaves are green'] mean is compatible with various distinct conditions for its truth (Travis, 2008, pp. 111–112).

Here, according to Travis, two conversational situations determine different ways to assess when the leaves of Pia's tree count as

green, but the meanings of the words in the sentence remain constant in both situations. The same meanings, the same logical form, and different truth conditions expressed. John Searle comes to a similar result in his discussions about the concept of background.

> The simplest way to see that representation presupposes a nonrepresentational Background of capacities is to examine the understanding of sentences. [...] The same literal meaning will determine different conditions of satisfaction, for example, different truth conditions, relative to different Background presuppositions, and some literal meanings will determine no truth conditions because of the absence of appropriate Background presuppositions (Searle, 1992, p. 178).

The suggestion here is that literal meanings determine truth conditions only against a background of nonintentional capacities and social practices. Recall Searle's remarks about the verb 'cut'. We rely on our current social practices to understand the satisfaction conditions of an imperative utterance like 'cut the grass'. These background practices allow us to specify the particular manner of cutting the grass that is relevant to obey the imperative. If the background were different, for example, if we cut the grass in the way we cut cakes, the satisfaction conditions would be different, and without some background there would be no satisfaction conditions. Such modifications of background, according to Searle, do not require a change in the meanings of the words that compose the sentence, or a change in its linguistic structure. So, the truth conditions (propositional contents) that determine the alluded satisfaction conditions can vary while the meanings of the component words remain unchanged.

In a similar spirit, François Recanati has held that we arrive from literal meanings at expressed truth conditions only trough a kind of pragmatic process called 'modulation'. Modulation is the creation of new meanings from encoded linguistic material by means of optional processes of interpretation. Modulated meanings fit to the conversational context making interpretation possible. In his book *Literal Meaning*, Recanati argues that the phenomenon of modulation undermines traditional truth-conditional theories. According to Recanati, modulation of literal meanings takes place at the sub-

sentential level, and for this reason the literal truth conditions that semanticist theories compose do not play any role in communication. A compositional theory that derives these literal truth conditions, also called minimal propositions, is simply useless.

In short, current contextualism threatens the truth-conditional conception of compositionality that we inherited from the semanticist tradition. But what is the fate of compositionality in the hands of contextualists? Can we preserve the idea of compositionality in a contextualist philosophy? Recanati (2009) has suggested that, despite the appearances, the pragmatic influences of context upon truth conditions do not conflict with compositionality. In order to sustain this suggestion, he adopts a two-step model of semantic interpretation, in which primitive expressions and syntactic operations are first interpreted in the standard semanticist way and then the resulting semantic values are modulated by means of optional pragmatic processes. This two-step model was formerly introduced by Peter Pagin and Jeff Pelletier (2007), as an attempt to develop a moderate version of contextualism that makes compositional semantics compatible with the contextualist framework of *Literal Meaning*. In the rest of this paper we aim to criticize the two-step model. In section 2 we will present Recanati's version of the model. In section 3 we will put forward our criticisms to this model, exploring briefly the consequences of its rejection.

II The two-step model of semantic interpretation

Following the semanticist orthodoxy, Recanati characterizes compositionality in terms of two types of rules: (1) lexical rules to interpret simple expressions and (2) compositional rules to interpret complex expressions recursively. In a Kaplanian framework, these rules are adapted to context-sensitive expressions in the following way: (1) If α is a simple context-sensitive expression, the associated lexical rule establish that the interpretation I of α in a context c is the value $f(c)$ of a function f that maps contexts into semantic values. If the simple expression is context-insensitive, f delivers a constant semantic value for all contexts. (2) If $*(\alpha_1, \ldots, \alpha_n)$ is a complex expression, governed by a syntactic operation $*$ applied to

the constituent expressions $\alpha_1, \ldots, \alpha_n$, the associated compositional rule establishes that the interpretation I of $*(\alpha_1, \ldots, \alpha_n)$ is the value $f^*(I(\alpha_1), \ldots, I(\alpha_n))$ of a function f^* that maps sequences of semantic values into semantic values. In short:

1. If α is a context-sensitive simple expression: $I(\alpha)_c = f(c)$. If α is a context-insensitive simple expression, then $f(c) = m$ for all contexts c; so that $I(\alpha)_c = f(c) = m$.

2. If $*(\alpha_1, \ldots, \alpha_n)$ is a complex expression, then $I(*(\alpha_1, \ldots, \alpha_n)) = f^*(I(\alpha_1), \ldots, I(\alpha_n))$

In this theory, the content of a complex expression is a function of the contents of its constituents and its mode of combination, and so standard compositionality is satisfied. But, at the same time, the content of a simple expression is relative to context. Besides, the theory is restricted to expressions that are context-sensitive in the way in which indexicals are context-sensitive. The theory only accounts for the context-sensitivity of expressions that, in virtue linguistic conventions, are in need of saturation. However, Recanati claims that there are contextual influences on truth-conditional content that go beyond saturation:

> Consider, as an analogy, the Rumelhart example I discuss in Literal Meaning:
>
> (9) The policeman stopped the car
>
> We naturally interpret this as meaning that the policeman stopped the car by addressing an appropriate signal to the driver, just as we naturally interpret 'John cut the cake' as meaning that John sliced it. As Rumelhart points out, however, a different interpretation emerges if we imagine a context in which the policeman is the driver of the car: such a context provides for a totally different 'manner of stopping the car' on the policeman's part [...] Transitive 'stop' means CAUSE TO STOP, and this can be fleshed out in all sorts of ways, yet the fleshing out process is different from the saturation process mandated by indexicals and other expressions whose standing meaning is gappy and requires contextual completion. Indeed we can construct a context in which (9) would mean that the policeman stopped the car in some way or other, indifferently. No such opinion exists for indexicals or under-specified expressions, which do not carry a definite content unless the free variable is assigned a definite value in context (Recanati, 2009, p. 19).

A semanticist might feel tempted to argue that the transitive uses of 'cut' and 'stop' introduce hidden variables ranging over manners of cutting and manners of stopping. Similarly, he might argue that 'green' introduces a variable for ways of count as green. If this analysis were correct, all of these words would be interpreted by means of saturation. Generalizing this line of thought, the semanticist could try to reduce every influence of context on truth conditions to saturation.[1] In contrast, Recanati asserts that many forms of context-sensitivity, including the use of transitive verbs, call for a different mechanism, namely modulation. In interpreting an ordinary utterance of 'The policeman stopped the car' we enrich appropriately the meaning of the verb 'stop', obtaining the manner of stopping that is contextually salient. There is no saturation involved. So, we need to distinguish the mandatory processes of saturation through which indexicals and free variables in logical form are assigned a contextual value from the optional processes of modulation through which an expression's semantic value is mapped into a distinct semantic value. Recanati suggests that we can make room for modulation by defining a function M of modulated-interpretation. The compositional and lexical rules are then reformulated as follows:

1. If e is a simple expression, $M(e) = g_i(I(e)_c)$, where g_i is some pragmatic function determined as relevant in the context c.

2. If $*(e_i, \ldots, e_n)$ is a complex expression, then $M(*(e_i, \ldots, e_n)) = g_i(f^*(M(e_1), \ldots, M(e_n)))$ for some pragmatic function g_i.

In this formulation, the I-interpretation of a simple expression e is subject to a modulation process represented by the function g_i. This pragmatic function delivers a modulated semantic value for e. For a complex expression $*(e_i, \ldots, e_n)$, the interpretation function f^* takes a sequence constituted by the modulated semantic values of $e_i, \ldots, e_n$ and delivers a semantic value. Finally, this semantic value is subject to a modulation given by the function g_i. The result is a two-step model in which each expression is interpreted by a traditional interpretation function I and then the output of this

[1] For a general defense of this strategy, see Stanley (2007).

function serves as input for the modulation function M. As an illustration, consider again Searle's example. In the sentential context 'cut the grass', we interpret the verb 'cut' as we do because the syntactic complement 'the grass' makes salient a particular manner of cutting given a relevant background of social practices. The linguistic context and the extra-linguistic background specify a function g_i that takes as input the literal semantic value of 'cut' and delivers as output the contextually salient manner of cutting. Thus, the pragmatic functions g are sensitive to the influences of both linguistic and extra-linguistic context.

III Logical forms and literal semantic values

Traditionally, semantic interpretation has been seen as a unidirectional bottom-up process, going from the interpretations of the parts to the interpretation of the whole. The meaning of the whole is determined by the meanings of its parts, but the meaning of a part cannot be determined by its linguistic surroundings. This traditional view excludes the possibility of whole-to-part (top-down) and part-to-part (lateral) semantic influences. Recanati emphasizes that his model has the virtue of admitting this kind of semantic flexibility while at the same time preserving the compositionality principle. In this model, although the modulated meanings of constituent expressions can be affected by linguistic and extra-linguistic influences, the modulated meaning of a complex expression is a function of its modulated semantic structure and the modulated meanings of its constituents. Now, since compositionality is here located at the level of modulated meanings, one might wonder why it is necesarry to posit a level of unmodulated semantic interpretation.

In our view, Recanati's *Literal Meaning* displays an inconsistent attitude toward literal semantic values. While Recanati strongly criticizes semantic approaches that make use of minimal propositions, there is no analogous complaint against minimal semantic values of simple expressions. The assumption that simple expressions, in contrast to sentences, have unmodulated semantic values runs through the whole book. Even the use of the term 'modulation' is symptomatic of this assumption, because minimal semantic values are the

things that are supposed to be modulated in modulation processes. The assumption is an integral part of the two-step model sketched in the previous section, where the pragmatic functions g, applied to simple expressions, take as inputs the unmodulated contents delivered by the function I. Can we consistently reject literal semantic values at the sentential level and accept them at the sub-sentential level? This alternative, we want to argue, goes against the spirit of contextualism. From a contextualist point of view, it does not make sense to say that words have a definite semantic value independently of the context of utterance. To see why let us focus on the following passage of *Literal Meaning*:

> Or consider the paradigmatic example I gave in the last section: the adjective 'red'. There is no particular incompleteness in the linguistic meaning of that word—it means something like 'being of the colour red' or 'having the colour red'. Insofar as 'red' refers to a specific colour (and it does) this, it seems, expresses a definite property—a property that could, in principle, go into the interpretation of a sentence in which the adjective 'red' occurs. (For example: 'Imagine a red surface.') But in most cases the following question will arise: what is it for the thing talked about to count as having that colour? Unless that question is answered, the utterance ascribing redness to the thing talked about (John's car, say) will not be truth-evaluable (Recanati, 2004, pp. 138–139).

In emphasizing the indispensability of modulation Recanati says that the adjective 'red' expresses a definite property. Such a property is supposed be the literal semantic value that the function I assigns to 'red'. The crucial question, however, is: what property is this? If there is a property associated with 'red', there must be a—probably vague—set of objects that are the instances of that property. But which objects are these? Let us return to the case of Pia's Japanese maple. Are the leaves of Pia's tree instances of the literal property assigned to 'green'? If we answer affirmatively, the understanding of 'green' in which painted leaves count as green would receive a privileged status. If we answer negatively, the understanding of 'green' in which painted leaves do not count as green by the mere fact that they were painted green would receive a privileged status. There is no reason to privilege one understanding over the other. Since the question of what count as being green does not have a definite answer prior

to the context of utterance, as Recanati himself acknowledges in the quoted passage, there cannot be a fixed set of objects that are the instances of the context-independent property of being green, and there cannot be a context-independent function that maps green objects into truth values, or a context-independent satisfaction relation that connects sequences of green objects with the open sentence 'x is green' (see Travis, 2008, p. 121). If Recanati is right in thinking that we require modulation in order to specify relevant ways of being red, he must be wrong in thinking that 'red' has a literal semantic value. It is not difficult to conceive of new cases in which modulation is required to determine which actions count as cuttings, or as stoppings, or as hangings, or which objects count as red, or as philosophers, or as Republican senators. For a contextualist this is not merely a matter of vagueness. In different contexts of use modulation give us different criteria to apply a given word, and usually there is no way to privilege one criterion over the others. Contextualism leads us to the conclusion that there are no metaphysical facts of the matter about context-independent semantic values.

At this point, it seems that we have arrived at a very radical conclusion: if contextualism is right, and truth conditions are expressed only in contexts of use, then sub-sentential expressions cannot have unmodulated semantic values. The semantic values of primitive words vary from one context to another, and there is no neutral literal value that serves as an input to the pragmatic processes of modulation. Of course, if simple expressions do not have context-independent semantic values, it is not possible to apply a Tarskian theory of truth to a language. These theories require a satisfaction relation that connects linguistic expressions with domain-objects, so that we can state the satisfaction conditions of such expressions. However, our previous conclusion was that semantic relations like satisfaction and reference could not be established in a context-independent way.

Given that there is no possibility to apply standard truth-theories, must we reject formal semantics as a legitimate enterprise? We don't think so. As we remarked in the beginning of this paper, the core idea behind the Davidsonian program lies in the notion of logical form. Even if we have semantic values assigned to sub-sentential

expressions only in pragmatic contexts, we can still study the way in which those semantic values give rise to truth-conditional structures. In order to undertake this study, we need no more than logical forms, because logical forms reveal the kinds of semantic contributions corresponding to expressions of different types. It seems to us that most valuable work in formal semantics can be interpreted as work on logical forms. A semantic theory of logical form is needed if we want to understand how truth-conditional contents emerge from modulated meanings. But if we assign logical forms to sentences—or to sentences in context—, we are forced to recognize a truth-conditional compositional structure. Because we cannot dispense with a theory of logical form, it is necessary to preserve a notion of truth-conditional compositionality.

References

Davidson, D. (1967). Truth and meaning. *Synthese*, 17. Reprinted in Davidson (1984):17–36.

Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford.

Pagin, P. and Pelletier, J. (2007). Content, context and composition. In Peter, G. and Preyer, G., editors, *Content and Context. Essays on Semantics and Pragmatics*, pages 25–62. Oxford University Press, Oxford.

Recanati, F. (2004). *Literal Meaning*. Cambridge University Press, Cambridge.

Recanati, F. (2009). Compositionality, flexibility, and context-dependence. In Hinzen, W., Machery, E., and Werning, M., editors, *The Oxford Handbook of Compositionality*.

Searle, J. (1992). *The Rediscovery of the Mind*. MIT Press, Cambridge Mass. and London.

Stanley, J. (2007). *Language in Context*. Claredon Press, Oxford.

Travis, C. (2008). *Occasion-Sensitivity. Selected Essays*, chapter Pragmatics, pages 109–129. Oxford University Press, Oxford.

Achourioti, T., Andrade, E.J., Staudacher, M. (eds.), Proceedings of the Amsterdam Graduate Philosophy Conference — Meaning and Truth, ILLC Publications X-2010-1, pages 78–103

Autism as a disproof of Grice on Meaning

Meredith Plug (King's College London)

meredith.plug@kcl.ac.uk

Glüer and Pagin (2003) argue that the existence of a subset of speakers on the autistic spectrum, a group of speakers that have a sophisticated level of language but have problems with attributing beliefs and thoughts to other people and themselves, present a counterexample to Grice's analysis[1] of "nonnatural meaning" in terms of higher order thoughts (e.g., beliefs about beliefs).[2] However, Grice takes this type of meaning to be related to communication, and it has been argued (Reboul, 2006) that the communicative problems that this group of speakers have indicates that these speakers can't exemplify this particular type of meaning, and that therefore these speakers do not pose a counterexample to his analysis. I argue that the way Reboul mounts her attack rests on a misunderstanding of Glüer and Pagin's argument (and that it is therefore not successful). Nevertheless, I believe that Glüer and Pagin do not make a sufficient case that the meaning of autistic speakers is a matter of nonnatural meaning, and so their argument, as it stands, is left open to attacks of a similar kind. To ensure the claim that (a subset of) autistic speakers are a counterexample to Grice's analysis an assessment is required of the intuitions we have concerning the applicability of the

[1] See for example Grice (1989a).

[2] Glüer and Pagin (2003, pp. 25f) note that Stephen Laurence made this suggestion before they did, albeit without working it out at length, in his Laurence (1998, pp. 209f).

term "meaning" in different circumstances (the intuitions that Grice refers to in his analysing process), and additional data on the grammatical ability of the subset of autistic speakers who are not (yet) able to attribute beliefs. I believe that with this addition, Glüer and Pagin's argument still stands.

Autistic spectrum disorder, which incorporates both Autism and Asperger Syndrome,[3] is a neuro-developmental disorder,[4] currently diagnosed on the basis of selective qualitative impairments in the areas of communicative, social and imaginative abilities, accompanied by patterns of restricted interests and repetitive or stereotyped behaviour.[5] The spectrum character of the disorder consists in the fact that expression of these impairments (and also further symptoms not currently incorporated in the diagnosis) varies significantly between individuals and also within individuals over time. Here we focus on linguistic ability and the ability to attribute representational mental states such as beliefs.

About twenty percent of children on the autistic spectrum do not develop functional language.[6] At the other end of the spectrum there are children whose linguistic development may be delayed,[7] but follows roughly the same path as the linguistic development of typically developing children, when we concentrate on formal[8] or

[3]Asperger Syndrome will be classed as part of the Autistic Spectrum in the forthcoming Diagnostic and Statistical Manual of Mental disorders (DSM-V) of the American Psychiatric Organisation, although this is not the classification the current Handbook adheres to (the DSM-IV: APA, 2000).

[4]As evidenced by brain dysfunction and structural and functional differences in regions of the brain (Surian et al., 1996).

[5]According to DSM-IV: (APA, 2000).

[6]Tager-Flusberg et al. (2005). In the past, the proportion of verbal versus nonverbal individuals on the autistic spectrum was thought to be 50/50 (Bryson et al., 1988).

[7]Or not—children with Asperger's do not have a specific language delay (APA, 2000).

[8]In particular: children on the autistic spectrum that develop functional language show the same learning curve in syntactic development as typically developing children (the same range, and the same developmental ordering of grammatical structures). Generally sentences are grammatically intact. There are also no specific problems with mastering questions, active passive transformations, negation, and clausal complement construction, for example. See Tager-Flusberg et al. (1990) for an overview. For a minority of autistic children, problems with

semantic[9] aspects of language. However, these High-Functioning Autistic or Asperger's children remain challenged in 'pragmatic' aspects of communication, even as adults:[10] They may have difficulty sustaining a conversation (as shown by abrupt terminations or shift in topics, and an inability to give and receive conversational cues, for example), they may have difficulty taking into account information that a conversational partner can be expected to have and not to have, are very literal minded, and may not understand jokes, irony and sarcasm, except through a conscious workaround.[11]

Both the pragmatic problems children on the autistic spectrum exhibit, and the impairments in the use[12] and understanding[13] of mental state terms such as "belief," "think," "know" and "pretend"

past tense constructions are found, but it is hypothesised that these children form a specific subgroup that have SLI ("Specific Language Impairment") co-morbid with their autism (Tager-Flusberg and Joseph, 2003).

[9]It is now thought that semantic organization is not significantly different from the semantic organization of typically developing children (Paul et al., 2007), although studies in this area have been few. For example; autistic children use semantic groupings in typical ways to categorize and retrieve words (Ungerer and Sigman, 1987; Lewis and Boucher, 1988). They are able to match a typical exemplar with a less typical exemplar (Tager-Flusberg, 1985), and they show the same pattern of overextension and underextension in relation to a prototype, as typically developing children do (Minshew et al., 2002). See also Toichi and Kamio (2001) for a more subtle priming study showing similar semantic relationships between words.

Vocabulary building also follows a roughly similar path as vocabulary building in typically developing children: autistic preschoolers imitate words in advance of naming things, their word comprehension is in advance of word production, and gesture production is used as a bridge between the two (Charman et al., 2003). Different word types, such as common nouns, predicates, and so on, are acquired in the same proportion to total vocabulary size as is the case with TD's (Charman et al., 2003). On the whole, the order of emergence of predicate categories in ASD children follows roughly the path as in typically developing children (Peralejo, 2008).

[10]For an overview of the research, see Tager-Flusberg et al. (2005).

[11]See in particular Happé (1993) and Surian et al. (1996) for autistic children's problems with certain kinds of pragmatic inferences.

[12]e.g., Tager-Flusberg (1992).

[13]e.g., the use of 'to know that such and such' without the understanding that this implies that such and such is true, or the use of "to believe that such and such," without understanding that this doesn't imply that such and such is true, or is false (Dennis et al., 2001).

(a notable exception to the fact that semantic development in autism compared to typically developing children is overall similar)[14] suggest that autistic children have problems with attributing mental states. These problems are well documented,[15] although the underlying reasons for the impairments are being debated, e.g., whether it is due to a conceptual deficit—no possession of the concept of belief or thought[16]—or mainly due to profound performance factors.[17] One of the most widely researched impairments is an impairment to attribute false beliefs. False Belief Tasks are designed to show whether a child is able to attribute a false belief to themselves or another person when this is demanded by context.[18] If a child is able to, she

[14]There are other exceptions, for example, the use of affective terms, problems with the use of deictic term (this,that), personal pronouns ("you want candy") and the use of articles: this can be shown alongside correct usage of pronouns). For an overview of the research, (Tager-Flusberg et al., 2005).

[15]To name just a few indications: autistic children show impairments in areas where belief attribution is arguably required, such as the understanding of deception. e.g., (Baron-Cohen, 1992) and surprise (Baron-Cohen et al., 1993), complex causes (such as beliefs) of emotion (Baron-Cohen, 1991) spontaneous pretend play (Lewis and Boucher, 1988) and the appearance/reality distinction (Baron-Cohen, 1989). There are also impairments in areas that are likely to be precursors to the ability to attribute beliefs, such as joint attention behaviours, (Charman et al., 2000), the understanding that something looks different from one viewpoint than from another (Hamilton et al., 2009), and the understanding that seeing implies knowing and not seeing 'ignorance' (Baron-Cohen and Goodhart, 1994).

[16]Some versions of the "Theory of Mind" Deficit Account of autistic spectrum disorders may posit this: they may for instance say that there is no "concept" because the module responsible for attributing beliefs has not yet matured or is faulty.

[17]Some versions of the Executive Function Deficit account of Autistic Spectrum Disorder may posit this: they may claim that there is (or there might) be a concept of belief, but that it can't be employed because it would require, for instance, too much control over one's attention span or ask too much of the memory system, or the calculations required to reason with it could be too difficult. (Note that executive function problems may also be thought to impair the formation of a concept of belief, depending on one views on concept development.)

[18]A typical False Belief Task involves presenting a situation to the child being tested where one of the protagonists is not privy to some information that the child is privy to. For instance, Sally leaves her marble in a certain location and then leaves the room. While she is away, Ann moves her marble from the first location to a second location. The child who witnesses these events is then asked, when Sally returns, where will she look for her marble? Success on the task

passes the task, and this is taken to be unambiguous proof that the child really is attributing a mental state, a belief, to the other person rather than just base her predictions of other people's behaviour on the basis of the state of the world or on the basis of her own belief about the world.[19] Failure to attribute false beliefs does not by itself indicate that mental states are not being attributed—theoretically it's possible that children might be failing the task because they mistakenly attribute a true belief;[20] but there is no evidence that might make this plausible as a general rule.[21] And so, when one interprets

depends on whether the child will predict the Sally's behaviour correctly—on the basis of the belief Sally has (that the marble is in the first location) which is different from the child's own belief, who has witnessed the moving of the marble and so knows that its in the second location instead. The classic False Belief Task was devised by Wimmer and Perner (1983), taking up a suggestion by Dennett (1978). See for the first variant of this task administered toautistic children (Baron-Cohen et al., 1985).

[19]Some may think that a prediction of other's behaviour based on the state of the world is also a kind of content attribution. However, such an attribution is done without an implicit or explicit understanding of the representational character of mental states. Grice's analysis of meaning requires not just that one attributes content, but that one intends to produce a belief in the audience (and even more than that): and this requires that one implicitly or explicitly understands the representational character of belief.

[20]My approach differs slightly from Glüer and Pagin, who in part seem to use a philosophical argument to establish the credibility of False Belief Task failure to delineate the subset of autistic speakers: "[...]the problem is *not* that there is only a problem with attributing *false* beliefs. The problem is that understanding what it is to have a false belief is an essential part of understanding what it is to have a belief at all." (Glüer and Pagin, 2003, p. 37) and: "to understand the difference between being true and being believed to be true, one must understand that a belief can be false, and this understanding is manifested by means of the ability to ascribe beliefs one takes to be false" (Glüer and Pagin, 2003, p. 27). This may be true, but it doesn't help to pinpoint an impairment in belief attribution (if that is what Glüer and Pagin mean here) because what is required is knowing when this understanding is not available (not when it is). False belief test failure in itself only suggests an impairment in explicit False Belief attribution, not an impairment in the understanding that beliefs are representational.

[21]At least, not until False Belief Tasks are almost passed. There is evidence that 'Diverse Belief' tasks (tasks that test for the understanding that different people may have different beliefs about the same situation at the same time—and therefore supposedly test for the understanding that beliefs are representational—without having to attribute a false belief) are passed just before passing False

False Belief Task failing from within the wider context just outlined above, one can assume that failing such tasks means (near enough)[22] that beliefs are not yet being attributed.[23, 24]

Glüer and Pagin use the data on False Belief Task performance to home in on a subset of autistic speakers who speak on a sophisticated level, and yet are not able to attribute mental states to other people. Such a group of speakers are likely to be a counterexample to Grice's analysis of meaning, given that Grice's analysis posits that in order to mean something, a speaker must intend to produce beliefs in their audience:

Grice's aim is to analyse our concept of "Meaning," in all its forms. In order to be capable of setting up an analysis, one must already be capable of applying the concept in particular cases. A

Belief tests in autistic children (Peterson et al., 2005), and one therefore has to be careful to interpret the transition from failing to passing False Belief Tasks as *exactly* pinpointing an understanding of the representational character of mental states. But there is no general evidence that would suggest an explicit or implicit understanding of the representational character of mental states in autistic children well before passing these tests. For instance, simplified False Belief Tasks are not passed before standard False Belief Tasks (Surian and Leslie, 1999), and nonverbal (or implicit) False Belief Tasks are not passed before standard False Belief Tasks. (Senju et al., 2009, ress). Also, False Belief Task failure should be interpreted in the context of the other evidence for impairments in the attribution of beliefs (see footnote 15), and lastly, there is specific evidence that spontaneous (untrained) passing of standard False Belief Task is correlated with enhanced pragmatic abilities (Eisenmajer and Prior, 1991), enhanced social skills in a naturalistic setting (Astington and Jenkins, 1995; Frith et al., 1994) and increased use and better understanding of mental state terms (Ziatas et al., 1998); although many impairments remain even after passing, on a more subtle level.

[22] With the caveat of footnote 21 kept in mind.

[23] I present a more thorough defence that False Belief Task failure can be so interpreted in my thesis (Plug, 2010).

[24] Autistic children are not alone in having problems with the attribution of false beliefs at a certain point in development. Typically developing children are unable to pass False Belief Tasks before the age of around 3.5. However, typically developing children are not likely to pose a problem for Grice's analysis of meaning because on the one hand, their language is not yet very accomplished (and is of a much lower level than the language of the autistic children considered in this paper). And on the other hand, it is far less clear in the case of typically developing children that False Belief Task failure is indicative of problems with the attribution of mental states such as belief.

conceptual analysis utilises one's intuitions concerning these cases to arrive at necessary and sufficient conditions for the concept to apply. Some of our intuitions regarding use are obvious. Grice seems to believe this is the case with a distinction between two ways in which we use the term "meaning" in our ordinary talk.[25] According to him it is reasonably clear, intuitively,[26] that we can distinguish our use of the term "meaning" in cases we might call "natural meaning;" a sense which is closely related to something being a natural sign for something else (e.g., "clouds mean rain"), from our uses of the term in cases we might call "nonnatural meaning," (e.g., " 'the cat is on the mat' means that the cat is on the mat")which is "related to communication."[27] It is the second type of meaning that he subjects to a more careful conceptual analysis.

I will discuss some of the intuitions regarding use that Grice reflects while analysing nonnatural meaning later. The *end result* of the analysing process, is that the most fundamental type of nonnatural meaning, Speaker Meaning, has the following necessary and sufficient conditions:

> For a speaker S, an utterance U, a response R and time t:
>
> S means something at t by uttering U, iff for some audience A, S uttered U intending
>
> 1. A to produce a particular response R.
> 2. A to think (recognize) that S intends 1.
> 3. A to fulfill 1. on the basis of the fulfillment of 2.[28]

In the case of the making of statements (rather than orderings for example)—and in our discussion we will limit ourselves to Grice's analysis of statements[29]—the response the speaker intends to produce in the audience must be a particular belief (the content of which

[25] See Grice (1989a, p. 214) and Grice (1989b, p. 291).
[26] Grice (1989b, p. 291).
[27] Ibid.
[28] This is roughly the formulation of Grice's initial analysis (Grice, 1989c, p. 92).
[29] Because the threat the existence of certain autistic speakers pose to Grice's program is easiest to present by focusing on his analysis of statements, and also

supplies the content of the statement).[30] Grice is aware that speakers seem to speak quite automatically, and are normally not aware that they possess the higher-order intentions (intentions about beliefs about intentions about beliefs) he proposes they possess. Therefore, Grice must take these intentions to be unconscious, tacit, or implicit. Nevertheless, speakers must have these intentions; the intentions must be *attributable* to the speaker.[31]

We have seen that autistic children who do not yet pass False Belief Tasks, are not able to attribute beliefs,[32] and therefore in their turn can't be attributed the intentions required: they can't intend to produce a particular belief in the audience on the basis of the audience's recognising that they intend to produce this belief in the audience.[33] Nevertheless, Glüer and Pagin argue that a subset of

because the analysis of statements is central to any analysis of meaningful speech. I set out the threat to his analysis of orderings etc. elsewhere (Plug, 2010). Such an argument follows similar lines, but must also refer to the difficulty a subset of autistic speakers have with second order intentions.

[30]According to Grice: "to ask what U meant is to ask for the intended effect" (Grice, 1989a, p. 220). In later writings, the intended effect is suggested to be a particular thought, an activated belief, a belief about a belief of the speaker, or an activated belief about the belief of the speaker, and so on (for an overview, see Neale, 1992, p. 38). There have been other variations on the formulation of the analysis (e.g., in later writings, the self-referential nature of the earlier formulation is replaced) but in all the proposed refinements at the core of the analysis remains the intended production of a thought or belief.

[31]There is some debate about what Grice believes the exact conditions to be under which one can attribute an intention to a speaker (or in other words, the conditions under which a speaker possesses the intention). One may for instance interpret Grice as supposing that it is not necessary that meaningful speakers must be able to form explicit intentions of the kind mentioned at the time at which something is meant, as long as it is possible to form these intentions at a later date and it is also recognised or in some exceptional cases "decided" (Grice, 1989a, p. 222) at that later date that these type of intentions were formed implicitly at the earlier date. If this reading of Grice is correct, additional empirical data on autistic children is required for Glüer and Pagin's argument to go through. I discuss this in my thesis (Plug, 2010).

[32]See footnote 19.

[33]The formation of such an intention also requires higher order thought: it requires thinking about the audience's thinking process, because Grice believes that the audience's recognition of the intention of the speaker is not just a cause for the audience but also a reason (Grice, 1989a, p. 221), and Grice believes one can't in general "form intentions to achieve results which one sees no chance of

these children who do not pass False Belief Tasks, have a "sophisticated"[34] linguistic ability. They cite a large metastudy by Francesca Happé that has shown that that some autistic children require the verbal age of a typically developing 9 to 11-year-old[35] in order to pass these tasks (this is of course with IQ and other possible influences partialled out).[36] Glüer and Pagin believe that it would be absurd to deny, on the basis of their linguistic level, that these speakers mean something[37] when they speak[38] and that they are communicators.[39] Are Glüer and Pagin correct to infer that autistic speakers mean something, from the data on linguistic development and ability that they put forward?

It is clear that autistic speakers mean something, in our ordinary sense of "meaning" (before reflecting on which types of meaning should be distinguished in our ordinary use). For instance, because vocabulary measures test for the correct usage of words,[40] it can

achieving" (Grice, 1989c, p. 98).

[34]Glüer and Pagin (2003, p. 47).

[35]To put it like this is slightly misleading, because the grammar and vocabulary measures are typed on large random groups that may also include some speakers with autism. But it is true near enough; these language measures have been checked extensively for reliability.

[36]Happé (1995).

[37]Glüer and Pagin phrase it differently: they say that it would be absurd to deny that these speakers are *language users*. Because Grice does not analyse what it is to be a language user, but what it is to mean something, Glüer and Pagin must take the ability to speak meaningfully (in the nonnatural sense, see below) to be entailed by being a language user. To avoid this issue I focus directly on whether, from the data presented, it can be concluded that autistic speakers *mean* something (in the nonnatural sense).

[38]In this paper I colloquially refer to autistic speakers "meaning something with their words," or meaning something "when they speak" but I do not want to suggest with this that Grice thought that speakers mean things only when they use words and sentences (that is, that Speaker Meaning is exhausted by linguistic meaning) or that Word Meaning is more fundamental than Speaker Meaning. It is only that in this paper, we happen to assess the fact that autistic speakers mean something on the basis of linguistic competence.

[39]Glüer and Pagin are aware that the claim that autistic speakers are communicators requires more argument than referencing linguistic ability. I will discuss their argument below.

[40]On the Peabody Picture Vocabulary Test - III (PPVT) (Dunn and Dunn, 1997) or its British equivalent, the British Picture Vocabulary Scale (BPVS)

hardly be maintained that these words as spoken by them have no content. But it is a different question whether this sense of meaning is the sense that Grice analyses. As we mentioned above Grice distinguishes two different uses of the term "meaning" in our day to day speech: "natural meaning" and "nonnatural meaning," which is "related to communication."[41] To make this distinction doesn't require a lot of reflection: Grice talks about it as a "reasonably clear intuitive distinction",[42] and he offers a couple of "recognition tests"[43] that allow one to see with what type of meaning one is dealing with, if in doubt. These are, firstly, that in the case of nonnatural meaning, unlike in the case of natural meaning, there is always someone or somebody meaning something. Secondly, that in the case of nonnatural meaning the 'something' that is meant is "nonfactive": what Grice means with this is that it would not be absurd to say that someone means something, but that it is not the case (or false). Lastly, what is meant can be paraphrased using quotation marks in the case of nonnatural meaning, but this can't be done with natural meaning.[44]

If some speakers would use their words in the right context, but not implicitly or explicitly understand that their words are representations, and that therefore sentences containing them could be either true, or false, it would not be unlikely that their "meaning something" would mainly consist in a correlation between the words and things in the world. If this was the case with our autistic speakers, it could be argued that they only mean something in the natural sense.[45] The data that Glüer and Pagin present, does not rule out

(Dunn et al., 1982), the two vocabulary tests referred to, lexical comprehension is tested by presenting an auditory word and asking the participant to point to or state the number of the corresponding picture (depicting objects, actions, etc.) from an array of four (numbered) pictures.

[41]Grice (1989b, p. 291).

[42]Ibid.

[43]Ibid. See also Grice (1989a, p. 214).

[44]There is one more recognitions test: In the case of natural meaning (and not nonnatural meaning), what is meant can be restated beginning with the phrase "the fact that... "

[45]An argument of that sort would have to deal with conflicting intuitions: for instance, the tension between our intuition that this is a case of "someone meaning

this possibility (the only studies cited that correlate verbal ability with False Belief Task performance are standard vocabulary measures, which only test object-"meaning" correlations)[46]—but Glüer and Pagin are aware that it must be ruled out, firstly, because they present vocabulary performance data in the context of other studies that suggest that linguistic development is overall comparable to the development of typically developing children (although delayed).[47] Secondly, they cite additional data suggesting that there are autistic speakers that understand negation and affirmation.[48] And thirdly: they cite additional data that vocabulary age in autistic speakers is in general correlated with grammar age[49]—in short, they aim to argue that when we are talking about a group of autistic children with a vocabulary age of 8, say, this group will on average have a general verbal age of 8 (and that this includes an understanding of negation, etc).

However, the problem with these three sets of data is that they concern the linguistic development of autistic speakers in general. The subset of autistic speakers we are concerned with, a set of sophisticated speakers who do not (yet) pass False Belief Tasks, could be a group of speakers with an *exceptional* linguistic development.[50] In particular, from the data it couldn't be ruled that despite their

something" and our intuition that words are close to signs for those who do not understand the representational character of words, suggesting that the meaning involved is natural.

[46]e.g., the large metastudy by Happé (1995) only investigates correlations between British Picture Vocabulary Scale level and False Belief Task performance.

[47]Glüer and Pagin (2003, p. 29).

[48]Glüer and Pagin (2003, p. 33).

[49]Ibid.

[50]Note that the subset of autistic speakers that is considered to be a counterexample to Grice's analysis is not a diagnostic subgroup in autism. This is why studies that investigate the correlations between grammar and verbal ability can't be expected to test this specific subset for alternative language profiles. For example, the study that Glüer and Pagin refer to (Jarrold et al., 1997) does not do so. Also, the subset is relatively small, and so a divergent language profile of language attainment in the subset would still be compatible with a relatively uniform profile of language attainment for autistic children in general. Lastly, the fact that linguistic development in general is similar, although delayed, does not help because we want to home in on a specific timeframe: linguistic ability before passing False Belief Tasks.

high vocabulary age, the speakers in our subset do not yet use negation and affirmation correctly, since precisely those children who do not yet pass False Belief Tasks could be deficient in this understanding. And so, there could be an exception to the rule that vocabulary age matches grammar age precisely in those children who have not yet passed False Belief Tasks.[51]

To rule out that the children in the subset mean something in the natural sense, it is necessary to present data on grammar development and ability in relation to False Belief Task performance. Such data exist: there are children with a grammar level of typically developing 6 to 9-year-old, that do not (yet) pass False Belief Tasks.[52] This level of grammar does not differ in fundamentals, from adult grammar,[53] and it means specifically that negation and embedded sentences have already been mastered (as these are in place in typically developing children from six years onwards).[54] Such a high level of grammar suggests that it is understood that words and sentences are representational, and that sentences can be true or false. Granted, if the only contexts in which the terms "true" and "false" were used correctly were simple contexts, this would not suggest representational understanding (because these words could be referring to other concepts than TRUTH and FALSITY). But sustained cor-

[51]There are children with autism with exceptional rote memory skills. Therefore, children who have excellent vocabulary age before passing False Belief Task (rather than the reverse) may have acquired a large vocabulary "associatively" by rote memory learning.

[52]See Tager-Flusberg and Sullivan (1994); Fisher et al. (2005). Note that because of the fact that Diverse Belief tasks are passed a little before False Belief Tasks (see footnote 21), it is best to go with a conservative choice in delineating the subset and considers autistic speakers that do not pass False Belief Tasks with a grammar age of 7, rather than 9, as a putative counterexample. As far as I am aware, there are no studies (yet) that test Diverse Belief correlations with linguistic ability.

[53]Grammar does not stop developing after age 6 or 7, but the grammatical constructions that are learned in middle childhood (between 6 and 10) are often exceptions, and are relatively infrequent, involving access to more than one type of grammatical knowledge (Goodluck, 1991, p. 98).

[54]In typically developing children, mature forms of negation appear after the age of four. After the age of five, there is an increase in the amount of complex sentences (e.g., embedded sentences) produced relative to simple sentences (Tager-Flusberg et al., 2005, p. 338).

rect usage of the terms in complex situations, and especially correct usage in context of talk about talk (about what is said or about the sentences uttered) rules out that these children are using the term 'false' to mean something other than FALSE, and shows that the representational character of sentences is understood. Therefore, autistic children with a grammar level of a typically developing 6 to 9 year old understand the representational character of language; (because they must have already mastered negation and embedded sentences). To corroborate this, there is also specific evidence that the use of negation in combination with communication verbs (such as "saying") is mastered before false belief tests are passed, in some autistic children.[55] It is clear, then, that no case can be made that these children must mean something with their words in the natural sense, when they speak.

The speakers in our subset do not mean something in the natural sense with their words, but from this it cannot be inferred without argument that they mean something nonnaturally: Grice distinguishes, on immediate intuitive grounds, two senses of meaning in our day to day use of the term, but this doesn't mean that more careful reflection on our use of the term would not distinguish a third (or even more) type(s) of meaning. A third type of meaning might be implicit in our use, of which we might not be immediately aware. Careful reflection might bring it out. That this is not a ridiculous proposition becomes clear when one realises that Grice offered at least some reason to suppose that nonnatural meaning might be partly characterised on the basis of a notion of communication. This reason is not a good reason—I am going to argue against it. But I think it cannot be dismissed out of hand, and that therefore an argument against this possibility should be added to Glüer and Pagin's argument.

What reason may one have to suppose that nonnatural meaning should be partly characterised on the basis of a notion of communication? First of all, as noted above, Grice sees it as *intuitively clear* that one of two uses of meaning is "related to communication" (namely, nonnatural meaning), and it is clear that on the analysis of the most fundamental type of nonnatural meaning ("Speaker Meaning"), a

[55]Tager-Flusberg and Joseph (2003).

speaker meaning something with her words entails that the speaker intends to communicate with the hearer (in an intuitive sense of *communication*). Also, Grice's systematic reflection on the intuitions regarding the use of the term "meaning" while analysing, suggests that the idea of communication was kept in mind during this process (as we will see below). So, it would not be farfetched to believe that nonnatural meaning is partly characterised on the basis of a notion of communication. This possibility needs to be further evaluated, especially because autistic speakers, and certainly speakers that do not yet pass false belief tests, have trouble with communication, including linguistic communication (as evidenced by their pragmatic problems, for instance). Therefore, the autistic speakers in our subset may not mean anything *nonnaturally* if doing so requires one to communicate, or requires having a sophisticated concept of communication, or requires having the intention to communicate.

One person who criticises Glüer and Pagin's argument on the basis of the communicative problems autistic speakers are known to have, is Ann Reboul (2006). Reboul argues that autistic speakers' communicative problems show that they do not mean (in Reboul's words: "do not assert"[56]) in the same way as typically developing speakers. Behind Reboul's argument, is the fact that she takes Glüer and Pagin to be arguing that "[h]igher order intentions [...] do not play a role in the process [of linguistic communication]."[57] She believes that Glüer and Pagin look at *particular instances* of assertion (something that is arguably both a matter of meaning something and of communication),[58] and aim to show that *that* particular instance could have taken place without the speaker possessing the required higher order intentions. So, Reboul suggests, their argument is analogous to "other epiphenomenal"[59] arguments such as Chalmers': in his argument for the epiphenomenal character of qualia, Chalmers

[56]Reboul (2006, p. 592).

[57]Reboul (2006, p. 590) does not properly explain why she interprets Glüer and Pagin's argument in this manner; but she observes it right after noting that Glüer and Pagin do not claim that typically developing speakers do not "have higher order intentions when they communicate" (Reboul, 2006, p. 590).

[58]Reboul (2006, p. 588).

[59]Reboul (2006, p. 591).

invites people to conceive of "twin earth," which is a world that is micro-physical identical to ours, and to conceive of our counterparts in it, who are behaviourally indistinguishable from us but who do not experience qualia. That such a world is conceivable, and therefore possible, implies that qualia are not causally effective in the production of behaviour, or so the argument goes very roughly. However—Reboul's argument continues—because Chalmers describes a hypothetical situation, he can stipulate that the behaviour in the two instances (on earth and on Twin Earth) is the same. Whereas, because Glüer and Pagin are describing a real situation, they have to *argue* that the particular instances of assertion are similar. But, focusing just on syntactic form, a sentence of that form may mean one thing coming out of the mouth of one person and another thing coming out the mouth of another.[60] To assess similarity, it becomes necessary, then, to refer to content, she says: the question is whether a sentence of a particular form would fulfil the same role in communication when spoken by autistic speakers as it would have done if it had been spoken by typically developing speakers. But the communicative problems show that for autistic speakers, assertions do not play the same role, and therefore, Glüer and Pagin's argument doesn't go through.

I agree with Reboul only that Glüer and Pagin offer a conceptual "possibility" that happens to be an actuality (the conceptual possibility would have been a sufficient counterexample), and that they therefore can't stipulate that what is meant by autistic speakers is meant "in the same way" as what is meant by typically developing children—it requires argument that they mean in the same way. But it is *not* necessary, for Glüer and Pagin's argument, that there is a similarity between *particular* assertions that autistic speakers make and typically developing children (would) make: that these assertions have the same content, play the same role, or whatever—what matters is that in both types of cases something is meant (in the nonnatural sense). Also, nothing Glüer and Pagin say entails that higher order intentions do not play a role in *communication* (they

[60]Reboul (2006, p. 592).

address Grice's analysis of meaning only[61]). Nor do they imply that higher order intentions could not play a role in the process whereby a speakers comes to mean something, or even in the act of a speaker "meaning something".[62]—it is just that nonnatural meaning cannot be analysed in the way proposed, if the counterexample is successful.

So, it is not that particular contents of the speech of autistic speakers and typically developing speakers need to be the same, but they need to be of the same type. Communicative problems may well show that particular contents are dissimilar, but we have yet to establish whether communicative problems have a bearing on what type of meaning autistic speakers exhibit.

Glüer and Pagin themselves are sensitive to the suggestion that communication is relevant to meaning something nonnaturally. They describe Grice's distinguishing between natural and nonnatural meaning in the following way: "In an effort to separate out nonnatural meaning from mere causal regularities Grice correctly turned to the intention to communicate [...]."[63] However, they say, even if an intention to communicate (partly) characterizes meaning something nonnaturally, this does not mean that autistic speakers don't mean nonnaturally. Although the autistic speakers in the subset do not possess the same concept of communication as typically developing speakers, they nevertheless can distinguish communicative situations on the basis of perceptual similarity, Glüer and Pagin argue, and so may want to create new situations of that perceptual kind.[64] And, autistic speakers do indeed initiate conversations,[65] and so it seems that they do intent to communicate in the more minimal sense just outlined.

However, Grice's notion of communication is not a minimal notion. When he discusses his theory of communication, he talks about the goals of communication (such as the giving and receiving of in-

[61]Although they sometimes write as if they are also addressing his theory of communication, see for instance Glüer and Pagin (2003, p. 47).

[62]A conceptual analysis of "swimming" may not refer to fins or arms, and yet fins or arms may play a role in a particular act of swimming.

[63]Glüer and Pagin (2003, p. 47).

[64]Glüer and Pagin (2003, p. 48).

[65]Glüer and Pagin (2003, p. 47).

formation, influencing and being influenced by others) as goals that are recognised by each participant in a communicative situation as a shared goal: involving a joint purpose, a mutually accepted direction.[66] What reason is there to assume that it is a minimal notion of communication that is related to nonnatural meaning, rather than Grice's own more substantial notion? I believe that whether and which notion of communication plays a part in the analysis of nonnatural meaning can only be settled by taking into account the intuitions concerning our use of the term "meaning" that come to the fore when one is analysing meaning. Since Grice offers an analysis, we can use the references he makes to intuitions during his analysis as a guide to see whether nonnatural meaning is to be characterised in terms of communication and if so, whether the intuition is reliable.

Firstly, although as we saw it is "intuitively clear," according to Grice, that nonnatural meaning is related to communication, by itself this doesn't imply that meaning is characterised in terms of communication (just that it is related to it). Secondly, because the distinction between natural meaning and nonnatural meaning is so intuitively clear, Grice is able to identify the "recognition tests" almost immediately, before subjecting the intuitions regarding the use of the word "meaning" to a more systematic analysis. As can be seen from the list mentioned above, none of these tests for nonnatural meaning refer to the intention to communicate, or explicitly use the concept of communication, or implicitly describe communication.

So, Grice doesn't refer (before commencing on a more systematic reflection) to any intuitions that might imply that nonnatural meaning is to be characterised in terms of communication. Now we need to establish whether he does while engaged in the more careful conceptual analysis (and what this entails about the correct characterisation of meaning): During the analysing process, Grice systematically tests our intuitions regarding the use of the term 'meaning' (in the nonnatural sense), by offering specific hypothetical scenarios and asking in each case whether one would or wouldn't want to apply

[66]See Grice (1989d, p. 28) for Grice's view that talk is adapted to serve a particular purpose, Grice (1989d, p. 26) for Grice's idea that participants in a talk exchange recognise a common purpose or mutually accepted direction, and Grice (1989d, p. 30) for what these purposes are.

the term "meaning" (in the nonnatural sense); all in the service of moving towards a general characterization of one's use of the term "meaning."[67]

Looking at Grice's analysing process, it's clear that Grice doesn't explicitly refer to communication, but his choice of starting point does suggest that the idea of communication (I take it that this idea always involves the idea of *others* as receivers, so to speak) is on his mind: Grice starts out asking whether intending to cause a belief in someone is sufficient to mean something.[68] Counterexamples show that it isn't, but the condition is retained as a necessary condition in the analysis of meaning (as related to the making of statements and the like, not orderings). However, Grice offers no systematic investigation into the necessity of this condition: he has offered no crucial hypothetical scenarios that suggest (directly) that we would indeed be inclined to withhold the term "meaning" in every case where the condition is not fulfilled.

It could be of course, that the necessity of the condition can be (indirectly) deduced from our intuitions regarding the use of the term "meaning" that are being highlighted by the hypothetical circumstances that Grice *does* consider. For example, Grice finds he wants to withhold applying the term "meaning" in a case where someone leaves a handkerchief of a particular person near a scene of a murder (in order to produce in the mind of the police the belief that that person is the murderer). Or, when someone shows a husband a photograph of his wife having an affair (in order to produce in the husband the belief that his wife is having an affair). On the other hand, Grice finds he doesn't feel reluctant to withhold the term "meaning" in a case where someone shows the husband a *drawing* of his wife having an affair (in order to etc.). Grice finds that what prevents him from applying the term "meaning" in the first two cases, is that the belief would have been produced in the audience anyway, even if there had not been an accompanying intention by the speaker (if the photograph had been lying around by accident, for instance). This leads him to add as a condition that the belief production in the

[67] Grice (1989a).
[68] Grice (1989a, p. 217).

audience must proceed via the recognition that the speaker intends to produce that belief in them.[69]

However, our intuitions to withhold applying the term "meaning" in the first two counterexamples, can also be explained in another way. I want to suggest that if it were possible for a speaker to present a photograph as being true (which requires the possibility that it could be false), or the handkerchief as being true, then we would not feel inclined to withhold the application of the term "meaning" in these cases. In such a case it would not be possible for the handkerchief or the photograph to be causing beliefs in the audience just by lying around—they would only do so if presented as true. I am not offering this as a step in an analysis of my own (I am not suggesting all matters of meaning nonnaturally are a matter of presenting as true or as false), but simply as an observation that, although Grice seem to have been led by the idea of communication in his analysis, it is not necessary to interpret our intuitions concerning the use of the term "meaning" as involving an idea of communication.

So, neither in making the intuitive distinction between natural meaning and nonnatural meaning, and offering recognition tests, nor in the more careful analysis of nonnatural meaning, have intuitions regarding the use of the term 'meaning' become apparent that would suggest that meaning is to be characterised in terms of communication. I believe it can therefore be concluded that nonnatural meaning is not to be characterised in terms of communication.

Should another type of meaning be distinguished in our use of the term "meaning," apart from the natural sense, the nonnatural sense, and the now rejected "nonnatural sense characterised in terms of communication"? Since nonnatural meaning is related to communication in general, there was some *reason* to think that nonnatural meaning should be characterised in terms of a notion of communication. But it is unlikely that another type of meaning should be distinguished in our use of the term 'meaning' that has so far completely avoided detection (and, incidentally, it is unlikely that Grice

[69]Grice (1989a, pp. 217–219). I pass over a step in Grice's argument: namely that the second counterexample shows that the belief production must proceed via the intention recognition and that the first counterexample just shows that there must be intention recognition. These details do not affect the argument.

himself would think this is a possibility, given that he proposed that one should not multiply senses beyond necessity):[70] because if the autistic speakers somehow did not fit comfortably in either the "natural" meaning or the "nonnatural" meaning category, there may be a reason to suspect that a third distinction was implicit in our use, applicable to the autistic speakers in our subset, but this is not the case: all the recognition tests that have been offered with regard to nonnatural meaning, comfortably fit the autistic speakers.

We have established that, without being able to communicate in Grice's substantial sense, or form the intention to so communicate (whether minimal or substantial), autistic speakers can understand that sentences that are uttered can be true or false, and can themselves utter sentences that they understand to be either true or false. It is just as difficult to establish whether autistic speakers are in fact presenting sentences as true, or as false, rather than just mouthing, or using words purely "associatively," as it is for any speaker—this depends on context. But if they do, it can be concluded that they, as speakers, mean something, that what is meant can be paraphrased in a sentence using quotation marks, and that it would not be absurd to say things like "she (the autistic speaker) means that the cat is on the mat, but the cat is not on the mat."

I have argued that Reboul's contention that the communicative impairments of the autistic speakers in our subset suggest that their particular assertions do not have the same meaning or do not play the same role in communication as syntactically identical sentences would have had, if they had been uttered by typically developing speakers, is not relevant to Glüer and Pagin's argument, because Glüer and Pagin aim to offer a counterexample to a Grice's analysis of a meaning type (not an argument that higher order intentions play no role in the formation of particular instances of meaning). However, I have tried to argue that the idea that autistic speaker's communicative impairments suggest that autistic speakers may not mean in the nonnatural sense, in other words, may not mean in the sense of meaning that Grice analyses, needs to be dealt with; it is not obvious that a speaker with a vocabulary age of between 9 and 11 means

[70] Grice (1989e, p. 47).

something nonnaturally, and therefore not obvious without further argument that autistic speakers pose a counterexample to Grice's analysis of nonnatural meaning. A subset of autistic speakers show advanced grammatical ability before being able to attribute beliefs, an ability that does not differ significantly from adult grammar, and where in particular correct use (in context) of negation and affirmation are in place, in embedded sentences such as "She said that x, but it is not true that x." I have argued that this rules out the hypothesis that the speakers in the subset mean in the natural sense. The same grammatical data also show that the autistic speakers' type of meaning perfectly passes the recognition tests, proposed by Grice to help one decide with which type of meaning one was dealing, if in doubt. No other intuitions, suggesting either that nonnatural meaning should be characterised in terms of communication, or that some other type of meaning should be distinguished alongside natural and nonnatural meaning, have been revealed by contemplating Grice's analysing process, and I have therefore concluded that the subset of autistic speakers do indeed mean nonnaturally, and therefore form a counterexample to Grice's account. And so, we can conclude that Glüer and Pagin's argument still stands.

References

APA (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. American Psychiatric Association, Washington, 4th edition.

Astington, J. W. and Jenkins, J. M. (1995). Theory of mind development and social understanding. *Connections between emotion and understanding in development*, 9(2/3):151–165.

Baron-Cohen, S. (1989). Are autistic children "behaviorists"? An examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders*, 19(4):579–600.

Baron-Cohen, S. (1991). Do people with Autism understand what causes emotion? *Child Development*, 62(2):385–395.

Baron-Cohen, S. (1992). Out of sight or out of mind? Another look at deception in autism. *Journal of Child Psychology and Psychiatry*, 33(7):1141–1155.

Baron-Cohen, S. and Goodhart, F. (1994). The "seeing-leads-to-knowing" deficit in autism: The Pratt and Bryant probe. *British Journal of Developmental Psychology*, 12(3):397–401.

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1):37–46.

Baron-Cohen, S., Spitz, A., and Cross, P. (1993). Can children with autism recognize surprise. *Cognition and Emotion*, 7:507–516.

Bryson, S. E., Clark, B. S., and Smith, I. M. (1988). First report of a Canadian epidemiological study of autistic syndromes. *Journal of Child Psychology and Psychiatry*, 29(4):433–445.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., and Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, 15(4):481–498.

Charman, T., Drew, A., Baird, C., and Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language*, 30(1):213–236.

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1(4):568–570.

Dennis, M., Lazenby, A., and Lockyer, L. (2001). Inferential language in high-function children with autism. *Journal of Autism & Developmental Disorders*, 31(1):47–54.

Dunn, L. M. and Dunn, L. M. (1997). *Examiner's manual for the PPVT-III peabody picture vocabulary test*. American Guidance Service, Circle Pines, 3rd edition.

Dunn, L. M., Dunn, L. M., Whetton, C., and Pintilie, D. (1982). *British Picture Vocabulary Test*. NFER-Nelson, London.

Eisenmajer, R. and Prior, M. (1991). Cognitive linguistic correlates of "theory of mind" ability in autistic children. *British Journal of Developmental Psychology*, 9(2):351–364.

Fisher, N., Happé, F., and Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry*, 46(4):409–419.

Frith, U., Happé, F., and Siddons, F. (1994). Autism and theory of mind in everyday life. *Social Development*, 3(2):108–124.

Glüer, K. and Pagin, P. (2003). Meaning theory and autistic speakers. *Mind and Language*, 18(1):23–51.

Goodluck, H. (1991). *Language acquisition: A linguistic introduction*. Blackwell Publishers, Oxford.

Grice, P. (1989a). *Studies in the Ways of Words*, chapter Meaning, pages 213–223. Harvard University Press, Cambridge, Mass.

Grice, P. (1989b). *Studies in the Ways of Words*, chapter Meaning revisited, pages 283–303. Harvard University Press, Cambridge, Mass.

Grice, P. (1989c). *Studies in the Ways of Words*, chapter Utterer's meaning and intentions, pages 86–116. Harvard University Press, Cambridge, Mass.

Grice, P. (1989d). *Studies in the Ways of Words*, chapter Logic and conversation, pages 22–40. Harvard University Press, Cambridge, Mass.

Grice, P. (1989e). *Studies in the Ways of Words*, chapter Further notes on logic and conversation, pages 41–57. Harvard University Press, Cambridge, Mass.

Hamilton, A. F. C., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition*, 113(1):37–44.

Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48(2):101–119.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Cognition*, 66(3):843–855.

Jarrold, C., Boucher, J., and Russell, J. (1997). Language profiles in children with autism: Theoretical and methodological implications. *Autism*, 1(1):57–76.

Laurence, S. (1998). Convention-based semantics and the development of language. In Carruthers, P. and Boucher, J., editors, *Language and Thought. Interdisciplinary Themes*, pages 201–218. Cambridge University Press, Cambridge.

Lewis, V. and Boucher, J. (1988). Spontaneous, instructed and elicited play in relatively able autistic children. *British Journal of Developmental Psychology*, 6(4):325–339.

Minshew, N. J., Meyer, J., and Goldstein, G. (2002). Abstract reasoning in autism: A dissociation between concept formation and concept identification. *Neuropsychology*, 16(3):327–334.

Neale, S. (1992). Paul Grice and the philosophy of language. *Linguistics and Philosophy*, 15(5):509–559.

Paul, R., Chawarska, K., Klin, A., and Volkmar, F. (2007). Dissociations in the development of early communication in autism spectrum disorders. In Paul, R., editor, *Language disorders from a developmental perspective: Essays in honor of Robin Chapman*, pages 163–194. Lawrence Erlbaum Associates, Inc., Mahwah.

Peralejo, J. (2008). A longitudinal study of lexical development in young children with autism spectrum disorders. Master's thesis, University of British Columbia.

Peterson, C. C., Wellman, H. M., and Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child development*, 76(2):502–517.

Plug, M. (2010). Autistic Speakers as a Counterexample to Grice and Davidson on Meaning. Unpublished PhD thesis, King's College London.

Reboul, A. (2006). HOT theories of meaning: The link between language and theory of mind. *Mind & Language*, 21(5):587–596.

Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., Osanai, H., and Csibra, G. (In Press). Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and Psychopathology*.

Senju, A., Southgate, V., White, S., and Frith, U. (2009). Mind-blind eyes: An absence of spontaneous theory of mind in Asperger Syndrome. *Science*, 325(5942):883–885.

Surian, L., Baron-Cohen, S., and Van der Lely, H. (1996). Are children with autism deaf to Gricean maxims? *Cognitive Neuropsychiatry*, 1(1):55–72.

Surian, L. and Leslie, A. M. (1999). Competence and performance in false belief understanding: A comparison of autistic and normal 3-year-old children. *British Journal of Developmental Psychology*, 17(1):141–155.

Tager-Flusberg, H. (1985). Basic level and superordinate level categorization by autistic, mentally retarded and normal children. *Journal of Experimental Child Psychology*, 40(3):450–469.

Tager-Flusberg, H. (1992). Autistic children's talk about psychological states: Deficits in the early acquisition of a theory of mind. *Child Development*, 63(1):161–172.

Tager-Flusberg, H., Calkins, S., Nolin, T., Baumberger, T., Anderson, M., and Chadwick-Dias, A. (1990). A longitudinal study

of language acquisition in autistic and down syndrome children. *Journal of Autism and Developmental Disorders*, 20(1):1–21.

Tager-Flusberg, H. and Joseph, R. (2003). Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences*, 358(1430):303–314.

Tager-Flusberg, H., Paul, R., and Lord, C. (2005). Language and communication in autism. In Volkmar, F. R., editor, *Handbook of autism and pervasive developmental disorders*, volume 1: Diagnosis, development, neurobiology, and behavior, pages 335–364. John Wiley & Sons, Hoboken, 3rd edition.

Tager-Flusberg, H. and Sullivan, K. (1994). Predicting and explaining behavior: A comparison of autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry*, 35(6):1059–1075.

Toichi, M. and Kamio, Y. (2001). Verbal association for simple common words in high-functioning autism. *Journal of Autism and Developmental Disorders*, 31(5):483–490.

Ungerer, J. A. and Sigman, M. (1987). Categorization skills and receptive language development in autistic children. *Journal of Autism and Developmental Disorders*, 17(1):3–116.

Wimmer, H. and Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1):103–128.

Ziatas, K., Durkin, K., and Pratt, C. (1998). Belief term development in children with autism, Asperger Syndrome, specific language impairment, and normal development: Links to theory of mind development. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 39(5):755–763.

Achourioti, T., Andrade, E.J., Staudacher, M. (eds.), Proceedings of
the Amsterdam Graduate Philosophy Conference — Meaning and Truth,
ILLC Publications X-2010-1, pages 103–115

Computing with Self-Reference

Stefan Wintein (Tilburg University)[*]

S.Wintein@uvt.nl

1 Introduction

Depending on one's notion of meaning, a Liar sentence is considered to be either *meaningful* or *meaningless*. An argument for the latter position may be that the Liar sentence is lacking meaning because it is *useless*. For instance, one might hold that the meaning of a sentence is to be explained by its canonical use in language, which is to make a statement, and that the Liar sentence fails to deliver a statement and is without meaning accordingly. Declaring the Liar to be meaningless is a way to escape its paradoxical consequences and to stay faithful to our naive theory of truth under the assumption that only *statement sentences* are to be plugged in into Tarksi's biconditionals. This *meaning-error approach* to the Liar was predominant among Medieval scholars and is nowadays defended by, for example, Goldstein and Blum (2008) and Thalos (2005).

In contrast, the *truth-error approach* is typically taken by philosophers who feel that whether or not a certain sentence type, such as, e.g., the Liar sentence, is meaningful depends on more general, meaning constitutive notions such as *reference* and *extension*. And from this more structural approach it is hard to see what exactly is wrong with the Liar. It cannot be the involved self-reference, for there are also unproblematic self-referential sentences nor can it be any reason

[*]Thanks to Reinhard Muskens, the participants of AGPC09 and to Frank Veltman, who commented on this paper during AGPC09.

purporting only to *intrinsic* features of the Liar for, as Kripke (1975) observed, the paradoxality of "Liar-like sentences" such as 'Most things Nixon said about Watergate are false' depend on *empirical* circumstances. In fact, the truth-error approach holds that from the perspective of meaning *nothing* is wrong with the Liar. Rather, according to this approach, the Liar reveals that it is our *naive theory of truth that is in error*. An adherent of the truth-error approach may regard the Liar as useful in the sense that an anomalous measurement result is useful in discrediting an otherwise plausible theory. However, this sense of usefulness—useful in teaching us something *about* (the truth fragment of our) language—is radically different from the sense in which we used 'usefulness' in the first paragraph—useful *within* language. In the latter sense then, the Liar seems to be *useless* on the truth-error approach as well. That the distinction between meaningful and meaningless sentences is independent of the distinction between useless and useful sentences seems to be recognized by the late Wittgenstein—whose thoughts are often summarized by philosophers via the slogan 'meaning is use'—as the following words testify.

> If the question is whether this [The Liar] is a statement at all, I reply: You may say that it is not a statement. Or you may say that it *is* a statement, but a useless one (Wittgenstein, 1939, p. 209).

In this paper I will show that this widely shared conviction that the Liar and other *ungrounded sentences*[1] are useless is wrong.

The setting in which this claim is established involves an *agent* (Section IV) who asks *yes-no questions* to an omniscient being, called an *oracle* (Section III), in order to gain information that fills his epistemic lacunas. I call such situations *query structures* (Section II) and I will prove that query structures can be solved more efficiently in the presence of ungrounded sentences than without such resources (Section V). Thereby, I illustrate the usefulness of such sentences or, in other terms, I establish the fact that self-referentiality increases computational power.

[1] Ungrounded sentences are those sentences that do not receive a truth value in the minimal fixed point of Kripke's fixed point construction.

II Query structures and resources

For sake of definiteness, L will denote a fixed language whose vocabulary consists of the set $P = \{p_1, p_2, \ldots\}$ of sentential atoms together with the set $\{\neg, \wedge, \vee\}$ of sentential connectives. Its set of sentences $Sen(L)$ is constructed in the usual way. We use $\vdash \subseteq \mathcal{P}(Sen(L)) \times Sen(L)$ to denote the standard sentential consequence relation. Throughout the paper, W will denote the set of all subsets of P. Members of W can be thought of as *models* for L, which we will also call *possible worlds*. A central notion in this paper is that of a *query structure*.

Definition 1. Query structure
We say that a pair $\mathbf{Q} = \langle \mathcal{B}, \mathcal{T} \rangle$, consisting of the *background knowledge* $\mathcal{B}$ and the *target knowledge* $\mathcal{T}$ of an agent, is a *query structure* just in case:

1. $\mathcal{B}, \mathcal{T} \subseteq Sen(L)$. $\mathcal{B} \cup \mathcal{T}$ is consistent. $|\mathcal{T}|$ is finite.
2. $\exists \sigma \in \mathcal{T} : \mathcal{B} \nvdash \sigma$ and $\mathcal{B} \nvdash \neg\sigma$. (non-triviality)

The agent tries to *solve a query structure*, by which we mean that he tries to determine the truth value of all the sentences of his target knowledge $\mathcal{T}$, i.e. he tries to *decide* $\mathcal{T}$. In order words, he tries to extend his background knowledge $\mathcal{B}$ to, say $\mathcal{B}^*$, where $\mathcal{B}^*$ is such that for every $\sigma \in \mathcal{T}$: $\mathcal{B}^* \vdash \sigma$ or $\mathcal{B}^* \vdash \neg\sigma$. The non-triviality condition imposed on query structures ensures that the agent's background knowledge by itself does not decide $\mathcal{T}$. In order to decide $\mathcal{T}$, the agent *queries* an oracle or, equivalently, he asks yes-no questions—modeled as declarative sentences—to the oracle. The yes-no questions that an agent is allowed to ask the oracle in order to solve a query structure are called the agent's *resources*. We will study how the resources of the agent influence his ability to solve query structures. The agent's *classical resources* are given by $Sen(L)$. The fun begins when we are more liberal in the allotment of resources to the agent; in general, the resources of an agent will consist of the sentences of a *referential T-enrichment* of L.

Definition 2. Referential T-enrichments
A *T-enrichment* L^C of L enriches the vocabulary of L with a *finite*

(possibly empty) set of constant symbols C—called 'the non quotational constants of L^C'—and a unary *truth predicate symbol* 'T'.Its set of sentences $Sen(L^C)$ is jointly defined with its set of constants $Con(L^C)$, which are the smallest sets[2] that satisfy the following conditions.

- $P \subseteq Sen(L^C)$, $C \subseteq Con(L^C)$.
- $\alpha \in Sen(L^C) \Rightarrow [\alpha] \in Con(L^C)$.
- $t \in Con(L^C) \Rightarrow T(t) \in Sen(L^C)$.
- $\alpha, \beta \in Sen(L^C) \Rightarrow \neg\alpha, (\alpha \wedge \beta), (\alpha \vee \beta) \in Sen(L^C)$.

A *referential T-enrichment* $\mathcal{L} = \langle L^C, \pi\rangle$ is a pair consisting of a T-enrichment L^C and a function $\pi : Con(L^C) \to Sen(L^C)$, called a *reference list* that satisfies

$$\pi([\sigma]) = \sigma \qquad \text{for all } \sigma \in Sen(L^C)$$

With $\mathcal{L}$=$\langle L^C, \pi\rangle$ we let $Con(\mathcal{L})$=$Con(L^C)$. $\mathcal{L}^{\mathsf{basic}}$ has C=$\{\lambda, \tau\}$, with $\pi(\lambda) = \neg T(\lambda)$ and $\pi(\tau) = T(\tau)$. Thus $\mathcal{L}^{\mathsf{basic}}$ contains a (strengthened) *Liar sentence*, '$\neg T(\lambda)$' and a *Truthteller*, '$T(\tau)$'.

III The oracle

The oracle is *omniscient*, by which we mean that he knows the truth value of each atomic sentence—equivalently, that he knows which world $w \in W$ is actual. The answer of the oracle to a question $\sigma \in Sen(\mathcal{L})$ is a *truthful* report of the oracle's abilities to fulfill the *assertion* respectively *denial commitments* of σ. Those assertoric commitments for the oracle with respect to $\sigma \in Sen(\mathcal{L})$ are determined by the *assertoric rules* of $\mathcal{L}$. The assertoric rules of $\mathcal{L}$ are mirrored, so to speak, by the inference rules of $\mathcal{L}$. We explain the idea of an assertoric rules that is mirrored by an inference rule via the following schematically depicted rule.

[2] Sentences that are constructed using connectives in $\{\rightarrow, \leftrightarrow\}$ are officially in the meta-language; the translation of them to "official sentences" is achieved in the usual manner.

(1) $$\frac{(\alpha \wedge \beta)}{\alpha, \beta}$$

An *inferential reading* of (1) would be along the following lines: "from $(\alpha \wedge \beta)$ you may infer α as well as β" An *assertoric reading* of (1) is as follows: "*if* one asserts $(\alpha \wedge \beta)$ *then* one is *committed* to be able to assert α as well as β". *Denying* a conjunction $(\alpha \wedge \beta)$ also brings with it assertoric commitments. $\mathcal{L}$'s assertion rule $(a\wedge)$ and denial rule $(d\wedge)$ for conjunctions are stated, in suitable notation, as follows:

$$(a\wedge)\ \frac{A_{(\alpha\wedge\beta)}}{A_\alpha, A_\beta} \qquad\qquad (d\wedge)\ \frac{D_{(\alpha\wedge\beta)}}{D_\alpha \ \mid\ D_\beta}$$

Note that $(d\wedge)$ testifies that we are considering assertoric rules for an *oracle*; a non-omniscient entity may very well deny a contradiction $p\wedge\neg p$ without being committed to be able to deny p or $\neg p$. With the assertoric rules for conjunction already stated, the other assertoric rules for $\mathcal{L}$ are stated below, where $t \in Con(\mathcal{L})$.

$$(a\neg)\ \frac{A_{\neg\alpha}}{D_\alpha} \qquad (d\neg)\ \frac{D_{\neg\alpha}}{A_\alpha} \qquad (a\vee)\ \frac{A_{(\alpha\vee\beta)}}{A_\alpha \ \mid\ A_\beta} \qquad (d\vee)\ \frac{D_{(\alpha\vee\beta)}}{D_\alpha, D_\beta}$$

$$(aT)\ \frac{A_{T(t)}}{A_{\pi(t)}} \qquad (dT)\ \frac{D_{T(t)}}{D_{\pi(t)}}$$

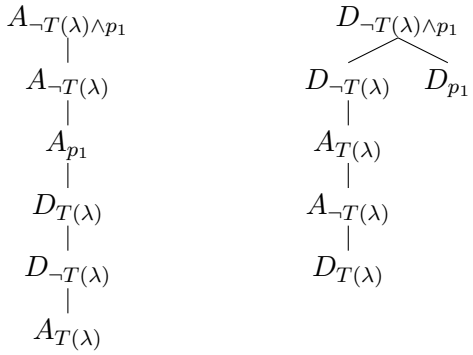
Thus, the assertoric rules are the usual tableau rules for $\wedge, \vee, \neg$ in terms of signed statements (see Smullyan (1995)), augmented with rules that govern the sentences of $\mathcal{L}$ which have the truth predicate as their main logical symbol. Expressions of form X_σ with X equal to A or D and with $\sigma \in Sen(\mathcal{L})$ we call *assertoric sentences*.

The assertoric rules state the commitments involved with asserting respectively denying the *compounded* sentences of $\mathcal{L}$. The commitments involved with asserting (denying) the atomic sentences in P are not stated in terms of the abilities of the oracle to assert or deny further sentences, but are stated in terms of his knowledge of

the truth value of the atomic sentences. In asserting $p \in P$, the oracle is committed to know that p is true, and in denying $p \in P$, the oracle is committed to know that p is false. Before the oracle reaches the conclusion that he can or cannot fulfill the assertion (denial) commitments of σ, he goes through a process of reasoning that is governed by the assertoric rules. In order to keep track of the process we introduce *assertoric trees*.

By the *assertoric trees* of σ we mean its *assertion tree* $\mathfrak{T}^{\sigma}_{A}$ and its *denial tree* $\mathfrak{T}^{\sigma}_{D}$. $\mathfrak{T}^{\sigma}_{A}$ keeps track of the assertion commitments σ while $\mathfrak{T}^{\sigma}_{D}$ keeps track of σ's denial commitments. Assertoric trees have occurrences of assertoric sentences as their points; $\mathfrak{T}^{\sigma}_{X}$ is the assertoric tree which has X_{σ} as its origin and which is obtained from its origin by a "tree expansion process" that follows the assertoric rules of $\mathcal{L}$. This tree expansion process does not differ from familiar tree-constructions based on (inferential) rules and is illustrated by the following example.

Example 1. Assertoric trees
Below we depict $\mathfrak{T}^{\gamma}_{A}$ and $\mathfrak{T}^{\gamma}_{D}$, where $\gamma := (\neg T(\lambda) \wedge p_1) \in Sen(\mathcal{L}^{\mathsf{basic}})$.



The example illustrates that, due to the finiteness C, an assertoric tree is a *finite* object.

The oracle can fulfill the assertion commitments of σ just in case $\mathfrak{T}^{\sigma}_{A}$ is *open relative to the actual world* $w_{@} \in W$ and he can fulfill the denial commitments of σ just in case $\mathfrak{T}^{\sigma}_{D}$ is *open relative to* $w_{@}$. The

notion of an assertoric tree being open/closed relative to a world is defined as follows.

Definition 3. Closure conditions for assertoric trees
Let $\sigma \in Sen(\mathcal{L})$, let $\mathfrak{T}^{\sigma}_{X}$ be an assertoric tree of σ, let $w \in W$ and let B be a branch of $\mathfrak{T}^{\sigma}_{X}$. B *is closed in* w iff 1. or 2. holds.

1. $\exists\, \sigma \in Sen(\mathcal{L}) : A_{\sigma} \in B$ and $D_{\sigma} \in B$.
2. $\exists\, p \in P :$ ($A_p \in B$ and $p \notin w$) or ($D_p \in B$ and $p \in w$)

B is *open in* w iff B is not closed in w. $\mathfrak{T}^{\sigma}_{X}$ is *closed in* w iff all its branches are closed in w and $\mathfrak{T}^{\sigma}_{X}$ is *open in* w iff $\mathfrak{T}^{\sigma}_{X}$ is not closed in w. A branch B is said to be *closed (open) a priori* iff it is closed (open) in every $w \in W$. An *a posteriori branch* is a branch that is neither open nor closed a priori. We write $Cl_w(\mathfrak{T}^{\sigma}_{X})$ and $O_w(\mathfrak{T}^{\sigma}_{X})$ to denote that $\mathfrak{T}^{\sigma}_{X}$ is closed, respectively open in w.

The oracle is *truthful*, by which we mean that his answers to yes-no questions disclose his (dis-)abilities to fulfill the assertoric commitments of the associated sentences. As the assertoric commitments of a sentence divide into the assertion and denial commitments, the oracle needs to have available *four* distinct reactions in his answering repertoire in order to be able to answer truthfully. For each $w \in W$ and $\sigma \in Sen(\mathcal{L})$, the *answering function* returns the (code for the) answer that the oracle would give to σ if w were the actual world.

Definition 4. The answering function
Let $w \in W$. The *answering function* $\mathcal{A}_w$:$Sen(\mathcal{L})\rightarrow\{0,1\} \times \{0,1\}$ returns the answer given by the oracle in w to $\sigma \in Sen(\mathcal{L})$. A pair (x_1, x_2) reflects whether the assertion tree of the sentence under consideration is open ($x_1 = 1$) or closed ($x_1 = 0$) and whether the denial tree is open ($x_2 = 1$) or closed ($x_2 = 0$). That is:

1. $Cl_w(\mathfrak{T}^{\sigma}_{A})$ and $O_w(\mathfrak{T}^{\sigma}_{D}) \Leftrightarrow_{def} \mathcal{A}_w(\sigma) = (0,1)$ ('no')
2. $O_w(\mathfrak{T}^{\sigma}_{A})$ and $Cl_w(\mathfrak{T}^{\sigma}_{D}) \Leftrightarrow_{def} \mathcal{A}_w(\sigma) = (1,0)$ ('yes')
3. $Cl_w(\mathfrak{T}^{\sigma}_{A})$ and $Cl_w(\mathfrak{T}^{\sigma}_{D}) \Leftrightarrow_{def} \mathcal{A}_w(\sigma) = (0,0)$ ('neither')

4. $O_w(\mathfrak{T}^\sigma_A)$ and $O_w(\mathfrak{T}^\sigma_D) \Leftrightarrow_{def} \mathcal{A}_w(\sigma) = (1,1)$ ('both')

Our answering function resembles the (Dunn-Belnap) valuation function familiar from work on four-valued logics (see for example Muskens (1999)) to some extent. However, in sharp contrast to this valuation function, our answering function is not compositional, as is illustrated by the following example.

Example 2. Non-compositionality
As the reader may verify, for every $w \in W$ we have that $\mathcal{A}_w(T(\tau))$ $= (1,1)$ and that $\mathcal{A}_w(\neg T(\tau)) = (1,1)$. Also, we have that $\mathcal{A}_w(T(\tau)$ $\wedge T(\tau)) = (1,1)$ while $\mathcal{A}_w(T(\tau) \wedge \neg T(\tau)) = (0,1)$ and $\mathcal{A}_w(T(\tau) \vee T(\tau))$ $= (1,1)$ illustrating that the answering function is not "answer compositional".

IV The agent

The agent can faultlessly apply the assertoric rules of $\mathcal{L}$, knows that the oracle is omniscient and that it provides truthful answers of which the agent knows the meaning, e.g., he knows that the answers 'yes' to σ indicates that the oracle can fulfill σ's assertion commitments while he cannot fulfill σ's denial commitments. The essential difference between the agent and the oracle is that the oracle knows which world is actual, whereas the agent does not. The oracle's answers to σ reveal whether the assertoric trees of σ are open /closed and this information can be converted into information about the truth value of sentences of L. The latter information is used by the agent to update his knowledge. We will define a knowledge update function $\mathcal{U}$, mapping each question-answer pair to the associated update.

With $p \in P$, we let $(A_p)^+ = p, (D_p)^+ = \neg p$, $(A_p)^- = \neg p$ and $(D_p)^-$ $= p$. Let B be an *a posteriori* branch of assertoric tree $\mathfrak{T}^\sigma_X$ and let $\mathbf{B}$ be the set of all atomic assertoric sentences Y_p that occur on B.[3] With $Y_p \in \mathbf{B}$, B^- is the disjunction of all the $(Y_p)^-$ and B^+ is the conjunction of all the $(Y_p)^+$. Intuitively, B^+ is what the agent learns when he finds out that B is open while B^- is what he learns

[3] As B is a posteriori, $\mathbf{B} \neq \emptyset$.

when he finds out that B is closed. In receiving an answer to σ, the agent is assured either that $\mathfrak{T}^\sigma_X$ is closed or that $\mathfrak{T}^\sigma_X$ is open. Such an assurance is *uninformative* iff all the branches of $\mathfrak{T}^\sigma_X$ are closed a priori or if $\mathfrak{T}^\sigma_X$ has a branch that is open a priori. If one of those two conditions apply to $\mathfrak{T}^\sigma_X$, we let the closed-update of $\mathfrak{T}^\sigma_X$—the information that the agent learns by finding out that $\mathfrak{T}^\sigma_X$ is closed— and the open-update of $\mathfrak{T}^\sigma_X$—the information that the agent learns by finding out that $\mathfrak{T}^\sigma_X$ is open—be equal to a tautology. Formally, $\mathcal{C}(\mathfrak{T}^\sigma_X) = \mathcal{O}(\mathfrak{T}^\sigma_X) = p_1 \vee \neg p_1$. Otherwise, we let $\mathcal{C}(\mathfrak{T}^\sigma_X)$ be the conjunction of all— ranging over the a posteriori branches of $\mathfrak{T}^\sigma_X$—the B^-, while $\mathcal{O}(\mathfrak{T}^\sigma_X)$ is the disjunction of all the B^+. Using the functions $\mathcal{C}(\cdot)$ and $\mathcal{O}(\cdot)$, we define the *update function* $\mathcal{U} : \mathit{Sen}(\mathcal{L}) \times (\{0,1\} \times \{0,1\}) \to \mathit{Sen}(L)$, mapping each question-answer pair to the associated knowledge update.

$$\mathcal{U}(\sigma,(1,0)) = (\mathcal{O}(\mathfrak{T}^\sigma_A) \wedge \mathcal{C}(\mathfrak{T}^\sigma_D)) \qquad \mathcal{U}(\sigma,(0,1)) = (\mathcal{C}(\mathfrak{T}^\sigma_A) \wedge \mathcal{O}(\mathfrak{T}^\sigma_D))$$

$$\mathcal{U}(\sigma,(0,0)) = (\mathcal{C}(\mathfrak{T}^\sigma_A) \wedge \mathcal{C}(\mathfrak{T}^\sigma_D)) \qquad \mathcal{U}(\sigma,(1,1)) = (\mathcal{O}(\mathfrak{T}^\sigma_A) \wedge \mathcal{O}(\mathfrak{T}^\sigma_D))$$

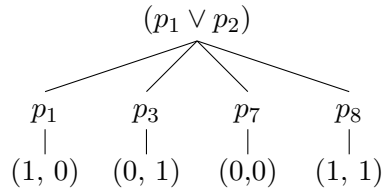
Example 3. Knowledge update (Example 1 continued) With γ as in Example 1, we see that $\mathfrak{T}^\gamma_A$ contains a single branch that is closed a priori. Hence, we have that $\mathcal{C}(\mathfrak{T}^\gamma_A) = \mathcal{O}(\mathfrak{T}^\gamma_A) = p_1 \vee \neg p_1$. As the denial tree has a branch that is closed a priori and an a posteriori branch, we have $\mathcal{C}(\mathfrak{T}^\gamma_D) = p_1$ while $\mathcal{O}(\mathfrak{T}^\gamma_D) = \neg p_1$. In terms of the update function, we have that $\mathcal{U}(\gamma,(1,0)) = \mathcal{U}(\gamma,(0,0)) = (p_1 \vee \neg p_1) \wedge p_1$ and that $\mathcal{U}(\gamma,(0,1)) = \mathcal{U}(\gamma,(1,1)) = (p_1 \vee \neg p_1) \wedge \neg p_1$. The example illustrates that, although the update function is defined for all question-answer pairs, this does not mean in general that all four answers to a sentence are possible answers; depending on the truth value of p_1, the oracle will either answer γ with 'no' or 'neither'.

V The power of self-referential truth

Let $\mathcal{L}$ be a referential T-enrichment. A *query strategy of n questions in $\mathcal{L}$*, $\mathcal{S}^n$, is a plan of the agent to ask n consecutive questions of $\mathit{Sen}(\mathcal{L})$ to the oracle, where the m^{th} question asked according to the plan may depend on the answers to the $m-1$ questions asked before. Formally, $\mathcal{S}^n$ is conveniently represented as a *4-tree* of height

n whose points are occurrences of elements of $Sen(\mathcal{L})$. A 4-tree is a tree in which each point that is not an endpoint has exactly 4 successors. In the case of a query strategy, the four successors of such a point σ represent the follow-up question to σ that will be asked conditional on the answer ('yes', 'no', 'neither', 'both') received to σ. More general, in *executing* a query strategy $\mathcal{S}^n$, the agent starts by asking the question corresponding to the origin of $\mathcal{S}^n$ and he asks further questions depending on the answers he received to previous ones. For instance, if the answer received by the agent to his first question was 'yes', he consequently asks the corresponding follow-up question. Similarly for the other answers and for questions "higher up" $\mathcal{S}^n$. The following example illustrates the notion of a query strategy and its execution.

Example 4. Query strategy of 2 questions in L



The depicted 4-tree corresponds to the following query strategy. First, the agent asks the question '$(p_1 \vee p_2)$' and then the asks a follow-up question which depends on the answer that the oracle gave to '$(p_1 \vee p_2)$'. Depending on whether the oracle answered '$(p_1 \vee p_2)$' with 'yes', 'no', 'neither' or 'both' the agent respectively asks 'p_1', 'p_3', 'p_7' or 'p_8' as a follow-up question.

As the oracle's answers to any question are determined by which world $w \in W$ is actual and as the questions which are asked in executing a query strategy $\mathcal{S}^n$ are determined by the answers received from the oracle to previous questions, executing a query strategy $\mathcal{S}^n$ in a world $w \in W$ delivers a *question-answer set* $\mathcal{S}^n(w)$ of n question-answer pairs. For instance, with $w_1, w_2 \in W$ such that $p_1, p_2 \notin w_1, p_3 \in w_1$ and $p_1 \notin w_2, p_2 \in w_2$ and with $\mathcal{S}^2$ as in Example 4, we have that $\mathcal{S}^2(w_1) = \{((p_1 \vee p_2), (0,1)), (p_3, (1,0))\}$ and that $\mathcal{S}^2(w_2) = \{((p_1 \vee p_2), (1,0)), (p_1, (0,1))\}$. With $\mathcal{S}^n$ an

arbitrary query strategy and with $w \in W$, we have that $\mathcal{S}^n(w) = \{(\sigma_1, \mathcal{A}_w(\sigma_1)), \ldots, (\sigma_n, \mathcal{A}_w(\sigma_n))\}$. The knowledge update due to $\mathcal{S}^n(w)$, $\mathbf{U}(\mathcal{S}^n(w)) \subseteq Sen(L)$, is then defined as follows:

$$\mathbf{U}(\mathcal{S}^n(w)) = \{\mathcal{U}(\sigma_1, \mathcal{A}_w(\sigma_1)), \ldots, \mathcal{U}(\sigma_n, \mathcal{A}_w(\sigma_n))\}$$

A query strategy $\mathcal{S}^n$ *solves* a query structure $\mathbf{Q}$ just in case, in each world $w \in W$, executing $\mathcal{S}^n$ in w allows the agent to decide the target set $\mathcal{T}$. Or, more precisely:

Definition 5. Solving a query structure
Let $\mathcal{L}$ be a referential T-enrichment, let $\mathbf{Q} = \langle \mathcal{B}, \mathcal{T} \rangle$ be a query structure and let $\mathcal{S}^n$ be a query strategy of n questions in $\mathcal{L}$. We say that $\mathcal{S}^n$ *solves* $\mathbf{Q}$ *in* $\mathcal{L}$ just in case[4] we have that, for each $w \in W$ and for each $\sigma \in \mathcal{T}$:

$$\mathcal{B} \cup \mathbf{U}(\mathcal{S}^n(w)) \vdash \sigma \text{ or } \mathcal{B} \cup \mathbf{U}(\mathcal{S}^n(w)) \vdash \neg\sigma$$

Definition 6. Query complexity, classical and magical
Let $\mathcal{L}$ be a referential T-enrichment, let $\mathbf{Q} = \langle \mathcal{B}, \mathcal{T} \rangle$ be a query structure and let $\mathcal{S}^n$ be a query strategy of n questions in $\mathcal{L}$. If $\mathcal{S}^n$ solves $\mathbf{Q}$ in $\mathcal{L}$ and if for no $1 \leq m < n$ there exists a query strategy $\mathcal{S}^m$ in $\mathcal{L}$ which also solves $\mathbf{Q}$ in $\mathcal{L}$, we say that *the query complexity of* $\mathbf{Q}$ *in* $\mathcal{L}$ *is* n. The *classical query complexity* of $\mathbf{Q}$ is its query complexity in L, while its *magical query complexity*[5] is its query complexity in $\mathcal{L}^{\mathsf{basic}}$.

We will now establish that the magical complexity for some query structures is strictly less than their classical complexity. We introduce the following abbreviations.

$$\theta_1 := p_1 \wedge \neg p_2 \wedge \neg p_3 \wedge \neg p_4, \qquad \theta_2 := p_2 \wedge \neg p_1 \wedge \neg p_3 \wedge \neg p_4,$$
$$\theta_3 := p_3 \wedge \neg p_2 \wedge \neg p_1 \wedge \neg p_4, \qquad \theta_4 := p_4 \wedge \neg p_2 \wedge \neg p_3 \wedge \neg p_1.$$

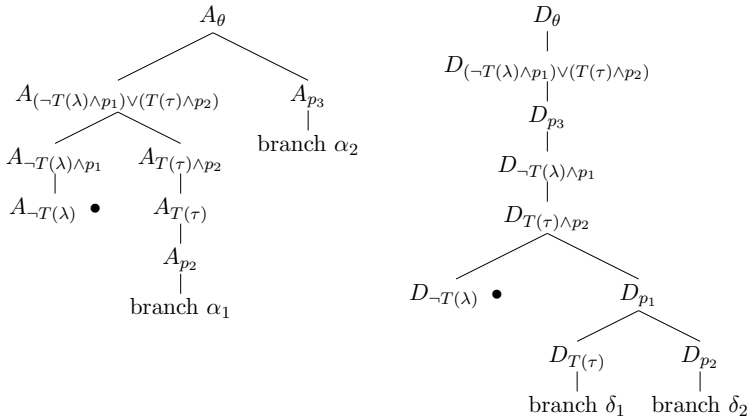
[4]Note that, when w is a world whose union with either $\mathcal{B}$ or $\mathcal{T}$ is an inconsistent set of sentences, $\mathcal{S}^n$ trivially decides $\mathcal{T}$ *in that world* w.

[5]The term 'magical' is used in honor of the logician and *magician* Raymond Smullyan, whose well-known "knight-knave riddles" formed the inspiration for this paper and whose work on analytic tableaux was used to formalize the inspiration.

The query structure *1-out-of-4* $= \langle \mathcal{B}, \mathcal{T} \rangle$ is specified as follows. $\mathcal{B} = \{\theta_1 \vee \theta_2 \vee \theta_3 \vee \theta_4\}$ and $\mathcal{T} = \{p_1, p_2, p_3, p_4\}$. Thus, the agent's background knowledge is such that he knows that exactly 1 out of 4 given sentences is true, although he does not have a further clue as to which sentence it is. The target of the agent is find out which of the four given sentences is true. The query strategy $\mathcal{S}^2$ of Example 4 solves *1-out-of-4* in 2 questions of L. Clearly, there does not exist a strategy in L that solves *1-out-of-4* in 1 question and so the classical query complexity of *1-out-of-4* is 2. The magical query complexity of *1-out-of-4* is 1, as is illustrated by the strategy $\mathcal{S}^1$ consisting of the sole question θ.

$$\theta := ((\neg T(\lambda) \wedge p_1) \vee (T(\tau) \wedge p_2)) \vee p_3$$

In order to illustrate that asking θ to the oracle allows the agent to decide $\mathcal{T}$, we display $\mathfrak{T}_A^\theta$ and $\mathfrak{T}_D^\theta$.[6]



The branches α_1, α_2 and δ_1, δ_2 are the (only) a posteriori branches of $\mathfrak{T}_A^\theta$ and $\mathfrak{T}_D^\theta$. In *1-out-of-4*, the agent's background knowledge is such that he knows that the actual world is represented by either θ_1, θ_2, θ_3 or θ_4 i.e. he knows that the actual world is such that exactly one of those four sentences is true. The following table states, for each "possible world" θ_i, the answer of the oracle to θ in that world and the associated knowledge update due to that answer.

[6] To save some space, we display in fact abbreviations of those trees; a bullet (•) indicates that we do not work out the the official steps after this point, as it is clear that the resulting branch(es) are closed a priori.

world	$\mathcal{A}_{(.)}(\theta)$	$\mathcal{U}(\theta, \mathcal{A}_{(.)}(\theta))$
θ_1	$(0,0)$	$(\neg p_2 \wedge \neg p_3) \wedge (p_1 \vee p_2 \vee p_3) \wedge (p_1 \vee p_3)$
θ_2	$(1,1)$	$(p_2 \vee p_3) \wedge ((\neg p_1 \wedge \neg p_2 \wedge \neg p_3) \vee (\neg p_1 \wedge \neg p_3))$
θ_3	$(1,0)$	$(p_2 \vee p_3) \wedge (p_1 \vee p_2 \vee p_3) \wedge (p_1 \vee p_3)$
θ_4	$(0,1)$	$(\neg p_2 \wedge \neg p_3) \wedge ((\neg p_1 \wedge \neg p_2 \wedge \neg p_3) \vee (\neg p_1 \wedge \neg p_3))$

It is easy to verify that the agent, when he receives an answer to θ and uses the associated information update as specified by $\mathcal{U}$ to update his background knowledge $\mathcal{B}$, is able to decide $\mathcal{T}$, which establishes that the magical query complexity of *1-out-of-4* is 1.

The techniques used in this paper are part of a framework for self-referential truth, developed by the author, which is called *assertoric semantics*. For a more detailed presentation, using the framework of assertoric semantics, as well as a for discussion of the philosophical consequences of this paper's result, the reader is referred to Wintein (2009).

References

Goldstein, L. and Blum, A. (2008). When is a statement not a statement? when it's a liar. *The Reasoner*, 2:4–6.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophical Logic*, 72:690–716.

Muskens, R. (1999). On partial and paraconsistent logics. *Notre Dame Journal of Formal Logic*, 40:352–373.

Smullyan, R. (1995). *First-order Logic*. Dover, New York.

Thalos, M. (2005). From paradox to judgement: towards a metaphysics of expression. *Australian Journal of Logic*, 3:76–107.

Wintein, S. (2009). Assertoric semantics and the computational power of self-referential truth. Submitted.

Wittgenstein, L. (1939). *Lectures on the Foundations of Mathematics, Cambridge 1939*. Hassocks, Harvestor, ed. Cora Diamond. Edited by Cora Diamond.