

Reconciling NSM and Formal Semantics*

Avery Andrews
ANU, August 2014
comments welcome.

Anna Wierzbicka's 'Natural Semantic Metalanguage' (NSM), and 'Formal Semantics' as pioneered by Richard Montague and pursued since that time as the majority approach to semantics in generative grammar, are usually seen as mutually exclusive research programs. Here, I will argue that they can and should be regarded as complementary to at least some extent, although this might require some rethinking of foundational assumptions. There are at least two major foundational issues, one being the role of mathematical formulation, the other the adoption of a 'realist' as opposed to 'antirealist' view of meaning.

Mathematical formulation is fundamental to formal semantics, but traditionally avoided in NSM. So one of the goals of this paper is to show how one can begin to apply some of the mathematical methods of formal semantics to NSM (emphasis on 'begin'). This enterprise should be of some interest to formal semanticists for at least two reasons, one 'good news', the other 'bad news'. The bad news reason is that since NSMs are subsets of natural languages, with restricted vocabulary and syntax, it follows that if it is impossible to do formal semantics for a proposed NSM for a language, it is impossible to do formal semantics at all for that language (and, very likely, any natural language, to the extent that the universality claims of NSM work out). The good news reason on the other hand hinges on the small size combined with universality of NSM: because NSMs are small, they are far more feasible as targets for formalization than entire languages, and because they are supposed to be constructed out of ('tectogrammatically') universal elements, with serious investigation of their universality, formalizing one of them would be a big step towards formalizing all of them, and making a case that natural language semantics is in general capable of formalization.

Why workers in NSM should be interested in formalization is less easy to explain (in terms of their point of view), but I suggest the following: a convincing demonstration that natural language semantics either can or cannot be formalized (in a sense we will be presenting as the paper proceeds) would be a very important result of great cultural significance, indicating for example whether or not the project of creating human-like Artificial Intelligence is possible. Especially interesting would be the negative result that NL semantics cannot be formalized. But one cannot demonstrate something of this nature with programmatic announcements and the exhibition of a few

* I would like to acknowledge helpful comments from Maïa Ponsnonnet and Zhengdao Ye. But they are not responsible for my errors.

sample difficulties. The failure of serious attempts with apparently reasonable prospects for success is required. The small size of NSMs makes them ideal laboratories for the investigating of this question.

Mathematical formalization is one major issue in reconciling NSM and formal semantics, which people promoting NSM need to provisionally accept; another, realism vs. anti-realism, is rhetorically prominent, but, I claim, essentially a distraction that can and should be set aside, at least for the present. Semantic realism claims that the central task of semantics is to connect language to the world, especially in terms of our ability to assess the truth of statements (see Lewis (1972) and Cresswell (1978) for classic developments of this position). Antirealism on the other hand claims that this project is impossible and senseless as a goal for semantics, which should instead focus on intuitively perceivable properties of and relations between sentences, such as anomaly and entailment (somewhat comparable to ‘grammaticality’ in syntax, although with significant differences in detail). This is the standard position of NSM, and in the generative world, received a comprehensive and influential formulation in chapter 1 of Katz (1972). Ray Jackendoff is perhaps the most prominent generative advocate of this position today. Larson and Segal (1995) proposed calling these ‘Logico-Semantic Properties and Relations’ (Katz seems to have provided no generic name for them). I will call them ‘Meaning-Based Properties and Relations’ (MBPRs) in an attempt to avoid complex technical terms when possible. Of the MBPRs, entailment and the closely related consistency and contradiction (all truth-based) are the most closely studied, with many others, such as presupposition and ‘appropriate answer to a question’ getting significant attention.¹

There are, I think, some extremely persuasive arguments that, philosophically, realism is correct. For example, as Frank Jackson occasionally observes (ANU philosophy discussions), we can talk about where to go for drinks after the seminar, go our separate ways, and reconvene 40 minutes later at the same pub. If language did not have a somewhat stable connection to the world, it would be impossible to use it in this and other ways so as to increase the probability of pleasant experiences and reduce that of unpleasant ones.

Nevertheless, there is a problem with realism for empirical semantics, which is that linguists (and philosophers) are for the most part in a very poor position to get useful information about language-world connections, beyond banal observations about how people deal with assertions such as ‘there isn’t any pastrami in the refrigerator’, and to launch a full-scale investigation of these connections does not seem to be very a sensible proposal. There are occasional exceptions, such as the work on the meaning of *most* presented in

¹Katz himself had little to say about entailment as such, focussing on ‘analyticity’, without saying anything that people appear to have managed to understand as to how this was supposed to be different from entailment.

Pietroski et al. (2009), but, for the most part, it would be absurd to attempt to study semantics in a fully realistic manner inspired, perhaps, by the study of bee-communication, by recording the philosophers' post-seminar discussion and then accessing NSA/FSB/ASIO phone surveillance records to see where they went afterwards. Rather, the overwhelming majority of actual evidence used in the investigation of semantics are the MBPRs, the same kind of evidence that antirealism claims are the proper subject matter of semantics. Exactly as Chomsky states with respect to grammaticality intuitions, this kind of evidence is rich, plentiful, and easy to access. We should therefore use it, rather than attempting to base semantics in a comprehensive and practical way on language-world connections, which are clearly mediated by complex cognitive abilities that linguists are not equipped to investigate on their own.

What I propose is that although in some philosophical sense, realism is essentially correct, for the empirical investigation of semantics this can for the most part at present be ignored, to focus on providing the best possible account of the MBPRs between utterances, sometimes accompanied by descriptions of situations constituting the context. These are accepted *de facto* as the main targets of attempted explanation by both realist and antirealist semanticists. However, 'for the most part' does not mean entirely, there are certainly things that can be learned by considering how (some) sentences are judged as true or false in certain kinds of (typically finite) situations. But this does not appear to be viable as the main focus of semantic research now.

The issue of realism feeds directly into the status of model theory, the premier tool of formal semantics, which is frequently justified by its claimed capacity to explain the relationship of language to the world. I will argue that even if it cannot to this, it can nevertheless serve as a 'context of inspiration', whereby informal reflection on aspects of everyday language use can provide motivation for features of mathematical theories of entailment, as well as the occasional enlightening experiment.

In the first section, I will discuss formal semantics, with special focus on certain features of model theory. The most important point will be that regardless of what one thinks of the project of relating language to the world, model theory is an extremely useful tool for investigating the MBPRs. Particularly relevant to NSM is the program of 'algebraic semantics', in which one uses model theory as the basic framework for formalizing entailment, but superposes upon it additional principles stated as laws governing the deductions held valid over certain symbols, such as, for example, the transitivity and asymmetry of 'bigger than'. This amounts to a way of 'having your cake and eating it too', which is useful for NSM because a significant number of primes appear to obey well-studied laws.

Then, in the second section, I will explore some ways in which formal semantics, especially model theory, can be applied to NSM, considering first

the rather surprising phenomena of truth-functionality, extensionality, and intensionality (as opposed to ‘hyperintensionality’, which I suggest is in fact the default expectation), and then surveying a range of entailment laws that seem to be obeyed by various NSM primes. This is only a rather small start on linking up these two research programs, but it is enough to reveal that it is not a completely impossible thing to work on.

1 Formal Semantics

We begin with a brief and incomplete sketch of model theoretic semantics (MTS), whose purpose is to show how it provides a mathematical theory of MBPRs independently of any claims about realism in semantics. Then we consider the deductive alternative, and discuss reasons why it has appeared to ‘fail to thrive’ in spite of being in many ways more consistent with the usual assumptions of generative grammar (and, also, NSM). Then, after returning to MTS to discuss its benefits in comparison to deductive approaches, we consider the further topics of algebraic semantics, and a notion of ‘support’, which can be regarded as an attenuated version of realism. The first two sections can be skipped or skimmed by readers with a basic knowledge of formal semantics. Much of the content of the later sections will also be familiar to such readers, but it is intended to indicate what kinds of adjustments might be useful for better integration of formal semantics with NSM.

1.1 Model Theory for Anti-Realists

A simple MTS for a fragment of a natural language can be constructed like this. We assume that our formal fragment contains n -place predicates for various integers n , together with some ‘logical words’ such as *and*, *not*, and *if*, (and, for a more advanced fragment than we will attempt here, *some* and *every*), and some proper names such as *John* and *Mary*. An ‘interpretation’ I for the language is then constituted by the following:

- (1) a. a set A to serve as a ‘universe of discourse’ (collection of entities)
- b. a member of A to serve as ‘extension’ of each proper name (in more advanced systems with quantifiers, variables will also range over the members of A).
- c. For each n -ary predicate, an n -ary relation, that is, a set of n -tuples on A , to serve its extension.

So, if our universe of discourse is $\{\mathbf{j}, \mathbf{b}, \mathbf{m}, \mathbf{s}\}$, and the extension of the two-place predicate *Love* is $\{\langle \mathbf{m}, \mathbf{j} \rangle, \langle \mathbf{j}, \mathbf{m} \rangle\}$, we can take this as representing

a situation where two people love each other, but a third is not involved, happily for all, presumably.

We now define ‘truth relative to an interpretation’ inductively, by extending (1) to an assignment of truth-values to the formulas constructed from the basic symbols of the language, together with further rules for the logical words. For the n -place predicates, we have:

- (2) If P is an n -place predicate, and $F = P(t_1, \dots, t_n)$ is a formula, then $I(F) = 1$ if $\langle I(t_1), \dots, I(t_n) \rangle \in I(P)$, otherwise $I(F) = 0$.

For example, if $I(\text{John}) = \mathbf{j}$, $I(\text{Mary}) = \mathbf{m}$, and $I(\text{Love}) = \{\langle \mathbf{j}, \mathbf{m} \rangle\}$, then $I(\text{Love}(\mathbf{j}, \mathbf{m})) = 1$, $I(\text{Love}(\mathbf{m}, \mathbf{j})) = 0$ (romantic purgatory).

According to this technical account, for 1-place predicates such as *Boy* and *Girl*, the interpretations will be sets of 1-tuples of elements of A , e.g. $I(\text{Boy}) = \{\langle \mathbf{j} \rangle, \langle \mathbf{b} \rangle\}$, $I(\text{Girl}) = \{\langle \mathbf{m} \rangle, \langle \mathbf{s} \rangle\}$, which are effectively equivalent to subsets of A , as usually presented in introductory logic textbooks.

For the logical words *and* and *not*, we assume syntax rules that combine *not* with one formula and *and* with two, putting parentheses around the result in the latter case to forestall ambiguity, with these rules:

- (3) a. If $F = \text{not}G$ is a formula, then $I(F) = 1$ if $I(G) = 0$, $I(F) = 0$ otherwise.
 b. If $F = (G \text{ and } H)$ is a formula, then $I(F) = 1$ if $I(G) = 1$ and $I(H) = 1$, otherwise $I(F) = 0$.

The standard logical word *or* on the other hand, is not accepted as a prime (and can be either defined or explicated in various ways). The case of *if* is more complex, and will be deferred.

These rules work by climbing up the syntax tree, assigning values to bigger constituents on the basis of those of the smaller ones, starting with the n -place predicates. So for example what values do these sentences get under the ‘romantic purgatory’ interpretation above?

- (4) a. $\text{Love}(\text{Mary}, \text{John}) \text{ and } \text{Love}(\text{John}, \text{Mary})$
 b. $\text{not } \text{Love}(\text{Mary}, \text{John}) \text{ and } \text{Love}(\text{John}, \text{Mary})$

In spite of its incompleteness (no account of quantifiers being why it falls short of constituting first order logic), this example demonstrates important features of how MTS works, but also a fundamental problem for its justification in much of the formal semantics literature.

MTS is supposed to help explain truth by showing how the truth of sentences is determined by how the world is, but, arguably, as discussed by LePore (1983), this is not happening here. Rather, we are connecting truth

and falsity of sentences to certain mathematical structures. To connect the sentences to the world, we would have to move on to connect the mathematical structures to the world, and there does not appear to be any well worked-out and generally accepted story about how to do this (although we will discuss something relevant that I will call ‘support’).

So, one can say, however persuasive the philosophical position of realist semantics might be, model theoretic semantics of natural language does not fully implement it. That is, it is in the end unexplained by formal semantics how a sentence such as *Everybody went to Mooseheads* is false if the people at the philosophy seminar actually all went to The Wig and Pen. This might appear at first to be a fatal objection, but if we pursue things a bit further we find that it isn’t, because many useful things survive it.

This is because what is typically done with a model theory, once it is set up, is to give a characterization of at least entailment, and perhaps some of the other MBPRs. A classic example of entailment would be the following syllogism where Σ , the ‘premises’ of the syllogism are above the horizontal line, S , its ‘conclusion’, below it:

$$(5) \quad \begin{array}{l} \text{Socrates is a man} \\ \text{All men are mortal} \\ \hline \therefore \text{Socrates is mortal} \end{array}$$

Generalizing, we can define entailment conceptually like this:

- (6) A set of sentences Σ entails a sentence S if any person who accepts the sentences of Σ as true ‘must’ also accept that S is true.

The exact interpretation of *must* here is something that one can debate, but I contend that it cashes out to a conviction that if somebody accepts the sentences of Σ but not S , there is absolutely no point in continuing a discussion with them. They can be written off as hopeless conversational partners. Note furthermore that one can recognize specific instances of entailment without having any kind of theory of how they arise (which is exactly our position with (5) above, since we have not suggested any theory of *all*).

In MTS, one provides a mathematical characterization of entailment as follows:

- (7) A set of sentences Σ entails a sentence S if every interpretation that makes all the sentences of Σ true also makes S true.

The description of (7) as a ‘characterization’ rather than a ‘definition’ is deliberate; the definition as given in (6) is supposed to tell us what kinds of things we ought to accept as proposed accounts of entailment, whereas what is described in (7) is a general format for producing mathematical models

of such accounts using MTS (there is another technique, inference rules, as we'll discuss soon).

An essential point is that the truth or falsity w.r.t. any specific interpretation becomes irrelevant here; what matters is the pattern of truth-assignments across all of the sentences of the language. The problem of connecting any particular assignment to the details of how things are in the real world then becomes something which we don't have to concern ourselves with (but can if we find an interesting aspect of it to study).

That MTS provides a way of characterizing one MBPR does not of course mean that it is the best way of characterizing that MBPR or any of the others. To judge its relative merits, we have to look at the alternatives, which, I contend, means deductive a.k.a. proof-theoretical formulations (Davidsonian 'truth conditional' semantics as developed in detail in Larson and Segal (1995) might be seen as an alternative, but Larson and Segal in fact endorse the use of a deductive account of the MBPRs, but don't make any specific proposal). Nevertheless, it does show that we can use MTS while being agnostic or hostile to realism: a substantial piece of MTS-based work that takes an agnostic point of view is Keenan and Faltz (1985).

1.2 The Deductive Alternative

In linguistics as well as in logic, where the study of these questions originated, the original accounts of what I am calling MBPRs were not formulated using anything like model theory, but rather with what can be reasonably called principles of deduction, whereby the relationships between sentences are described in terms of their forms. For example, glossing over many details, a basic start on syllogisms can be made by proposing that any argument of the following form is 'valid', that is, if we accept the premises above the line, we 'must' accept the conclusion below it:

$$(8) \quad \frac{\begin{array}{l} \text{Every } P \text{ is } Q \\ e \text{ is } P \end{array}}{\therefore e \text{ is } Q.}$$

Introductory logic textbooks present deductive systems that you can do useful things with, and Kneale and Kneale (1962) is the standard source for the development of the subject from classical antiquity to the mid twentieth century, shortly before the beginnings of formal semantics.

In contemporary linguistics, this general form of approach (with completely different and highly unsatisfactory details) was introduced by Katz and Fodor (1964), in their attempt to characterize 'analyticity' in terms of a subset relation between elements in the semantic representation of sentences, and related notions such as 'contradictoriness' along similar lines. Amongst the deficiencies of this formulation was that it had no account of n -place

predicates and their arguments, and therefore was incapable of distinguishing the meanings of *John loves Mary* and *Mary loves John*. Later revisions such as those in Katz (1966) and Katz (1972) remedy this deficiency and are considerably better, but perhaps in part because of the bad impression made by the original version (Partee 2003), they never attained a significant following. Katz (1972:ch 1) also, as we have already discussed, proposed that a major goal of semantics should be accounting for the MBPRs.

Although this goal has been widely accepted, at least implicitly,² neither Katz' technical execution nor any of the others that have appeared in 'representational semantics' have gone very far. Jackendoff's last inference rule was for example proposed in Jackendoff (1983), and similar approaches in linguistics do not appear to have done any better, at least prior to the rise of recent investigations of 'Natural Logic',³ which has significantly different goals than standard formal semantics. Lakoff's 1970 proposal to use standard logical deduction rules applying to generative-semantics 'deep structures' likewise does not appear to have led to substantial results.

The general failure to thrive of all of these approaches is something of a puzzle, because not only were they on the scene prior to MTS-based formal semantics, but they are also more in accord with the basic ideas of generative grammar: a deductive account of entailment does, after all, constitute a theory of how it is in principle possible for a native speaker of a language to have an entailment intuition, namely, by constructing the deduction and interpreting this as a basis for accepting the entailment. A model theoretic account on the other hand can't do this, because the definition involves *all possible interpretations* (which will be infinite in number in all serious cases), which the language user obviously cannot run through in his head. Rather, an MTS account of entailment constitutes what in mathematics would be called a 'nonconstructive' characterization of entailment, a characterization that is in some sense logically or intuitively valid, but doesn't tell you anything directly about how to recognize or construct the things that it characterizes (but whose details can often but do not always provide hints for the construction of such techniques).

Nevertheless, in linguistics, MTS started developing rapidly, and has continued to do so for decades. My suggestion is that the basic reason for this is that proof theory (the currently standard name for the mathematical investigation of deductive systems) is considerably harder than model theory, at least with respect to linguistic applications, and that this difference in difficulty was more extreme in the early days of formal semantics, when proof theory was less developed, and there were far fewer resources available from

²Explicit acknowledgements seem to be rare, absent for example from Chierchia and McConnell-Ginet (2000:ch 1), although many of the same MBPRs are discussed in a similar way.

³As presented for example in <http://nlp.stanford.edu/projects/natlog.shtml>

which to learn anything about it. Linguists trying to make up deductive systems on their own, in particular, were I think almost certain not to get very far. To develop this point, we need to examine in closer detail how such accounts work. We will do this with conventional logical rules of deduction, since these are very well understood.

Deductive accounts of an MBPR such as, for example, entailment, work by formulating the property or relation in question as a relation over syntactic structures. In classic symbolic logic, these structures were ‘formulas’ in invented artificial languages, with no formalized connection to natural language sentences, while in formal semantics, they are linguistic structures that are abstract to at least some extent (involving, at a minimum, some account of constituent structure). In this presentation we will not worry about the details of the formulas/abstract structures.

In contemporary work, a popular notation for expressing entailment relations is the ‘natural deduction tree format’, where the premises are written above a line, and the/a conclusion below. From premises consisting of two clauses combined by *and*, for example, we can derive either of the conjuncts as a conclusion; such a rule is called an ‘elimination’ rule in Natural Deduction:

$$(9) \quad \frac{A \text{ and } B}{A} \text{ and-elim} \qquad \frac{A \text{ and } B}{B} \text{ and-elim}$$

Or, given *A* and *B* as premises, we can produce their conjunction as conclusion, called *and*-introduction:

$$(10) \quad \frac{A \quad B}{A \text{ and } B} \text{ and-intr}$$

Such rules can be applied in combinations to produce additional results; for example we can derive *B and A* from *A and B* like this:

$$(11) \quad \frac{\frac{A \text{ and } B}{B} \text{ and-elim} \quad \frac{A \text{ and } B}{A} \text{ and-elim}}{B \text{ and } A} \text{ and-intr}$$

The possible results get more interesting as we add more rules, so that for example if suitable rules and representations for quantifiers and their associated ‘variables’ are adopted, we can derive results such as the following three-premise argument, which the Ancient Greeks could not accommodate:

$$(12) \quad \begin{array}{l} \text{Every boy loves every girl} \\ \text{John is a boy} \\ \text{Mary is a girl} \\ \hline \therefore \text{John loves Mary} \end{array}$$

So deduction is certainly in general a workable approach to characterizing MBPRs, but suffers from at least two substantial problems (at least for its adoption by linguists), one of which is the requirement for very large numbers of arbitrary decisions, and the other a discouragingly high level of mathematical difficulty.

1.3 Problems with Deduction

The arbitrariness problem consists in the fact that there are a very large number of ways in which deductive systems can be set up. Anybody with a reasonable amount of basic knowledge about logic will be aware of at least the following different major techniques:

- (13) a. Hilbert Systems (the oldest, and most still most powerful, but not very intuitive).
- b. Natural Deduction, in any of a large number of variant formulations (usefully surveyed by Pelletier (1999); the main alternative to the ‘tree-style’ format used above, common in logic textbooks, is ‘Fitch-style’; there is also ‘sequent style’ natural deduction).
- c. Gentzen Sequents (different from sequent-style natural deduction, although both were invented by Gentzen).
- d. Beth Tableaux.

There are presumably additional more obscure options known to specialists.

And within these broad approaches there are many further sub-variants. For example within ND or Gentzen Sequents, beyond simple presentational variants, there is a more substantive choice between ‘additive’ and ‘multiplicative’ formulations of many of the connectives, such as the conjunctions (the difference comes out when some of the ‘structural rules’ are removed, as discussed in the literature on ‘substructural logic’). Nowadays, a great deal is known about the equivalences and other relationships between all these systems, but it still takes some time to learn a reasonable amount of this stuff, and it would have taken much longer in the 1970s, when formal semantics was taking shape, because there was much less in the way of accessible textbooks and tutorial papers relating the variants (such as, for example Restall (2000)). Being confronted at the outset with a blizzard of arbitrary-looking decisions is particularly discouraging for generative grammarians, who are trained to try to find some kind of empirical basis for all of the formula-tional decisions which they make for themselves (as opposed to inheriting from some previous authority such as Chomsky or Montague).

The other problem, difficulty, compounds with arbitrariness. In order to have empirical content, a theory of entailment has to characterize some

collections of premises as entailing some possible conclusions, others as not. In a deductive approach, the former is relatively easy: to show that some premises entail a conclusion, produce a derivation of the former from the latter (actually, in fact, what is easy is to check a putative fulfillment of this goal; finding a proof can in fact be rather hard). But the latter requires showing that there is *no* proof of the conclusion from the premises, which can be difficult (especially for linguists without a lot of mathematical aptitude and training).

This problem is exacerbated by the fact that linguists can't confine themselves to the somewhat limited range of deductive patterns employed in mathematical discussion, but must consider a much wider range of naturally occurring ones, such as that *John and Bill each carried a piano downstairs* entails both *John carried a piano downstairs*, and *Bill carried a piano downstairs* but *John and Bill carried a piano downstairs (together)*. does not. How can you be sure that the new deductive rule you're contemplating won't cause your system to melt down in contradictions? MTS seems on the basis of experience to ameliorate the problem of arbitrariness ("there are different views about how to represent certain things, but people tend to agree on what kinds of facts might settle them" – Barbara Partee, p.c., not verbatim), and provide a real solution to the problem of accidental introduction of contradictions, since it requires an impressive degree of ineptitude to accidentally produce an interpretation-extending scheme that assigns both 1 and 0 to some formula.⁴

1.4 Advantages of MTS as a theory of MBPRs

As noted above, one of the things we can do with an MTS account is formulate a mathematical theory of entailment. Superficially, doing this might seem more convoluted than pursuing a deductive approach, but it turns out to have significant advantages. Most importantly, it is often easy to prove non-entailment, by constructing an interpretation where the premises come out true, but the conclusion false. For a ridiculously simple example, assuming the usual truth-table semantics for *or* (part of MTS) we can easily demonstrate that the following entailment does not hold:

$$(14) \frac{A \text{ or } B}{A} \text{ or-elim}$$

What makes MTS 'safe' in this respects is that it is typically relatively easy to calculate whether a sentence is true, false (or has some other relevant

⁴There is a subject of 'paraconsistent logic' where one does this on purpose, but that is not a topic for beginners (the goal is to construct logics which can tolerate contradictions, for use in situations such as database management where they will inevitably arise in the data, and one doesn't want the entire system to become useless as a result).

semantic value) for a small interpretation, so that we can check whether such an interpretation makes the premises of a deduction true but not the conclusion: if we find one, the putative entailment does not hold (which is what we want for (13)).

Proving that entailments hold on the other hand can be a bit more difficult, since it involves mathematical arguments to the effect that any interpretation that makes the premises true makes the conclusion true as well, but these are often at the relatively modest level of difficulty of sophomore abstract algebra proofs, which people capable of getting through graduate programs in linguistics can attain. A potential issue is that they often use natural language counterparts of the symbols and formation rules used in the proof, so have an appearance of circularity. We might, for example, justify the ‘soundness’ (the property of always leading from true premises to a true conclusion, relative to some choice of MTS) of the *and*-intr rule as follows:

- (15) Suppose the two premises A and B are true under I (ie I evaluates them as 1). Then I will evaluate their conjunction A and B as true. Therefore, since this will hold for any choice of I , *and*-intr is sound.

I think a close investigation of what’s really going on in proofs of this nature might be rewarding, but to begin with, one can regard them as a strategy for justifying proposed rules in a format that has been subject to very heavy testing in the last century or so of development of mathematics (when contemporary techniques for proof were developed and stabilized).

More benefits can be derived if a collection of deduction rules can be found that are provably ‘complete’ as well as sound; completeness means that every model-theoretically valid entailment also has a deductive proof. Then, one can use whichever technique, deductive or MTS, is easiest or otherwise most suitable for the task at hand. In particular, the existence of a model-theoretic argument for an inference pattern proves that it has a formal deductive proof, even if nobody has managed to find one yet; apparently this is sometimes useful (Wolfgang Schwartz, p.c.).

The amelioration of the arbitrariness problem by MTS on the other hand is more ‘empirical’. In principle, it seems like there should be a stupendous variety of ways in which MTS interpretations can be set up, resulting in the same kind of bewildering variety of options that afflicts the deductive approach, but in fact, people only seem inclined to worry about a manageably small number. The fact that there isn’t a clear story about why this is the case should not inhibit us from benefiting from it.

1.5 Algebraic Semantics

MTS and deduction can function individually as alternative forms of account of MBPRs, but also jointly, in what has come to be called ‘algebraic

semantics’, pioneered by Godehard Link, Erhard Hinrichs and others.⁵ In algebraic semantics, one starts with an MTS framework, and adds additional constraints that all interpretations must obey, often ones that have been extensively studied in branches of algebra such as especially the theory of lattices and ordering. This allows the investigator to have their cake and eat it too, in the sense that they can freely propose deductive laws that appear to be true, using the MTS infrastructure to remain sure that the system as a whole is not contradictory (if it is, no sentence will have a model that satisfies it, so producing a single one that does is sufficient to demonstrate consistency).

We can for example propose that the *some of* relationship between amounts of stuff (the extensions of count and mass nouns) obeys laws such as these:

- (16) a. Transitivity: If x is some of y and y is some of z , then x is some of z .
- b. Asymmetry If x is some of y , then y is not some of x (implying that nothing is some of itself, which seems empirically correct for the English expression *some of*).

These are the laws for a ‘strict partial order’; quite a lot is known about their consequences, and those of their variants, as part of lattice theory.

In the second section of this paper, we will discuss a number of examples of principles of this nature that appear to apply to some of the Wierzbickian primes.

1.6 ‘Support’

The final useful feature of MTS that I will discuss here is its relation to the notion of ‘support’ of natural language sentences by various kinds of collections of sensory information, and the related concept of extensionality. In textbooks, the former issue is often discussed in a somewhat limited way in connection with for example Cresswell’s 1978 picture of a door with respect to establishing the truth or falsity of sentences such as *The door is open* and *The door is closed*. For the realist, this is a simple example of the real-world significance of language; for the anti-realist, less so, because the response will in general be to a limited amount of information served up by the senses and other cognitive abilities in what might be reasonably regarded as a ‘language of thought’. One might therefore treat them as a form of entailment, but since these phenomena do not involve relationships between actual, utterable, sentences, I think it is better to treat them separately, calling this second notion ‘support’.

⁵Link (1998) contains many examples; Landman (1991) provides a textbook-style presentation.

Support as described here, but not under this name, has been the subject of substantial psychological investigations, such as Johnson-Laird and Byrne (1991, 1993), and, for a recent example, Pietroski et al. (2009), which investigates hypotheses about the semantics of *most* in terms of how sensory information is accessed. Viewed as a theory of support, conventional MTS can be interpreted as the claim that many kinds of propositions such as those involving *and* and *every* are assessed in terms of combinations of multiple pieces of information that can be expressed in terms of atomic formulas. So for example in order to assess *every child is wearing a hat*, we scan the scene before us, and whenever we notice an x such that $Child(x)$ is true, we try to find a y such that both $Hat(y)$ and $Wearing(x, y)$ are true. If all of these searches are successful, we assess the sentence as true, otherwise not (and running the risk of proclaiming a falsehood if we miss a short, hatless child lurking behind a taller one).

One interesting characteristic of support is that all of the standard deduction rules can be motivated by experiences involving support, in that they seem to lack finite counterexamples.⁶ For example, if somebody says in good faith *every crab in the bucket is green*, and you reach in and pull out one that appears to be red, it always appears to be the case that there is some ‘tricky’ feature to the situation. Perhaps there were many crabs in bucket and they didn’t shift them around enough to find the red one hidden under the others; perhaps there is something strange about the conditions (such as UV light and gene-engineered fluorescence), or some other explanation. But something more plausible than the amazing discovery that the rule of universal instantiation is wrong always seems to turn up. We can conclude from this that experience with finite model checking is plausibly the basis of many or even all entailment intuitions, which perhaps seem ‘necessary’ because of the extremely large amount of empirical testing that they have undergone.

Closely related to the kinds of experiences covered by the term ‘support’ are the phenomena of truth-functionality and extensionality. Viewed intuitively, the sentences *John loves Mary* and *Mary does not love John* are two unfortunately connected universes of feeling. But to assess the truth of their conjunction *John loves Mary and Mary does not love John*, we can ignore all of the complexities of the situations supporting the two sentences, and need to consider only the resulting truth-values, computing that of the conjunction by means of the truth-table. Similarly, with *every crab in the bucket is green*, the truth value of the result does not depend at all on the details of the speaker or hearer’s conception of crabs, buckets or the color green, but only on the set theoretical relation between the collection of crabs in the bucket

⁶This is my interpretation of a p.c. from Barbara Partee of a remark made to her by Johann van Benthem.

and the subset of those crabs that are green (for truth, identity is required). In the next section we will see that truth-functionality and extensionality are limited to a very small number of primes. But they are nevertheless remarkable phenomena, worthy of close investigation.

Support is a notion which is not, as far as I can see, a subject that is suitable for systematic empirical investigation by linguists (or philosophers), although useful studies of specific topics can be done with the aid of psychologists. But it can be regarded as providing a ‘context of inspiration’ whereby reflections on the use of language in everyday situations can be used to suggest mathematical ideas which, through the techniques of model theory, can be used to construct rigorous formal accounts of the MPBRs.

1.7 Conclusion

This concludes an account of why antirealists should take MTS seriously even if they don’t accept the philosophical arguments that are usually produced to support it. In the next section, we examine NSM more closely, primarily with the goal of formulating relevant deductive principles. The discussion we’ve just completed then indicates that a model theoretic semantics for NSM is worth thinking about as an aspirational goal, in spite of the programmatic rejection of the goals of MTS that is a salient feature of the current NSM literature.

2 Natural Semantic Metalanguage

The Natural Semantic Metalanguage (NSM) program was initiated by Anna Wierzbicka in the early 1970s (Wierzbicka 1972), and has been under development by her and many colleagues ever since. The collection of NSM publications has become very large; a fundamental one is Wierzbicka (1996), while a recent major one is Goddard and Wierzbicka (2013).

We start with a quick sketch, and a small sample study to give a sense of how it works (and why it is methodologically not as different from formal semantics as one might think), and then proceed to consider a variety of more specific topics where formal semantics has something to say about the behavior of NSM primes.

2.1 Quick Sketch of NSM

The basic idea of the program is in conception very different from formal semantics. It is to state meanings in a universal, intuitively intelligible manner, using a small number of words, and, more recently, constructions, found ‘in some form’ (there are issues here) in all languages. This might seem

completely impossible, and there certainly are problems, but it seems to me that in the mid-1980s, especially with the appearance of *Lexicography and Conceptual Analysis*, the program crossed a threshold in terms of being able to provide much more convincing ‘explications’, the standard term for NSM accounts of meanings, than had previously been attained.

The basic units out of which the explications are constructed are called ‘(semantic) primes’ (originally, ‘primitives’), each of which is supposed to have an ‘exponent’ in every language. So the prime BECAUSE has exponents *because* and *because of* in English. Exponents can be multiword expressions, and primes can share exponents (BECAUSE and AFTER often do in Australian languages, for example). Although the primes are supposed to be universal, their exponents are obviously not, and of course the superficial forms of the grammatical constructions will also differ (so that the universal syntax for NSM is basically a format for ‘tectogrammatical’ structure in the sense that this term is sometimes used in the categorial grammar community). So NSM can be seen as an abstract universal language, essentially a kind of Language of Thought (Fodor 1987), which is furthermore manifested in a concrete subset of every natural language, yielding what can be called ‘the NSM’ of that language.

Although these basic ideas have remained constant, contemporary versions of NSM have significant differences from earlier ones:

- (17) a. The set of primes is now much larger: around 65 instead of the original 13 (at the growth rate of one every year or so, there would appear to at least a century’s worth of headroom before the size of the inventory gets uncomfortably close to 200, which seems to me to be a plausible approximate maximum size for the inventory.⁷)
- b. There is now attention to the task of disciplining the syntax as well as the word-choice in explications, so that extending the approach beyond its original and still primary target of lexical semantic to compositional semantics is no longer impossible in principle (although very underdeveloped in practice), and the account of lexical semantics is constrained by this practice, since proposed explications really are abandoned because they use a syntactic construction without any evident rendering in some language.
- c. There is a concept of ‘semantic molecule’, which is an intermediate-level concept that is first explicated and then used in further explications (it is currently unclear whether they can be eliminated or not; the programming language Prolog provides an example of predicates that can be defined but not eliminated from the places where

⁷The more it grows, the less likely it is that there really is a complete set.

they are used). Molecules arguably first made their appearance in Wierzbicka (1985) as a practical technique to make explications of complex concepts such as *mouse* and *lion* intelligible, and were later elevated to the status of an explicit component of the theory.

- d. In addition to word-explications, NSM is used to formulate ‘cultural scripts’, often involving the GOOD and BAD primes, which supplement many aspects of meaning that can’t be managed in a convincing way with definitions. Such scripts allow us to reconcile the apparent universality of words meaning at least something similar to English *good* and *bad* with the simultaneous existence of enormous differences in the extensions of these terms.⁸

Molecules and semantic scripts are conceptually very important innovations, because their inclusion means that NSM no longer clearly aspires to be able to replace words with eliminative paraphrases stated with the primes, a goal whose feasibility has been widely, and, I think, correctly, questioned.

Instead one may think of it as a program for producing what are in effect teaching materials, so that if you learn the molecules, you will be able to read scripts and explications, thereby learning concepts and behavioral norms implicit in other languages and cultures with greater hope of not mindlessly importing all of the assumptions built into your own. This seems far more feasible than constructing eliminative paraphrases, and is furthermore an enterprise where ‘mere’ progress as opposed to perfection seems likely to be useful, even in real world practice.

2.2 NSM Sampler

To give a bit of the flavor of NSM, I’ll consider some explications for the *have*, *make*, *get to* and *cause to* causative constructions of English. For the first two, I will propose explications of my own, including comparison of my original explication of *make* with that of Wierzbicka (2006:171-178). I will then present and make some remarks about her explication of the remaining ones. Examples of this constructions are:

- (18) a. Mary caused John to arrest Susan
- b. Mary had John arrest Susan
- c. Mary made John arrest Susan
- d. Mary got John to arrest Susan

⁸But there is, evidently, far more cross-cultural agreement about what feels good and bad, than about what is.

I propose that *cause*, not discussed by Wierzbicka, has the simplest explication:

(19) X caused (Y to W) [= Z]

X did something

Because of this, after this, Z

Here we follow the standard NSM practice of using upper case letters to represent ‘variables’ in the explications for which material in the syntactic context will be supplemented, and we also make no serious attempt to deal properly with concepts of tense and aspect, instead using ad-hoc and informal notations. We also use an informal notation to describe the ‘subject to object raising’ construction used here with *cause*, which does not imply Agent status of Y. This is part of the syntax-semantics interface problem for NSM, which so as yet largely unstudied (for a preliminary attempt at describing predicate-argument semantic composition for NSM using glue semantics, see Andrews (2006)).

The prime exponents here are *because*, *do* and *something*. A very wide range of actions on Mary’s part, and kinds of causal connections that might lead to Mary’s arrest are compatible with the explication, understood as a piece of ordinary English. For example Mary, jealous of Susan over, say, Bill, might plant some weed in Susan’s car and deliver an anonymous tip-off to the police, who randomly assign John to investigate and perform any consequent arrests. John need have no awareness that Mary is the instigator of all this, or even that any such person exists. A significant fact is that although this construction can be used in place of all the other ones, this would be rather unnatural; I will claim that this is because of a form of Elsewhere Blocking: the *cause to* construction is ‘dispreferred’, and normally used only (typically in remote/educated/bureaucratic style; the explication might be expanded to convey this) when the conditions of the other constructions are not satisfied.

Moving on to (b), the *have* construction appears to entail that John is in some sense in Mary’s ‘chain of command’. My original explication, constructed without reference to Wierzbicka’s, was (modulo some presentational edits):

- (20) A. X said something to Y because X wanted Y to do Z
B. X did this because X knew that when X says something like this,
Y will think that it would be good if Y does this (Z)
C. Because of this, after this, Y did this (Z)

In the first component, we have exponents of two new primes, SAY and WANT, while in the second, we get KNOW, LIKE, THINK and GOOD. These may be taken as a kind of expansion of the first clause of (19) which

attempts to capture the ‘chain of command’ sense of causative *have*, by entailing that Mary (X) must communicate with John (Y) so that he does something volitionally (if not necessarily fully voluntarily; there can be compulsion involved), and that, furthermore, there is some preexisting basis for Mary to have an expectation that John will do what he wants when she tells him. This does not have to be a formal chain of command with penalties for noncompliance: for example a committee member of my local park-care group might ‘have’ me put up a poster somewhere on the basis that, as a (often rather indolent) member of the group, I in principle support its activities and so will probably fulfil reasonable requests from people who are taking the trouble to be active members of the committee. The final clause is the same as before.

Something that has been completely finessed here is the treatment of pronouns and reference, which I suspect of being a problematic area where some substantial changes to NSM will have to be made.⁹ For example the determination of the antecedence of *do this* is highly context dependent in NSM in complex ways; by NSM principles, no formalization is required, merely lack of ambiguity in practice. A formal approach on the other hand will require a specific disambiguation technique. I will not give some other instances of this below, without attempting to resolve the issue. (One possibility would just be to add indexes and ignore complaints to the effect that they aren’t naturally interpretable.)

Wierzbicka’s explication (2006:176) is:

- (21) Person X had person Y do Z:
- a. X wanted something to happen (to W)
 - b. because of this, X wanted Y to do Z (to W)
 - c. because of this, X said something to someone
 - d. because of this, Y did Z
 - e. X could think about it like this:

when I say something like this (about something like this)
Y can’t say ‘I don’t want to do this’

The new primes here are SOMEONE/PERSON, and NOT. (a-c) can be seen as roughly corresponding to my A. A significant difference is that in Wierzbicka’s discussion, the desire for Z to be accomplished is characterized

⁹NSM normally handles coreference for people by deploying the phrases ‘this person’ and ‘this other person’; but it is unclear to me how this could account for the six readings of *when a bishop introduces a bishop to a bishop, she blesses her*, which could presumably be distinguished in most manual sign languages. Grammatical gender also provides resources for distinguishing participants which will be available to different degrees in the NSMs of different languages, creating similar issues for cross-linguistic transportability of explications.

as an important feature of the meaning of *have*, as opposed to for example that of *make*. So for example (18b) entails that Mary has a genuine interest in Susan being arrested, whereas in (18c), the point might merely be to mess up John by forcing him to be unpleasant to someone he likes (it doesn't matter who). In the explication, this is expressed by a., but not quite correctly, I suggest, since X is described as wanting something to happen to an unspecified W, without saying what (why can't W be the same as Y, and the 'something' be the infliction of anguish upon them?). I therefore suggest replacing a. with:

(21) a'. X wanted someone to do Z

Wierzbicka's d. is on the other hand very close to my C, but her d. notes something that I missed, which is that X can achieve their effect by talking to someone other than the person who ultimately performs Z. So we want to retain this.

Her e. on the other hand is performing roughly the function of my B, but there are a number of tricky issues to deal with. B. is stated as giving a reason for what X does, which is not appropriate once we have accepted that desire for someone to do Z is the reason. On the other hand, e. merely gives a 'way in which X can think' about what's going on, which doesn't seem strong enough. I suggest that revised e. should start like this:

(21) e'. X did it like this because X thought like this:

...

The next problem is to fill in the ellipsis dots. Wierzbicka proposes that Y thinks that they can't say that they won't do it, whereas I proposed that they think it would be good if they did it. It now seems to me, however, that both of these, along with the alternative that they think that it would be bad if they didn't do it, ascribe too much to X's characterization of the state of mind of Y (that is, I can't think of any entailments that distinguish these possibilities). I suggest instead:

(21) e'. X did it like this because X thought like this:

If I do it like this, Y will do Z

This implies nothing more than that the communicative means chosen would have the desired effect, which seems correct. But we encounter the problem of the expression of (future) tense in NSM, for which the current proposal is 'inflectional allomorphy' as described in Goddard and Peeters (2006:22-23). Without going into the details, we can think of *will* here as a convenient abbreviation for more complex expressions involving *after this*.

Putting the pieces together, the proposed revision of (21) is (note the omission of original b., no longer required):

- (22) Person X had person Y do Z:
- a. X wanted someone to do Z
 - c. because of this, X said something to someone
 - d. because of this, Y did Z
 - e. X did it like this because X thought like this:
 If I do it like this, Y will do Z

While this is surely not perfect, the discussion serves illustrate the technique of improving explications by adding and changing clauses in order to correct the entailments of the explication. I will also mention the assumption that there is a single, correct explication for each word in the lexicon of a speech community. It is probably more ‘reasonable’ to assume that there is a possibly rather wide range of variation. But even if this is true in the end, it is also probably a more productive heuristic strategy to assume at the beginning that there is only one, and see if it can be found, and not give up looking for it too easily.

Moving on to the next two verbs, *make* and *get to* do not imply a ‘chain of command’, in that Mary is not said to know that she can cause John to do something simply by telling somebody something. Furthermore, they both either imply or suggest at least some potential reluctance on John’s part to what is requested, which causative *have* does not (he might detest or resent Susan, and be delighted to get to take her in, or be a co-conspirator in a plot with Mary and Susan which requires Susan to be in a prison cell for a while, for some reason). *Make* differs from *get to* in that it implies no choice on John’s part, whereas with *get to*, John’s compliance is in the end a voluntary choice, even if some reluctance might be present.

A significant difference between these two and *cause* is that they seem to require that John perform the caused event as a true, if perhaps reluctant, agent. So the (a) below is consistent with John not knowing that he tripped the alarm, while (b, c) aren’t:

- (23) a. Mary caused John to trip the secret alarm
- b. Mary made John trip the secret alarm
 - c. Mary got John to trip the secret alarm

For example in (a), John might be a kidnapper, who Mary tricks into inadvertently triggering the alarm which will hopefully cause her to be rescued. In (b, c) on the other hand, John might be a co-hostage with Mary who basically thinks it’s too dangerous to trip the alarm, but she induces him to do it anyway (deploying greater coercive powers in (b) than in (c)). *Make* and *get to* also, I believe, involve some extended action.

Wierzbicka’s (2006:181) explication for causative *make* is:

- (24) Person X made person Y do Z =
- a. X wanted Y to do Z
 - b. because of this X said (did) something to Y
 - c. because of this, Y thought like this: I have to do it
 - d. because of this, Y did Z
 - e. Y didn't do it because of anything else

Something to note is that 'have to' is not a prime, but needs to be treated as a compression of something like *if I don't do this, something very bad will happen*. This explication is fairly similar to what I came up with independently. Note that clause (e) encodes the reluctance of Y to perform the action, by indicating that Y had no other reason to do it. Perhaps this is not quite strong enough.

Get to also at least implicates some reluctance, but the sense of coercion implied by *make* is absent. An interesting kind of example is the 'request formula' used for example by hairdressers: '*Can I get you to lean your head back/take your glasses off/...*'. Note that it would be absurd to report your compliance by saying '*the hairdresser got me to take my glasses off*'. I suggest that the wording of the formula softens the request by overstating the extent to which complying with it could be construed as a problem. *Get to* also involves verbal communication, I believe, whereas *make* could involve only a bit of gesturing (with a gun, for example).

Wierzbicka's explication (2006:178) is:

- (25) Person X got person Y to do Z:
- a. X wanted Y to do Z
 - b. X knew that if Y did not want Y to do it, Y would not do it
 - c. X thought about it like this: "If Y wants to do it Y will do it"
 - d. because of this, X did (said) something to Y
 - e. because of this, after this, Y wanted to do Z
 - f. because of this, Y did Z
 - g. because of this, X could think like this:
 "I wanted something to happen
 it happened"

I'd suggest that g. is unnecessary (I can't think of any entailments that it clearly adds). the other components express various aspects and consequences of X's recognition that Y's performance of Z is voluntary, in a way that is not the case with *make* or *cause*, and at least 'more voluntary' than with *have*.

To conclude the discussion of causatives, we repeat the point that *cause* is mostly used in situations where none of the other, semantically much

richer, items in the English causal inventory apply. This tour certainly does not settle all questions about these causative words (and there are others discussed by Wierzbicka that we have not considered here), but is hopefully enough to provide some sense of how NSM works, and explications develop.

2.3 Comparison to Formal Semantics

Although NSM certainly is very different from formal semantics, there is one important commonality: they both depend primarily on the same kind of data, intuitions of entailment. In the case of NSM, these have so far been almost entirely intuitions of the entailments of individual words. When one has proposed a specific explication, one then tests it by thinking up situations, described verbally, and comparing how the word and the explication apply. Discrepancies are taken as evidence that the explication is wrong, and requires adjustment (in practice, this happens all too often).

These discrepancies are mismatches in entailments, such as whether a causative verb or explication entails that the ‘Causee Agent’ does or doesn’t want to perform the Caused Action. However, there is a massive difference in how the entailments are to be explained by the two kinds of theories. Since the formulas/abstract linguistic structures used in formal semantics are not natural language, they can only derive any kind of meaning they might have via formal apparatus associated with them (unless one assumes that, with use, the notations become naturally intelligible in the way that normal language are, which is I think sometimes assumed implicitly, but not systematically justified). This could be taken as at least one aspect of the sense of David Lewis’ (1972:169) famous assertion that ‘semantics with no treatment of truth is not semantics’. Namely, it is not enough just to propose some abstract notations for displaying meanings. And even for anti-realists, the apparently rather central role of truth in the concept of entailment should suggest that a theory of semantics without some kind of story about truth is at best incomplete, and probably rather lame (TRUE is in fact one of the NSM primes).

With NSM explications, the situation is different: since these are in natural language lightly supplemented with a bit of whitespace formatting similar to what is familiar and widely used in placards, menus, billboards and other kinds of public writing display, it is meaningful as is, without requiring any additional formal work (and reference indexes, if they turn out to be required, seem to occur naturally in manual sign languages). If somebody claims that it is in general impossible to assess the entailments of an explication without a formalization, they are thereby claiming that most of the data of natural language semantics, including formal semantics, does not exist, which is absurd.

It is plausible to argue (Maïa Ponsonnet, p.c.) that this difference is

not absolute, but a matter of degree: NSM is somewhat conventionalized, in ways that make it not always entirely straightforward for beginners, and people do seem to find it easy to ‘pick up’ the intuitive meaning of various kinds of logic notations. But nevertheless, I think the difference in degree is rather large: full logical forms are usually far more bewildering to beginners than explications are, and the simple clauses out of which explications are constructed have a very high degree of immediate intuitive intelligibility.

NSM explications thus constitute an intriguing case of an explanatory theory which doesn’t need any kind of mathematical formulation in order to work in a useful way. Which does not, of course, imply that attempts to formalize it wouldn’t lead to interesting results. And in particular, a serious but failed attempt to do formal semantics on NSM would suggest that formal semantics is impossible, and perhaps lead to some insight into why. This is I think quite a sufficient reason to investigate the prospects.

In the sequel, I will consider two aspects of the situation. In the first place, I will review some logical properties of the primes, with attention to truth-functionality, extensionality, and ‘intensionality’ in the sense of applicability of possible worlds semantics (as opposed to ‘hyperintensionality’ in the sense of Pollard (2008), as manifested with some of the mental state primes, such as THINK, SAY and KNOW).

2.4 Truth-Functionality

Sheer truth-functionality is a very limited phenomenon in NSM. There appear to be only three or four words and constructions in the prime set that exhibit it, being TRUE, NOT, and possibly IF...THEN, along with the grammatical construction whereby multiple sentences are combined into an explication or sub-explication, which I call ‘box composition’, which effectively performs the role of logical conjunction.

I will begin with the relatively new prime TRUE, which seems to first appear in 2002, (Wierzbicka 2002, Wierzbicka and Goddard 2002) and still has received relatively little discussion. It appears to be used as a predicate to NP subjects, or as an attributive to (SOME)THING. For example, in (Wierzbicka 2002:411), the explication of *Ivan skazal nepravdu* starts out:

- (26) Ivan said something
He knew that this <something> was not true

The angle brackets indicate that the ‘something’ word here is sometimes included in such explications, sometimes not, for reasons that could use further reflection. Omitting it does not make the explication ungrammatical or unintelligible.

Truth-functionality reveals itself in discourses such as:

(27) Susan understated her tax liability. This is true.

The antecedent of *this* is the first clause, and the second clause is true if and only if the first clause is. Truth functionality likewise obtains if the second clause is changed to *this is not true*, but the discourse becomes contradictory, whereas, in the original version, the first clause seems to function as an initial admission that is expected to lead up to some kind of plea for indulgence, such as *but nobody ever told her that her grandmother had set up a Swiss bank account in her name*.

In English we can also say *it is (not) true that Susan understated her tax liability*, but it is not at all clear that this is a universally available valence of the TRUE prime. A potential explication of this use might be along the following lines:

(28) Some people think like this:

Susan understated her tax liability

This is (not) true

This attempts to capture the idea that the construction is not primarily used to add a new assertion to the common ground, but comment on something that is already in it, at least as a possibility. At any rate, we will not take clausal subjects to be a valence option of the prime TRUE, but something provided for the English word by means of additional explicatory material.

There is on the other hand no prime 'FALSE'; English *false* is not a prime exponent. To see this, consider the oddness in ordinary discourse of something like:

(29) I thought that Cindy was in her office. This was not true/?false

The negation of TRUE is banal here, *false* seems strange. I suggest that this is because the meaning of *false* involve components that refer to rhetorical postures whereby the non-truthfulness of the assertion is claimed to have some kind of wider significance:

(30) I used to think that Poverty of the Stimulus arguments motivated a Morphological Blocking Principle. Now I believe that this is false.

A tentative starter explication of *false* might run along the following lines:

(31) Some people think like this:

X

This is not true

It is bad if people think this way.

It is good if people know this

We don't here account for the fact that the first component appears to be a presupposition rather than an assertion.

We have already seen uses of NOT together with TRUE, and it is clear that its meaning flips truth (not 'truth-values', in NSM *per se*, because MTS is not intrinsically part of NSM, but rather a tool that will might prove useful for investigating it). We can ask if NSM is bivalent in the sense that every sentence is either true or false, and the answer presumably has to be 'no', because of the paradoxes. Interestingly, many of the explications and cultural scripts in Wierzbicka (2002) involve formulations of bivalence as one of their ingredients, indicating that it is not being thought of as a basic features of NSM, but something that needs to be stated explicitly when needed. For example, the full *Ivan skazal nepravda* explication is:

- (32) Ivan said something
He know that this something was not true
 people can say two kinds of things to each other
 things of one kind are true
 it is good if people want to say things of this kind to other people
 Ivan did not say something of this kind
 things of the other kind are not true
 it is not good if people want to say things of this kind to other
 people
 Ivan said something of this other kind

Note that although this explication implies that true and not true are important kinds of things that people can say to each other, it does not entail that there are not others, such as questions, and paradox sentences that can be neither true nor not true, such as *I shave everybody who does not shave themselves*, given plausible ideas about what kinds of entailments hold, and that true and not true are mutually exclusive.

Aside from bivalence, the other logic-based question about negation would be the status of double negation, that is, does NOT NOT S entail S. Due to the restrictions we've imposed on the valence of TRUE, we can't produce something close to a standard double negation formula without introducing extraneous elements, but can contemplate a discourse like this:

- (33) A: John is not washing the dishes
 B: that is not true!

This appears to entail that John is washing the dishes, arguing that double negation is valid for NSM, and therefore for ordinary language.¹⁰

¹⁰And therefore that intuitionistic logic, useful as it is, is not a true representation of ordinary language meaning.

To progress further with formalizing truth-functionality in NSM, it might seem most natural to proceed deductively, but, as discussed in previous section, the model-theoretic approach seems easier. It would appear to at least be convenient (perhaps there are other ways) to have at least two ‘truth-values’, conventionally either $\{1, 0\}$ or $\{\mathbf{T}, \mathbf{F}\}$ (as noted above, the nonexistence of a FALSE prime does not imply that we can’t use a ‘false’ truth-value in an MTS for NSM, since MTS can be regarded as a tool for investigating MBPRs in NSM, rather than an intrinsic part of NSM). So then, clauses in an explication will at least sometimes get evaluated as 1 if they are true, 0 if they are not true (and the logic might be ‘three-valued’ with an ‘unspecified’ value for some combinations, paraconsistent with some sentences being having both \mathbf{T} and \mathbf{F} values, and many other possibilities). We then propagate these values to instances of THIS that are anaphoric to those clauses, propagate them unchanged to predications with TRUE, and flip them if NOT is present. To fill out this sketch as a proper formal theory, it would be necessary to provide an account of anaphora, which will not be attempted here.

The next truth-functional construction is simply the operation whereby one clause is added to one or more others to form a multi-clause explication. Logically, this is simply conjunction (in spite of frequent NSM assertions to the effect that *and* is not the exponent of a primitive), and has the truth-table of logical conjunction. In particular, if an explication is true of a situation, every component in it must be true, and if every component in an explication is true, whatever it is that it explicates must be true, if the explication is correct. In order to proceed deductively here, we would need to figure out some way to deal with the anaphoric references, which seems like it might be challenging, but evaluating anaphors to the reference of their antecedents, which we have already discussed, and treating box-composition as having the truth-conditional semantics of conjunction would be a plausible way to start.

Our remaining at least partly truth-conditional construction is IF . . . THEN. The truth conditional aspect is revealed by the obvious falsity condition for (33), where the *if*-clause (protasis) is true and the main clause (apodosis) false:

(34) If Avery’s door is open, he is in his office.

The problem case is the classical one whereby for the ‘material conditional’, the whole thing is true if the protasis is false. This debate has raged for literally millenia (“Even the crows on the roof caw about the nature of conditionals” - Callimachus); a relatively recent survey is Bennett (2003).¹¹

¹¹With respect to which I note that it would be very interesting to see how many of the complex and problematic scenarios discussed by Bennet work out in the same way for the standardly recognized conditional constructions of other languages, especially non-European ones without significant cultural heritage from Ancient Greece.

We obviously have no chance whatever of settling this issue here, but we can see the problem by contemplating this conversation:

(35) A: If Avery's door is open, he is in his office.

B: ?*Yes indeed, it's closed

B': ?Yes indeed, it's open and he's there

The intuition here is that to 'support' a conditional we need evidence for a general rule that lacks exceptions, for which answer B is completely irrelevant, and B' is insufficient, constituting only a single case.

On the other hand, the idea that the conditional is fully truth-functional is supported by uses like this:

(36) If that guy knows anything about biology, then I'm the Pope [spoken by somebody who's clearly not the Pope, perhaps, Jerry Coyne]

The idea here is that the obvious falseness of the apodosis means that in order for the whole sentence to be true, the protasis must be false, so the net effect is to assert that that guy does not know anything about biology.

A final case that is not quite truth-functional, but relevant anyway, is the WHEN construction. Consider a sentence such as:

(37) When Avery's door is open, he's in a good mood.

A simple MTS account of sentences like this can be given by having the semantic values of sentence be functions from time-periods to truth-values. We evaluate this sentence by looking at all the (relevant) time periods, and checking for each the material conditional 'Avery's door is open \supset Avery is in a good mood'. If there are no time periods for which the conditional is false, and, perhaps at least some where its protasis is true, we judge that the sentence is true (I have a feeling that the sentence is not so clearly true if the door is never open). So although this construction goes beyond mere truth-functionality, requiring quantification over times, it does seem to involve a major truth-functional component.

2.5 Extensionality

Extensionality is in essence truth-functionality extended beyond propositions to things. The clearly extensional words are the 'quantifiers' ONE, TWO, SOME and ALL. The other two quantifiers, MUCH/MANY and LITTLE/FEW, might also well be extensional, but involve complex contextual effects that I do not want to look at here. The basis of extensionality is that many Natural Language nouns, verbs and adjectives, in typical situations, pick out sets or sets of n -tuples of entities that they are 'true of'. So in a

given back-yard scene, there might be entities (things that you can point to saying *this*) that are birds and others that are dogs, and also some that are barking, and also certain ‘ordered pairs’ (two entities, with designated ‘first’ and ‘second’), such that the first is chasing the second, the first is watching the second, etc.

Extensionality refers to the fact that in sentences such as those below, the truth of the whole depends only on the patterns of overlap between the sets of entities for which the properties and relations hold:

- (38) a. Some dog is barking
b. Every dog is barking
c. Every dog is chasing some bird
d. Some bird is watching every dog
e. Two birds are chasing one dog

The richness of whatever the concepts of birdhood, dogginess, barking or chasing are is completely irrelevant; only the overlaps matter.

So for example in the intransitive case of *some* (an elaborated exponent of the SOME prime), the *dog* set must have a nonempty overlap with the *bark* set, while in that for *every*, and elaboration of ALL, the dog set must be a subset of the bark set. The transitive examples are too complex to work through here, but dealt with in introductions to formal semantics or symbolic logic, and it is clearly what *n*-tuples are in the sets of associated with *chase* and *watch* that matters, rather than anything else about their ‘meanings’. Extensionality is thus inherently surprising for the same reason that truth-functionality is. A number of related properties of ‘quantifiers’ were presented in Barwise and Cooper (1981), and have been mainstays of work in formal semantics ever since.

2.6 Intensionality and Hyperintensionality

‘Intensionality’ can be regarded as a weakened form of extensionality, including truth-functionality. To get a sense of what is involved, consider a sentence such as:

- (39) *It is possible that Bond might be in Zurich*

The truth of this is clearly independent of whether Bond actually is in Zurich. Rather, what seems to be involved is the plausibility of Bond’s being in Zurich given other knowledge. This is a delicate matter. With the arrival of the Boeing 707, it became possible for somebody to be almost anywhere in the

world with real roads within 48 hours of their being spotted in any urban center, but, still, a sincere utterance of (39) seems to require a bit more than this level of plausibility; roughly, I think, enough of a probability to be worth deploying resources to cope with Bond's presence in Zurich if he's really there.

The conventional way to deal with this kind of intuition in formal semantics has been 'possible worlds', developing some proposals of Carnap (1947). In contemporary implementations, an interpretation now contains a set of objects called 'possible worlds', alongside the set of entities, and a 'proposition' is a function from possible worlds to truth-values. Sentences are furthermore evaluated 'at' a possible world, so that *Bond is in Zurich* can be evaluated to 1 in some, 0 in others (as well as the same in all), and is viewed as true if it evaluates to 1 in the chosen world we're evaluating at (thought of as representing the 'real world'). 'Possible', an elaborated exponent of the MAYBE prime, is then a 'modal operator' which produces a proposition that evaluates to 1 in the world of evaluation if it evaluates to 1 in one of the 'possible' worlds, perhaps subject to further restrictions such as 'accessibility' (for which, see a textbook on modal logic). We can have multiple sets of possible worlds useful for different sorts of things, so that if (39) appears in a context such as a Bond movie, our 'possible worlds' might be worlds deemed allocating some resources to deal with in case they turn out to be the actual one.

All of this might seem rather pointless until we get to the next grade, hyperintensionality (as this term is used for example by Pollard (2007)). To get a sense of this, consider the case of Fred, who is blind, not well educated, not very smart, and has some other ill-understood cognitive challenges. Fred thinks that he was abducted by aliens last night, and that they have told him that they will destroy the world if he is not holding an equilateral triangle at 7:30AM this morning (too early to recruit a friend in time to help). He is for various reasons unable to perceive equilaterality with his senses, but thinks he remembers that equiangular triangles, which he can detect, are also all equilateral, but is not quite sure.

Now suppose we have independent knowledge of the alien's activities. We can now say either:

- (40) a. Fred must be holding an equilateral triangle at 7:30AM this morning or the world will be destroyed.
- b. Fred must be holding an equiangular triangle at 7:30AM this morning or the world will be destroyed.

(On the assumption that the local geometry of space is close enough to Euclidean that the relevant theorem hold in it, up to the aliens' ability or desire to make the measurements precise.) This is because we know that the

two types of triangle are the same, so that (a) will be true in all the relevant ‘possible worlds’ that (b) is and vice-versa.

By contrast, the following are not equivalent:

(41) a. Fred knows that he is holding an equilateral triangle.

b. Fred knows that he is holding an equiangular triangle.

It seems perfectly possible that (a) could be true and (b) false, the same if we replace *knows* with *thinks*, and even more so if we replace it with *says*. So, these are hyperintensional rather than just intensional. There are numerous theories about how to manage hyperintensionality in formal semantics; we won’t try to go into this here, but merely note that it seems to be a relevant distinction, which effects a classification of the primes.

In particular, it would appear that MAYBE and CAN are clearly intensional, and also, even if IF is not truth-functional, it is at least intensional rather than hyperintensional, even in the counterfactual conditional construction that we haven’t discussed (and won’t):

(42) a. If Fred is holding an equiangular/equilateral triangle at 7:30AM, the world will not be destroyed.

b. If Fred had not been holding an equiangular/equilateral triangle at 7:30AM, the world would have been destroyed.

Likewise, BECAUSE, HAPPEN and DO, illustrated for the first:

(43) Because Fred was holding an equilateral/equiangular triangle, the earth was not destroyed.

On the other hand, THINK, SAY, KNOW and WANT are hyperintensional. Since FEEL has only a nominal, not a propositional valence, its status is currently unclear to me.

So we now have the intensional/hyperintensional classification applying to 10 primes, along with the three clearly truth-functional ones, which is beginning to constitute a noticeable fraction of the prime list. One might still be disdainful of these concepts on the basis that they don’t apply to many primes, but this would be wrong because the real surprise is that they apply to any at all. Meanings are, after all, very complex, and our default expectation should be that the semantic behavior of combinations of primes should depend on all of the semantic properties of the primes that are combined. But this turns out not to be the case: in a significant number of instances, most of the meaning of a prime can be discarded for ascertaining the truth of the combinations that it appears in, a very strange phenomenon indeed. (Conceptually, it’s not so different from the notion of ‘phase’ in current generative syntax.)

The final topic I will look at is ‘laws’, essentially the application of algebraic semantics to NSM.

2.7 Laws

A major role in mathematical reasoning is played by algebraic principles which relationships may or may not obey. The most important ones that people have found it useful to study are:

- (44) a. Transitivity: For all x, y, z , if xRy and yRz , then xRz (e.g. *richer than*)
- b. Reflexivity: For all x , xRx (e.g. *identical to*)
- c. Irreflexivity: For all x , not xRx (e.g. *richer than*)
- d. Symmetry: For all x, y , if xRy , then yRx (e.g. *similar to*)
- e. Asymmetry: For all x, y , if xRy , then not yRx (e.g. *richer than*)
- f. Antisymmetry: For all x, y , if xRy and yRx , then $x = y$ (e.g. *least as much as* for amounts; this law has relatively few convincing NL examples, although it is very useful in mathematics).¹²

Relations obeying combinations of these (and other) principles have various interesting properties (see Partee et al. (1993) for an introduction oriented towards the needs of linguists). So-called ‘equivalence relations’, for example, a feature of introductory abstract algebra, obey Transitivity, Symmetry and Reflexivity, and having the interesting property of splitting up a set exhaustively into non-overlapping subsets called ‘equivalence classes’ (*has the same birthday as* is an example). So that, if *loves* was an equivalence relation (rather than obeying no laws whatsoever), group marriages would probably be more stable than they actually seem to be.

We can then see that many of the relational primes obey various of these laws, although there are also unclear cases, such as whether NEAR is reflexive. Here are some examples:

- (45) a. PART: Asymmetric and Irreflexive, possibly Transitive (not entirely clear)
- b. THE SAME: Reflexive, Symmetric, Transitive (arguably the prototypical equivalence relation; all the others are ‘the same in some respect’)
- c. OTHER: Irreflexive, Symmetric, (not Transitive)
- d. MORE: Transitive, Asymmetric (and therefore Irreflexive)

¹²Note the subtlety that *least as tall as* is not antisymmetric; two people can be at least as tall as each other but still be different people. Their heights however will be the same.

- e. BEFORE: Transitive, Asymmetric (and therefore Irreflexive)
- f. AFTER: Transitive, Asymmetric, (and therefore Irreflexive)
- g. ABOVE: Transitive, Asymmetric, (and therefore Irreflexive)
- h. BELOW: Transitive, Asymmetric, (and therefore Irreflexive)
- i. INSIDE: Transitive, Asymmetric, (and therefore Irreflexive; assuming that something can't be inside itself; if we challenge this, then it becomes Antisymmetric and Reflexive).
- j. NEAR: Symmetric, certainly not Transitive, probably Irreflexive (this tends to produce disagreement in the classroom).
- k. FAR: Symmetric and Irreflexive; neither Transitive nor Reflexive.

Presumably the real reason for the obedience of the primes to the various laws is their connection to the human sensory apparatus; it is also worth noting that the ultimate basis of most of the mathematics is thinking about sensory experience without a very close dependence on the form of ordinary language, whatever else may be going on under the hood. So further investigations of Primes and Laws would very likely be rewarding. And MTS with the 'algebraic semantics add-on' would seem to be a safe and promising way to investigate this.

3 Conclusion & Further Issues

We have seen various ways in which NSM and Formal Semantics are at least somewhat compatible in their aims and methods. Although there are many problems remaining to be solved in order to fit them together as truly complementary methods for investigating the same subject-matter (coreference without variables in multi-participant discourses, and the semantics of color and other unimodal sensory terms, where ostension appears to be required to get any substantive account of meaning, strike me as two particularly pressing ones), the discussion here shows that some progress can be made.

The most important practical ingredient supplied by formal semantics that is lacking in NSM is mathematical formulation; however NSM has its own special feature that we have not given much explicit discussion of, which we can describe as 'immediate intuitive intelligibility'. While entire NSM explications can be difficult to take in, it is important to note that the individual components are not. Somebody who claims to have difficulty in understanding what *this is good* or *somebody did something bad to someone* can I think be dismissed as a writeoff just as surely as someone who rejects a

simple syllogism. It should be interesting to find out whether mathematical development and immediate intuitive intelligibility can be combined into a single theory of meaning in natural language.

References

- Andrews, A. D. 2006. Semantic composition for NSM, using LFG + Glue. In K. Allan (Ed.), *Selected Papers from the 2005 Conference of the Australian Linguistic Society*. <http://www.als.asn.au/proceedings/als2005.html/> (accessed 19 Feb 2010).
- Barwise, J., and R. Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4:159–219.
- Bennett, J. 2003. *A Philosophical Guide to Conditionals*. Oxford University Press.
- Carnap, R. 1947. *Meaning and Necessity*. Chicago, IL: University of Chicago Press.
- Chierchia, G., and S. McConnell-Ginet. 2000. *Meaning and Grammar: An Introduction to Semantics (second edition)*. MIT Press. first edition 1990.
- Cresswell, M. J. 1978. Semantic competence. In F. Guenther and M. Guenther-Reutter (Eds.), *Meaning and Translation: Philosophical and Linguistic Approaches*, 9–27. London: Duckworth.
- Fodor, J. 1987. *Psychosemantics*. MIT Press.
- Goddard, C., and B. Peeters. 2006. The Natural Semantic Metalanguage (NSM) approach. In Peeters (2006).
- Goddard, C., and A. Wierzbicka (Eds.). 2013. *Words and Meanings: Lexical Semantics Across Domains, Languages, and Cultures*. Oxford: Oxford University Press.
- Jackendoff, R. S. 1983. *Semantics and Cognition*. MIT Press.
- Johnson-Laird, P. N., and R. M. J. Byrne. 1993. *précis of deduction*. *Behavioral and Brain Sciences* 16:323–380.
- Johnson-Laird, P. N., and R. M. Byrne. 1991. *Deduction*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Katz, J. J. 1966. *The Philosophy of Language*. New York: Harper and Row.

- Katz, J. J. 1972. *Semantic Theory*. Harper and Row.
- Katz, J. J., and J. Fodor. 1964. The structure of a semantic theory. In *The Structure of Language: Readings in the Philosophy of Language*. Englewood Cliffs NJ: Prentice-Hall. first published in 1963, *Language* 39:170-210.
- Keenan, E. L., and L. M. Faltz. 1985. *Boolean Semantics for Natural Language*. Reidel.
- Kneale, W., and M. Kneale. 1962. *The Development of Logic*. Oxford University Press.
- Lakoff, G. 1970. Linguistics and natural logic. *Studies in Generative Semantics* #1, Ann Arbor MI.
- Landman, F. 1991. *Structures for Semantics*. Kluwer Academic Publishers.
- Larson, R., and G. Segal. 1995. *Knowledge of Meaning*. MIT Press.
- LePore, E. 1983. What model theoretic semantics cannot do. *Synthese* 54:167–187.
- Lewis, D. 1972. General semantics. In D. Davidson and G. Harman (Eds.), *Semantics of Natural Language*, 169–218. Dordrecht: D. Reidel.
- Link, G. 1998. *Algebraic Semantics in Language and Philosophy*. Stanford CA: CSLI Publications.
- Partee, B. H. 2003. Reflections of a formal semanticist. In *Compositionality in Formal Semantics: Selected Papers of Barbara Partee*. Oxford: Blackwell Publishers. longer version available as Partee (2005).
- Partee, B. H. 2005. Reflections of a formal semanticist. URL: http://people.umass.edu/partee/docs/BHP_Essay_Feb05.pdf. longer version of Partee (2003).
- Partee, B. H., A. ter Meulen, and R. E. Wall. 1993. *Mathematical Methods in Linguistics*. Kluwer. Corrected second printing.
- Peeters, B. (Ed.). 2006. *Semantic Primes and Universal Grammar: Empirical Evidence from the Romance Languages*. Amsterdam: Benjamins. anulib electronic.
- Pelletier, F. J. 1999. A brief history of natural deduction. *History and Philosophy of Logic* 1–31. slightly shortened version downloadable from: [//citeseer.ist.psu.edu/pelletier98brief.html](http://citeseer.ist.psu.edu/pelletier98brief.html).

- Pietroski, P., J. Lidz, T. Hunter, and J. Halberda. 2009. The meaning of 'most': Semantics, numerosity, and psychology. *Mind and Language* 24:55–585.
- Pollard, C. 2007. Hyperintensions. To appear in *Journal of Logic and Computation*. URL: <http://www.ling.ohio-state.edu/~hana/hog/pollard2007-hyper.pdf> (viewed June 2008).
- Pollard, C. 2008. Hyperintensions. *Journal of Logic and Computation* 18:257–282. <http://www.ling.ohio-state.edu/~pollard/cvg/hyper.pdf> (accessed 19 Feb 2010).
- Restall, G. 2000. *An Introduction to Substructural Logics*. Routledge.
- Wierzbicka, A. 1972. *Semantic Primitives*. Frankfurt: Athenäum Verlag.
- Wierzbicka, A. 1985. *Lexicography and Conceptual Analysis*. Karoma.
- Wierzbicka, A. 1996. *Semantics: Primes and universals*. Oxford.
- Wierzbicka, A. 2002. Russian cultural scripts: The theory of cultural scripts and its applications. *Ethos* 30:401–432.
- Wierzbicka, A. 2006. *English: Meaning and Culture*. Oxford University Press.
- Wierzbicka, A., and C. Goddard. 2002. *Meaning and Universal Grammar - Theory and Empirical Findings, vols I and II*. Philadelphia, PA: John Benjamins.