

GREG CARLSON

'MISMATCHES' OF FORM AND INTERPRETATION

Abstract. Much work in language acquisition has focused on the meanings of major lexical items (chiefly nouns, adjectives, and verbs). This chapter instead focuses on the types of issues and problems that arise when one focuses on the meanings of function words in a language (such as tense markers, conjunctions, articles, etc.). The issues that arise are very different from those surrounding the acquisition of major lexical items. It is pointed out that the function items interact with the syntax in complex ways that form a special challenge for any current view of language and meaning acquisition.

1. 'MEANING' AND SEMANTICS

The theme of the conference 'Semantics meets acquisition' that led to this volume has an amusing ring to it as it hearkens back to 'B' movie titles such as 'Wolfman meets Frankenstein'. This latent reference to movie monsters turns out to be apt, in a certain sense, as when I think about the twin issues of acquisition and semantics, and how to put them together, it does seem a frighteningly hard problem indeed. Were this presented to me as an abstract problem in a form that I didn't recognize as really about learning and meaning, I'm sure I would throw up my hands and soon declare the problem insoluble. But this of course would be a misjudgment, as it is contradicted daily by simple facts of the world.

In this paper, I initially take a fairly superficial, perhaps even naive, perspective on matters of meaning and learning. I am going to assume that early language learners have, at best, access to knowledge of surfacy kinds of linguistic information (such as word boundaries), and some knowledge of context. I will present in overview style some of the challenges learners might face in trying to construct a consistent form-to-meaning mapping. (Throughout I take the point of view of someone trying to understand a language). One way to begin thinking about the issue is 'from the top', so to speak. The *experience* of extracting information from natural language utterances is a global one — the experience is that of understanding something you didn't before the utterance event occurred, and that's about it. This does not distinguish for instance among presuppositions, conveyed meanings, implicatures, literal, or metaphorical meanings, nor any other

1

V. Van Geenhoven (ed.), ????????, ?—?.
© yyyy Springer. Printed in the Netherlands.

types of information derived from your encounter with the utterance, e.g. location, gender, emotional state of the speaker, etc. Some take this intuition about the unity of our experience at face value — this I take it is the underpinnings of “holism of meaning” (Chierchia & McConnell-Ginet, 2000) but I and many others believe that messages extracted from natural language are susceptible to analysis (the ‘atomist’ position) and it becomes clear that analysis leads to a view that meaning *in toto* is composed of a variety of distinguishable factors.

Let us draw a parallel. Upon hearing a single word, say, the English word *cats*, one has the experience of hearing a noise and pairing it with a certain type of animal, very roughly. There is nothing, I believe, in this experience that comes identified as also experiencing ‘a word’, ‘two morphemes’, ‘a stem’, ‘the feature [-sonorant]’, ‘Noun’, and so on and so forth. Yet, upon analysis it becomes clear that this experience is somehow informed by a constellation of such factors, that all these factors or factors like them contribute their part to the whole. The experience of meaning is, I believe, likewise amenable to such analysis, and when one considers the factors it becomes clear that *meaning experienced* in its broadest sense results from a combination of similar factors, factors that do not wear their rank on their sleeves but which become apparent upon consideration through the lens of analysis. When we use the term ‘semantics’, we intend a certain *component* of meaning, that component which is in some sense referentially based and which is connected most intimately with the syntax of natural language: I’m going to refer to this as the truth-conditional aspect of meaning, a phrase I use here for convenience rather than in its fullest theoretical sense. This is the aspect of meaning which is absent from otherwise meaningful objects and events, such as the dark emotive colors in a painting, the rattling sound in my car, music, and, apparently (though I want to be a bit careful here), animal communication systems.

However, language clearly conveys meaning in ways in common with such other things, as well. Consider a point emphasized in Grice’s work on conversational implicatures (Grice, 1975). He takes pains to point out that these implicatures apply to actions in general, not just the linguistic actions of executing utterances. For instance, one can congratulate someone by patting them on the back or shaking their hand, or one can do it linguistically by saying something like *Way to go there, Bob* or by using the stodgy performative utterance *I hereby congratulate you on your success*. Meanings of actions then, including linguistic actions, contribute one component to the meaning of the whole. Another type of meaning that is not commonly discussed in truth-conditional approaches is that of connotative or social/emotional meaning, associated primarily with words and set phrases. To learn a language is to learn, in part, facts like *rump* is a somewhat cruder

way of making reference to a certain body part than *hind end*, and that *derrière* (in English) is almost affectedly silly, in most contexts, despite common reference. Such social/emotional meaning is omnipresent in language, is an essential part of its overall meaning, and seems most highlighted in poetry, song lyrics, and corporate presentations; but it is a type of meaning clearly present in nonlinguistic artistic objects and events as well. A further component of meaning arises from background cultural knowledge. For instance, the significance of the color red varies from culture to culture. And, in language, it is not a good idea (in English) to wish someone a refreshing night's sleep by saying *Rest in peace*, as this is a formulaic phrase that is associated with gravestones.

The purpose here is not to enumerate or catalogue the variety of meanings that the use of natural language gives rise to. Rather, it is to make the point that when we begin to talk about the semantics of a quantifier or the scope of tense marking, and how they might be acquired, we are already a long distance out of the starting gate in considering the general issue of meaning and language. Meaning comes at us — and people learning a language — from a variety of different directions, at a large numbers of levels, and only one among them is the subject of the kinds of semantic theories that are intimately connected with the syntax of language. Learners must somehow identify this level.

2. LEXICAL SEMANTICS

How this is accomplished remains something of a puzzle. But let us assume it gets done. Even then, when we restrict consideration to just this semantic, truth-conditional aspect of meaning, the difficulty of the problem of learning hardly abates. Obviously, perhaps most obviously, one must learn the meanings of the words of the language (or, a significant subset of them, at any rate), and there are many terrifically interesting learning issues that have been explored within this domain, at least in the area of learning meanings of the content words, noun, verbs, adjectives, in the main (e.g., Gleitman, 1990; Bowerman 1980; Clark, 1993; Bates *et al*, 1979). One absolutely immediate problem that comes up here is that of ambiguity. I would like to point out that the problem, even at the lexical level, is of mind-boggling proportions. In the early 1960's the linguist Charles Fries did a count of how many meanings were listed in the Oxford English Dictionary for the 500 most common words of English; there were in excess of 14,000, or about 28 different meanings per word, on average. Granted, many would be regarded as different 'senses' rather than different meanings proper, and a good many of them are low-frequency or even archaic usages that are learned later in life, if at all. further, in context none are remotely as ambiguous as in citation form (though again,

the question is how learners could marshal the context to narrow down the possibilities in the first place). On the other hand, many ambiguities are not included in this count. Type/token ambiguities are systematically associated with nouns (e.g., the ambiguity of *All the machines at the arcade are for sale*, whether it is those actual machines or other individual machines of the same design). Metonymic reference is not reflected there, as in the practice commonly cited of waitpersons referring to customers by their orders, resulting in ambiguities in sentences like *The ham sandwich is ready to eat*. Many ambiguities of thematic role assignment, which are extremely common, are missing. *John shoveled the cement* has the cement either as location cleared, or what was moved, *Sally packed the suitcases* can have the suitcases being put into things, or things being put into them, etc. Nor are many figurative meanings, on the whole, included in this count. It is hard to see offhand, or even on reflection, how any of this helps a language learner. To pair a new meaning with a word for which you already have a meaning, it appears one must notice there is an error, in the first place, which requires tremendous sensitivity to context *and* what is appropriate in a context. In the second place one must also localize the error: if one hears *That actor is a ham* and notes that the utterance is wrong in context for the 'smoked meat' sense of *ham*, why not conclude the error is due to *actor*, *that*, *is*, or *a*? Or perhaps it is a scope ambiguity, a topic/focus distinction?

The problem of learning words that are ambiguous has hardly gone unnoted, and in fact has received a huge amount of attention in the language learning literature (e.g., for a recent overview see Siskind, 2000). The chief focus, to this point, has been on the learning of ambiguous lexical items (nouns, verbs, adjectives, prepositions). But understanding an utterance of a sentence or discourse involves, of course, much more than just understanding the meanings of lexical items, and resolving their ambiguities in context. It involves consideration of the ways the lexical items are combined with one another, and here, as is well-known, the problem of ambiguity is likewise omnipresent — one could responsibly say that it has been a central problem in psycholinguistic and computational research in the past few decades.. As we know the linear order of words can make an essential difference in meaning (*Dogs chase cats* vs. *Cats chase dogs*) but in many other instances there is no difference (*Mary put out the cat* vs. *Mary put the cat out*, Scrambling structures that appear in most languages, etc). But perhaps most interestingly, a vital part of learning this combinatorial semantics is ferreting out the contributions of all those 'little' words to the meaning of the whole. Considering the contributions of these not so apparently referential things is a central focus of semanticists: what is the meaning of a past tense, a modal, an indefinite article, a reciprocal expression, *if*, *how*, *which*, what does an infinitive marker do, a plural ending, negation, interrogative pronouns?

3. NON-LEXICAL SEMANTICS

Consideration of these *functional* elements of meaning (Lebeaux, 1988; Carlson, 1983) introduces issues about the mapping between forms and meaning that are either absent or obscured when one concentrates primarily on the semantics of lexical items, or grosser aspects of sentence meaning such as argument structure. Consider the instance of the Classical Latin conjunctive particle *-que* (e.g., Hale & Buck, 1966). Latin had this alongside the conjunction *et*, but the syntax of the two was not the same. *Et* appeared, from a semantic point of view, 'right where it is supposed to', that is, between the elements conjoined, like most conjunctions we're used to seeing. The enclitic *-que*, on the other hand, appeared attached to the end of the first word of the phrase conjoined. Thus in (1) *-que* appears after the first word but signals that the whole phrase is a conjoined element, and not just the word *two*:

- (1) *duasque ibi legiones conscribit*
 two-and there legions enrolled
 '... and there he enrolled two legions.'

In a slight wrinkle probably driven by prosodic considerations, it appeared attached to the second word if the first was a monosyllabic preposition:

- (2) *ob easque res*
 because these-and things
 '... and because of these things'

If one treats *-que* as having the meaning of a conjunction, and compositionally combines it with whatever it is combining with on the surface syntax, one would not be able to get these meanings. Instead, one must in some sense 'raise' it up to a higher position in the tree structure, and put it in its rightful place. (This is a lot like Quantifier Raising, of course, with the notable difference that in the case of *-que* one does not wish to leave a variable in situ). Of course in such examples *-que* is not in any 'wrong' place — to put it elsewhere would be wrong — because the grammar says it is to be put where it is. But from a compositional semantic point of view, one needs to do some rearranging that one does not need to do with *et*, *and*, and *und*.

This particular example is not simply an idle little curiosity to note and tuck away. The position of *-que* is of course a Wackernagel position phenomenon — a 'second position' phenomenon — a syntactic-morphological position so common it has a name (Anderson, 1993). But the

phenomenon of having to ‘rearrange’ functional elements semantically extends well beyond Wackernagel position particles such as these. Consider how common it is to treat tense, for instance, both syntactically and semantically as a higher-level operator, and for good reason. A very common type of example from English Verb Phrase Ellipsis will illustrate this point--the deleted VP in (3) does not carry the tense information of the (underlined) antecedent VP, even though tense is expressed as an inflection on the verb in that antecedent:

- (3) John *wrote a paper* because he had to (*wrote a paper).

Or, it appears plurality must be dissociated from the noun it appears attached to, by similar evidence:

- (4) John has two *dogs* and Fred has one (*dogs).

This listing of functional elements that appear ‘out of place’, from the point of view of where in the structure the meaning is contributed, is easily extendible to the point where one can easily conclude that it is a common and possibly essential feature of language that learners must somehow master. Is this something we are born knowing already? That would help, it seems, but how can one tell?

The appearance of items that do not seem to be ‘in the right place’ is but one of the issues learners must face in dealing with the meanings of functional items. Coalescence phenomena between adjacent functional elements are extraordinarily common — it is the classic definition of an inflectional language as opposed to an agglutinative one. Coalescence may also occur with otherwise free morphemes, as with the preposition/article coalescence found in Germanic and Romance; thus, French *du* is in some sense the equivalent of *de+le*. From the commonsense point of view taken here, a learner is required to assign a composite meaning to such elements as *du*, but not to other elements such as *le*, or any lexical items. A very similar process that well could cause increasing difficulties is when a sequence of two formally identical functional elements is reduced to one (a variant of haplology). This does not, to my knowledge, occur with lexical items (thus, *a bare bear* does not reduce to *a bear*, meaning, ‘a BARE bear’). Consider the case of Japanese *-no*, noted by Kuno (1973), Radford (1977), and others. It has two quite distinct functions, as a possessive postposition and as a pronoun (meaning some thing like ‘one’). If these are juxtaposed, as in (5a), you get an ungrammatical sentence. But there is a nonperiphrastic way of expressing this, namely, (5b), with only one instance of *-no*. But both the possessive meaning and the pronominal meaning remain:

- (5) a. **Kore wa anata no no desu ka*
 This TOP you POSS one be
 ‘Is this yours?’ (lit. ‘Your one’)
- b. *Kore wa anata no desu ka*

Again, this is hardly a funny little isolated fact. One can multiply examples by the dozens in familiar and unfamiliar languages alike, and, as usual, when one looks for something like this, it seems to be everywhere. The Swahili negative past *ku* occurring right next to the infinitival marker *ku* reduces to a single *ku*- prefix, yet both meanings remain. In certain Turkish word forms (in the instance of NP’s like *their books* where both the possessors and things possessed are plural) two plurals ‘ought’ to appear in a row, but only one appears; there are, however, two plurals, semantically. The special problem that examples like these raise is that, from a surfacy point of view, you have one element with two meanings, or the same meaning assigned two different scopes, as in the Turkish example. But I thought it was almost an axiomatic fact of perception that a single form could *not* be assigned two different meanings. Not only does this apply to lexical items. *He sat by the bank* cannot mean he sat by the river *and* a financial institution but this applies to perception more generally — this is Necker cube stuff. This would seem a prime case of putting the learner squarely behind the eight ball; yet, there it is.

We not only have the case of one form with two meanings to be concerned about, but also its converse. Two (or more) forms that add up to a single meaning. One reflection of this is discontinuous morphology. For instance, Nida (1978) cites the Kekchi examples in (6):

- (6) a. *oççoçc* b. *roççoçce’p*
 ‘house’ ‘their house’

French *ne ... pas* would be a possible candidate for a more familiar example. But far more commonly this is found in agreement or concord forms: an agreeing plural article, two plural adjectives, and a plural noun add up one simply one plurality, not four. A definite article combined with the definite form of a noun still add up to one definite. Multiple negations, as given in the Old English example in (7) add up to a one single negative:

- (7) Ac he ne sealde nanum nytene ne
 and he NEG gave NEG beasts NEG
 nanum fisce nane sawle.
 NEG fish NEG souls

‘And he did not give beasts or fish souls.’

Such examples are so familiar we might easily overlook the language learning problem: if we build a signal-detector that generates an associated meaning upon encounter with a certain form, we are going to get extra meanings generated which are not parts of the actual interpretation. Note that the strategy of treating certain forms as meaningless, and localizing the meaning to just one of the forms, may work in some instances but not generally. Let us take a really simple example, the English phrase *These houses*. Here we have two plural forms, so let us treat the one on the noun as ‘real’. The problem is that *These have wooden doors* has a plural subject, semantically and in all other respects, and so does *Houses have wooden doors*.

It also appears on occasion that sounds are not paired with meanings. We are all used to work on expletives, so I’ll draw on examples from another domain, that of Classical Latin semi-deponent verbs. Latin had a productive inflectional passive marker that normally signalled passivization (i.e., the subject is semantically the direct object), but in many semi-deponents while the present tenses were formed from the usual active paradigms, the perfect forms required the passive morphology, but without a corresponding effect on passive meaning. Here is a textbook example:

- (8) a. *audeo* b. *ausus sum*
 ‘I dare’ ‘I dared’ (not, ‘I was dared’)

Or, consider the habitual markers that appear in contrafactuals in some languages. In (9) is an example from Hindi cited in Iatridou (2000), due to R. Bhatt:

- (9) a. ??Meera do baje bhaashaN de
 Meeraa 2 o'clock speech give
 rahii ho-tii (hai)
 PROG be-HAB (PRES)
- b. agar Meera kal do baje
 if Meera yesterday 2 o'clock
 bhaashaN de rahii ho-tii...
 speech give prog be-HAB
 ‘... if Meera had been giving a speech yesterday at 2:00 ...’

Here, there is no discernible semantic contribution of the habituality marker HAB in (9b), while in (9a) its presence does make a semantic contribution,

which makes the point-time adverbial sound strange (as generalizations are often odd if given point-time readings); but not so in (9b). English pluralia tanta (*scissors, pants*), or dependent plurals (as in *Unicycles have wheels*) would be possible examples of a plural making no semantic contribution. I will not go on, but language seems to have many instances of interpretable elements that, in given constructions, bear no such or seemingly any meaning.

Or, what they can do is bear *other* meanings instead. An illustrative case is the Spanish spurious *se*, first discussed to my knowledge in the generative literature by Perlmutter (1971). In sequences of Spanish clitics, if the third person indirect object clitic appears before a third person direct object clitic, it is realized as *se*, which is normally taken to be a reflexive form (though it has other functions as well). However, the meaning is not (necessarily) reflexive:

- (10) *Se lo mandas. *Le/*Les lo mandas.*
 'You send it to {him/her}/them.'

Again, this might at first appear a funny little fact, but forms that are, from a transformational point of view, mapped to other forms in syntactic context are extraordinarily common. Consider sequence of tense phenomena, where a past tense appears in a subordinate clause, but it has a reading cotemporaneous with the interpretation of the higher tense as if, semantically, it were a present tense. In preposition/pronoun inversion in Germanic (now lost in English except in frozen forms like *thereupon* or *therefore*), a (neuter) personal pronoun seems expressed instead as a locative, as in German *damit, darauf*. In Modern Greek, we find in certain contexts imperfectives that appear to contribute perfective meaning, as in wishes and contrafactuals. Again, we are not looking at some spotty little curiosities, but rather some features which detailed analysis and study show recur time and again.

4. LEARNING

Even if one has a meaning, or a sets of meanings, paired with phonological forms, there remain serious issues about learning to put them together to form appropriate coherent meanings for the whole. The question I would like to pose is what minimal assumptions about the learning process can we make and account for the daily fact that language is, indeed, learned? Perhaps the most minimal assumption one could make is that this organization is also reflected in the learning of other domains (this is the 'there is nothing special about language' view). For example, perhaps there are organizational principles in learning to structure visual scenes, which encode this same

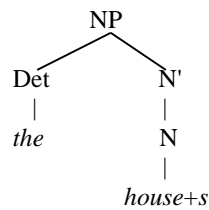
arrangement of matters. My own impression of the state of such research, however, is that principles of learning based on work in other domains has proven of limited value to the learning of a full language. While there may be some spotty successes in, for instance, learning the location of word or morpheme boundaries, or learning certain word meanings demonstratively, it has yet to be shown that such mechanisms yield anything like a system capable of learning language, much less extracting appropriate full sentence meanings for anything beyond *Man bites dog* types of structures (and even those ignore the tense).

Lack of success is hardly an argument that something cannot be done eventually, and this must remain a possibility. As long as one views the learning task of language as a matter of learning the syntax of the language, that is, as an arrangement of forms, it is a seductively easy step to take to believe that learning language form arrangements is just like all the other form arrangements we learn as well that are not constituted of language. But when one focuses on meaning, particularly entire sentence meanings, the ground shifts, and I believe this step becomes much more difficult to take. This is because it is in a language, and only in a language, that one finds the essential ingredients of a semantics — truth, reference, and predication — and that these are inextricably bound up with the syntax of a sentence. No other naturally occurring object, including (in all likelihood) animal communication systems, has these properties and characteristics. It is in studying semantics, then, particularly non-lexical semantics, that the unique properties of language become most starkly evident, and thus it becomes increasingly reasonable to believe that special domain-specific language learning measures are called for.

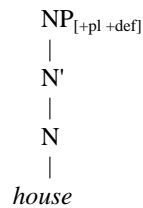
Let us return then to the empirical issues raised above, and ask again what sorts of minimal assumptions might be necessary for the learning of a language. At this point I can at best speculate, but perhaps there is a general direction that may have some value. It has been long noted among grammarians that most (if not all) languages make a distinction between types of vocabulary items. This distinction has been characterized in various ways (e.g. ‘full word’ vs. ‘particle’, ‘open class’ vs. ‘closed class’), but I will choose to use ‘lexical item’ vs. ‘function item’. Lexical items consist most of prominently of nouns, verbs, and adjectives (with prepositions and adverbs sometimes included), and function items are all those little words that accompany them, such as articles, conjunctions, infinitive markers, the copula, auxiliary verbs, pronouns, and so forth. As suggested by the terms ‘open class’ and ‘closed class’, the lexical items tend to allow the relatively easy addition of new members of the category, via productive processes of the language or borrowing. Function items, on the other hand, lack the internal productivity of the language to add new items, and are generally

much more resistant to borrowing (for example, current English lexical vocabulary is about 60% of non-native origin, but among the function items nearly 100% is of native Germanic origin). Whether a simple two-way distinction is an appropriate characterization remains unclear (e.g., prepositions, a closed class, nevertheless appear to have some lexical characteristics as well). This said, I am going to outline things in these binary terms.

Among the closed-class items are not only words, or free forms, but also inflectional morphology and clitics. As the discussion in the previous section indicates, these function items have syntactic/semantic properties not shared by open-class lexical items. Chief among these is that their interpretation may take place at a "higher" level of syntax than the interpretation of their host word (in the case of inflectional morphology and clitics). It is not fully certain, but appears a fact that their interpretation may not appear 'lower' in the syntax than surface constituency would indicate. In Carlson (1983), a point of view is developed which is designed to express this distinction between lexical items and function items in which the function items form a part of the *structure* in which the lexical items appear. One formalization of this idea is to treat the function items as expressions of features that appear on the syntactic categories. A rough representation of what a plural definite noun phrase might look like is not the usual:



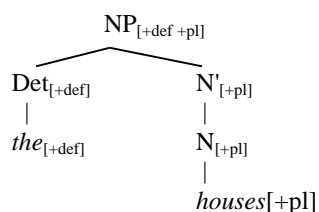
but rather:



It is the features that are then interpreted by the semantics, and not the phonological forms of the expressions of those features. Thus, *the* and *-s* as

phonological forms do not have meanings assigned to them (unlike the phonological form *house*, which does), but the features themselves do have meanings associated with them. One reflection of this is that many function items might have (and commonly do have) a nonanaphoric null expression, which provides no phonological form for interpretation but nevertheless provides a feature as surely as when there is phonological expression.

In order for these features to receive expression, they may be ‘passed down’ to nodes or heads of phrases (this is simply the reverse of ‘feature percolation’) to determine the forms. Thus, in the rough example, the representation might look as follows:



The features then are phonologically interpreted ‘at the bottom’, and semantically interpreted ‘at the top’. In Carlson (1983), the use of features was inspired by work in GPSG, but the basic idea can easily be translated into other frameworks, some of which may not make use of discrete features such as those exemplified.

What, then, would be possible assumptions that language learners might bring to the task of learning how to interpret language? One thought is that language learners have some type of foreknowledge of this basic architecture of a language (however represented, it need not be a Cartesian idea, but could be more indirectly expressed in the architecture of the learning mechanism itself), combined with some understanding of the types of meanings that can be expressed by features versus those that may be expressed by lexical items. For instance, feature meanings tend to be context-sensitive in a way that distinguishes them from meanings of lexical items. Once this is understood, one of the first tasks is to identify those elements in the speech stream which express lexical items versus those that serve as expressions of features, and it is at this task level that a relatively weak and non-domain-specific learning strategy could well turn out to be of considerable value, making use of such notions as intonation, stress and tone assignment, opportunities for phonological assimilation, statistical patterning, etc. However, on this view, the more general mechanisms themselves do not provide an account of the *significance* of what is found, that is provided by the larger and domain-specific prior understanding of the nature of the object being learned.

These comments are of course highly speculative, and their value is found only in the extent to which they might provide a productive means of accounting for how a person might end up learning the semantics of a language, despite its challenging, even daunting, complexity.

5. AN EXERCISE IN DEFINITE ARTICLES

The point of this final section is to illustrate, using a certain domain of data, the types of problems of matching forms and interpretations that learners might encounter. The data we are going to examine first has to do with nouns that appear without any articles before them. As there is a vast literature in semantics on the topics of mass terms and bare plurals, I am going to focus on a somewhat lesser-studied topic: singular count noun forms without articles. While these are the norm in many languages that lack plural marking and articles, such as Chinese and Japanese, in Indo-European languages, for instance, they appear less often and in fewer contexts (see Kallulli, 1999; Borthen, 1998; Schmidt & Munn, 1999; Dayal, 1999). English presents in its own special issues with the bare singular construction. An excellent overview of the English data and many theoretical matters is found in Stvan (1998).

The basic facts seem to be these. Bare singulars are both lexically and positionally restricted. So we have contrasts such as those in (11):

- (11) a. They put him in *jail/prison*/**penitentiary*.
 b. I took my son to *school/college*/**university* (Am. English)
 c. The men were found on *shore*/**beach*.

These lexical restrictions are subject to considerable dialect (such as American vs. British English) and social variation (e.g., if one is a member of the art world, one is more likely to say or hear *She is in studio*, which to my inartistic ears sounds strange).

Syntactically, they often follow certain verbs and prepositions (as in (11) above), but may appear occasionally as subjects of certain verbs:

- (12) a. *Prison* has little to offer in the way of recreation.
 b. *College* is a good place to learn.
 c. *School* makes Jimmy very happy.

They may not be modified (unlike bare plurals); this follows from Stvan's arguments that these are, in fact, full one-word phrases and not simply nouns:

- (13) a. They sent him to *(big) jail.
 b. I watched it on television *(that had a 31" screen)

Bare singulars can also appear, to a certain stylistic effect, in conjunction constructions and a couple others. In these particular instances the lexical restrictions are eliminated or reduced:

- (14) a. University and high school alike require much study.
 b. Neither television nor radio has become serious educational tools.

Impressionistically, these structures appear to share many of the positional constraints of bare plurals in Spanish and Italian that have been analyzed as properly governing and empty D position (Contreras, 1986), and they seem to share the lexical constraints of incorporated and incorporation-like structures found in other languages.

The main semantic observation is that bare singulars appear to be non-referential, in the following way. Consider a situation in which Bob is watching television. There's a definite TV he is then watching, and one can refer back to that TV, e.g., by continuing *and then he turned it off and went to bed*. However, if we use this sentence as the antecedent of VP ellipsis, consider the result:

- (15) Bob was watching television, and Fred was, too.

There is no reading of this where both had to have been watching exactly the same television set. This is exactly what you would expect if *television* were treated as a narrow-scope, nonspecific indefinite.

The meanings are actually more subtly articulated than just this. Stvan (1998) lists three main meanings. One is an 'activity' meaning. If someone is *in bed*, for example, it does not mean just a person is located in a certain place, but is also partaking of a certain activity or relationship that is canonically associated with that location — in this instance, sleeping or at least trying to. If one is *in prison*, then one is incarcerated, not simply in that location. This 'activity' meaning is not always present, and a 'familiarity' meaning may appear instead. If John is *at home*, or *in town*, or *out of state*, there is no special activity or relationship implied. However, a seemingly deictic meaning is intended, i.e., he is at *his* home, he is in *our* town, or out of *this* state. It is not just any home, any town, or any state. In these instances, the meaning of the bare singular, a maximally unmarked construction, is actually *more* complex than similar constructions with determiners (such as being in *a* prison, or in *a* bed), not less. Finally, there is the generic reading that lacks both of these dimensions. If tape recordings are not admissible *in court*, or if religious teaching is questionable *in prison*, all prisons or courts (of law) are under discussion.

Now let us move to the title of this section, cases involving overtly definite noun phrases. If we use VP ellipsis in such cases, identity of reference is preserved. This is intuitively obvious because definite articles are used to pick out something salient or familiar or unique in the context; the English word *the* signals this:

- (16) a. Bob attended the old brown school, and Sam did, too.
 b. Max liked the Brecht play a lot, and Susan did, too.

Naturally, they went to the same school, liked the same play, because of the definiteness. Or so the usual story goes.

However, certain unmodified, lexically selected nouns seem to work differently (the basic intuitive observations can be found in Birner & Ward, 1994):

- (17) a. Sam is in the hospital again, and so is Mary.
 b. I heard about the riot on the radio, and Sharon did, too.

Sam and Mary need *not* be in the same hospital, and while Sharon and I did hear about the same riot, our radios may well have been different, as if that phrase is an indefinite. But we have this definite article, what are we to make of that?

Note that in the presence of restrictive modification, this indefinite reading disappears, and the truly definite reading emerges as the only option:

- (18) a. Sam is in the Red Cross hospital again, and so is Mary.
 b. I heard about the riot on the radio sitting on my desk, and Sharon did, too.

Suddenly, for American English speakers, Sam and Mary must be in the same hospital, and have heard the very same radio. These examples illustrate two hallmarks of the bare singular construction: lack of modification and lexical restriction (e.g. *riot* vs. *radio*). The other hallmark, syntactically restricted distribution, may likewise be established. Thus, we have some instances of definite articles which lack their usual semantic effect.

It seems many noun phrases with definite articles work this way in English, but others (such as *the riot*) do not. There is a certain amount of work, e.g., by Longobardi (1994) and Vergnaud & Zubizarreta (1992), that introduces the notion of 'expletive article', one put in to produce a noise in an otherwise empty D position. What I am suggesting here is that in instances such as (17), there is a reading (the most natural one) where the definite

article is expletive, that is, we *really* have an instance of a bare singular in each case, the semantics of which is similar to that of bare plurals.

It is misleading to think of such constructions as ‘idioms’, being surprisingly highly productive and lacking the figurative or clearly non-compositional meanings often associated with clear instances of idioms. One way of thinking about what is going on here is to think of the construction as a complex predicate of the type examined in Snyder (2001) regarding first language acquisition (Snyder also includes a cross-linguistic comparison of a variety of languages). Snyder considers the acquisition of a series of English constructions that have been suggested by syntacticians to be complex predicates, that is single predicates made up of two or more syntactic parts. For instance, one among them is the resultative construction, as in *He hammered the metal flat*, where *hammer flat* appears to function as a single predicate. Verb-particle constructions are another example. Based on theoretical concerns as well results of the cross-linguistic study, Snyder notes that there is a close connection between the availability of productive root compounding in nominals, and the appearance of what are analyzed as complex predicates in a language. Snyder’s study with children indicates that there is a highly positive correlation in children learning English between the time of appearance of both productive root compounding, and of the verb-particle construction (as a representative of complex predicates). If the non-referential definites discussed above form a complex predicate with the verb, then there should also be a positive correlation between acquisition of compounds, complex predicates, and this particular reading of definites. Whether this (speculative) prediction is correct is presently unknown, but it would provide one starting place for further investigation.

To return to the overall point, how could one see through the complexity of language to learn such facts, though? There would appear to be considerable usefulness in such notions as meaningful things that occasionally mean nothing, and things like null determiners, as such seem a persistent part of the design of language. But how could this ever actually help a learner acquire an appropriate syntax and interpretations? From a surfacy point of view, things like this appear to be roadblocks to learning, as if messages which easily could have been encoded in a straightforward signal-to-meaning relationship have been cleverly garbled and disguised. But I am suggesting a perspective that is perhaps the opposite of this — that such ‘mismatches’ in fact contain cues and clues to meaning that enable learning the whole system, particularly the non-lexical semantics, and are not there to hinder it, as a surfacy signal-to-meaning notion would suggest.

6. CONCLUSION

Natural languages, from a commonsensical point of view, seem treacherously designed. We have some things that mean nothing, and nothings that mean something. We have two things meaning one thing, and one thing meaning two things. We have things in disguise, meaning in highly constrained contexts what something else means that it normally contrasts with. We have things, even if the meaning is a single, normal seeming meaning, that are put in the 'wrong' place and have to be figured instead for another.

Things like this are learned by scores of millions annually, and one task of language learning research is to produce sensible and effective ideas about how this could be. It would appear that such work would be most effective if embedded within a larger and comprehensive framework articulating the form-to-meaning mapping of the general sort worked on by semanticists and syntacticians, if only to provide some appreciation of the sheer difficulty of the task ahead.

University of Rochester

7. ACKNOWLEDGEMENTS

I wish to thank the many workshop participants who commented on the presentation. Special thanks go to Veerle Van Geenhoven and Tom Roeper. This material is based upon work supported by the National Science Foundation under Grant No. 0082928.

8. REFERENCES

- Anderson, Stephen. "Wackernagel's Revenge: Clitics, Morphology, and the Syntax of Second Position." *Language* 96 (1993): 68–98.
- Bates, Elizabeth, L????? Beghini, I????? Bretherton, L????? Camaioni and V????? Volterra. *The Emergence of Symbols: Cognition and Communication in Infancy*. New York: Academic Press, 1979.
- Birner, Betty, and Gregory Ward. "Uniqueness, Familiarity, and the Definite Article in English." In *Proceedings of the 20th Annual meeting of the Berkeley Linguistic Society*, 93–102. Berkeley, CA: BLS, 1994.
- Bloom, Lois. *One word at a Time*. The Hague: Mouton, 1973.
- Borthen, Katja. *Bare Singulars in Norwegian*. Cand. phil. thesis, Norwegian University of Science and Technology, 1998.
- Bowerman, Melissa. *The Structure and Origin of Semantic Categories in the Language-learning Child*. In M. Foster and S. Brandes (eds) *Symbols as sense*. New York: Academic Press, 1980.
- Carlson, Gregory. "Marking Constituents." In Frank Heny and Barry Richards (eds.) *Linguistic Categories: Auxiliaries and Related Puzzles, Volume 1: Categories*, 69–98. Dordrecht Reidel, 1983.

- Chierchia, Gennaro, and Sally McConnell-Ginet. *Meaning and Grammar* (2nd Edition). Cambridge, MA: MIT Press, 2000.
- Clark, Eve. *The Lexicon in Acquisition*. Cambridge: Cambridge University Press, 1993.
- Contreras, Heles. "Spanish Bare NP's and the ECP." In Ivonne Bordelois, Heles Contreras, and Karen Zagona (eds.) *Generative Studies in Spanish Syntax*, 25–49. Dordrecht: Foris, 1986.
- Dayal, Veneeta. "Bare NP's, Reference to Kinds, and Incorporation." In Tanya Matthews and Devon Strolovich (eds.) *Proceedings of SALT 9*. Ithaca, NY: Cornell University, 1999.
- Gleitman, Lila. "The Structural Sources of Verb Meanings." *Language Acquisition* 1 (1990): 3–55.
- Grice, Paul. "Meaning and Conversation." In Peter Cole and J???? Morgan (eds.) *Syntax and Semantics 3: Speech Acts*, 41–58. New York: Academic Press, 1975.
- Hale, W. and C. Buck. *A Latin Grammar*. Alabama: University of Alabama Press, 1966.
- Iatridou, Sabine. "The Grammatical Ingredients of Counterfactuality." *Linguistic Inquiry* 31 (2000): 231–270.
- Kallulli, Dalina. *The Comparative Syntax of Albanian: On the Contribution of Syntactic Types to Propositional Interpretation*. Ph.D. Diss., University of Durham, 1999.
- Kuno, Susumo. *The Structure of the Japanese Language*. Cambridge, MA: MIT Press, 1973.
- Lebeaux, David. *Language Acquisition and the Form of Grammar*. Ph.D. Diss., University of Massachusetts, Amherst, 1988.
- Longobardi, Giuseppe. "Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form." *Linguistic Inquiry* 25 (1994): 609–669.
- Nida, Eugene. *Morphology: The Descriptive Analysis of Words*. Ann Arbor: University of Michigan Press, 1978.
- Perlmutter, David. *Deep and Surface Structure Constraints in Syntax*. New York: Holt, Rinehart, Winston, 1971.
- Radford, Andrew. "Counter-Filtering Rules." In J???? Green and S???? Harlow (eds.) *York Papers in Linguistics* (7), 7–45. ?????: ?????, 1977.
- Schmitt, Christina, and Alan Munn. "Against the Nominal Mapping Parameter: Bare Nouns in Brazilian Portuguese." In *Proceedings of NELS 29*, ???–???. ???: ????, 1999.
- Snyder, William. "On the Nature of Syntactic Variation: Evidence from Complex Predicates and Complex Word Formation." *Language* 77 (2001): 324–342.
- Stvan, Laurel. *The Semantics and Pragmatics of Bare Singular Noun Phrases*. Ph.D. Diss., Northwestern University, 1998.
- Siskind, Jeffrey Mark. "Learning Word-to-Meaning Mappings." In P???? Broeder and J???? Murre, (eds.) *Models of Language Acquisition: Inductive and Deductive Approaches*, 121–153. Oxford: Oxford University Press, 2000.
- Vergnaud, Jean-Roget, and Maria Luisa Zubizarreta. "The Definite Determiner and the Inalienable Constructions in French and English." *Linguistic Inquiry* 23 (1992): 595–652.