Chapter 6

Conclusion: The boundaries of linguistic competence

The question I began this study with was, How can a competence grammar account for one of the systematic aspects of variation, namely, non-categorical grammatical constraints? The preceding chapters have sought to answer this through a case study of copula absence in AAVE. Chapters 2 and 3 (among other things) elaborate the competence grammar that I am starting with. Chapter 4 presented preliminary evidence that non-categorical constraints are actually a problem for competence grammar: The results of the study suggest that people have knowledge of these constraints, as can be seen by the interaction between non-categorical constraints and social value/social meaning.

Chapter 5 begins with a discussion of the idea that variation doesn't exist solely in a formal plane, but is infused with social meaning, reviewing some of the literature on social meaning and variation. Chapter 5 goes on to ask how existing approaches to non-categorical constraints can handle the social aspects of sociolinguistic variation including the results of the experiment reported in Chapter 4.

In this chapter, I will suggest that the key to a successful account of these facts lies in recasting the notion of linguistic competence. In brief, I will argue that the tighter coupling of competence and performance (or knowledge of language and use of knowledge of language) already adopted in 'performance-plausible' models of competence grammar (Bresnan 1978, Pollard and Sag 1994, Sag and Wasow 1999) leaves open the possibility of a more inclusive competence grammar. I will propose an account of the effects of non-categorical constraints observed in the experiment that involves adding three types of information to the competence grammar: social value associated with signs, (a certain kind of) probabilistic or at least frequentistic information, and over-specified types. All three extensions have already been proposed in the literature. The resulting model of grammar, not unlike the usage-based model (Langacker 1987, 1990; Kemmer and Israel 1994; Kemmer and Barlow 2000), is a first pass at a model of linguistic knowledge that fits into what Eckert (2000) describes as a social theory of language.

The structure of this chapter is as follows. Section 6.1 considers the question of what grammar is a model of, whether it concerns the language of the community or the language of the individual on the one hand, and whether it should reflect 'externalized' or 'internalized' language on the other. Sections 6.2 through 6.4 discuss the specific types of information that I propose to add to the grammar to account for non-categorical constraints and their interaction with social meaning, reviewing existing proposals along the same lines. Section 6.5 reviews existing criteria for considering phenomena a matter of grammar and proposes some possible alternatives. Finally, Section 6.6 discusses how this fits into a social theory of language (Eckert 2000).

6.1 What grammar is a model of

In order to discuss whether certain phenomena belong 'in the grammar' it is necessary to ground the linguist's grammar by deciding what it is to be a model of. There are two dimensions of possibility: First, a grammar may model the language of a community or the language of an individual. Second, within models of the language of the individual, one may distinguish 'intensional' models that are meant to reflect the actual internalized knowledge of the individual and 'extensional' models that attempt only to generate the set of form/meaning pairs that the individual would accept. The next two subsections treat each of these dimensions in turn.

6.1.1 Community grammar v. individual grammar

Weinreich et al. (1968) argue that the structure in variation is only apparent when the language of the whole community is studied. What looks random and meaningless at the level of the individual speaker, according to them, becomes part of a pattern when that speaker is considered with the other speakers in his/her community. Thus the variable rules that they present are cast as rules of community grammar—partial descriptions of the language of a community.

Chomsky (1986) takes a radically opposed position, arguing that only grammars as theories of the internalized language of individuals (I-language) can have sound scientific basis. A grammar that generates the set of sentences in a language is a theory of that language. But how is a linguist to find out what the relevant set of sentences is? For individual speakers, one could (with infinite time and a speaker of infinite patience) discover the set of sentences accepted by that speaker. Since speakers from the same community don't necessarily accept the same set of sentences, it is not at all clear what set of sentences constitutes the language of a community (E-language). Taking the set of all sentences accepted by at least one speaker in the community means running the risk of including contradictory sentences: pairs of sentences of which any given speaker will accept only one. From a Chomskyan viewpoint, this would potentially obscure parameters of variation. On the other hand, taking the set of all sentences accepted by all of the speakers in the community would surely lead to an impoverished system. Further, as has often been noted, the notion of community is not well-defined; it is not possible to decide exactly which speakers constitute a 'speech community'. Such a listing would be required, in principle, to enumerate sentences in the way imagined here. Chomsky concludes that linguists should be concerned with the language of individuals, since these systems at least are well-defined:

> Theories of E-language, if sensible at all, have some different and more obscure status [than theories of I-language] because there is no corresponding thing in the world. (1986:27)

The thing-in-the-world that corresponds to a theory of I-language is an individual's internalized knowledge of the language. That is, the underlying system behind the patterns of grammaticality is the individual's own cognitive state. There is no cognitive state that contains an underlying system for the language of the whole community, no community mind for the community grammar to be an aspect of.

Part of the issue here is that Chomsky and Weinreich et al. differ in more than where they take the structure of language to reside. Chomsky defines language in terms of grammaticality, and studies the patterns of grammaticality across different forms. Weinreich et al. are studying differences in form that do not affect grammaticality, and are investigating how the patterns of use of the variant forms articulate with the structure of society. Thus what looks ill-defined to Chomsky from the point of view of patterns of grammaticality is the only way to study what Weinreich et al. are interested in.

Further, it's simply not true that successful theories of systematic phenomena must be models of some thing-in-the-world that embodies the underlying system. For example, theories of macro-economics describe the systematicity of the combined commercial acts of whole countries. This does not require that there be a national mind which contains some representation of the underlying system.

Thus is might seem that it is possible to study the language of a community, and to posit a grammar of that language, as long as that grammar is not construed as a model of linguistic competence. However, this does not mean that sociolinguistic patterns are strictly outside the purview of competence grammar. Rather, the thing in the world that corresponds to community grammar is individuals' (imperfect) understanding of how other people speak. In this case, it doesn't really matter that speech communities are vague or ill-defined. Every speaker has some well-defined (although ever changing) set of linguistic and social experiences. From these experiences, speakers build their imperfect models of the speech of other people in their social landscape. Through the process of iconization (Irvine and Gal 2000) forms or patterns of usage become (for the speaker) iconic of types of people, personality characteristics, stances, etc. Note that this process involves analysis of or abstraction over both the linguistic and the social aspects of the speaker's experience. The result of this abstraction is that individuals come to see their communities as well-defined.

There is a danger of missing the crucially social aspect of language if one focuses on grammar as a model of linguistic competence. That is, individuals do not construct grammars any which way, but rather on the basis of their experience of language use in their communities. As Harder (2000) argues, individuals' mental representations of the linguistic system involve or invoke a notion of the collectivity. Focusing on lexical semantics, Harder argues that, from the point of view of any individual speaker, the meaning of most words is a fact about the way the world works.[1] I will return to these points in Section 6.2.2 below. At the same time, it is important not to overemphasize the role of the actual patterns of language in the community. That is, the fact that individual speakers represent their knowledge of language in terms of the expectations of their community doesn't entail that individuals are correct in their understanding of the expectations of the community.

6.1.2 Linguistic competence and psychological reality

In the preceding discussion, I have contrasted Chomsky's and Weinreich et al.'s positions along the community/individual dimension. Chomsky's position also represents a strongly cognitivist view:

> Statements about I-language ... are true or false ... The I-language L may be the one used by a speaker, but not the I-language L′, even if the two generate the same class of expressions ... (Chomsky 1986:23)

This is where the distinction between 'intensional' and 'extensional' models of language comes in. One and the same language (defined as a set of sentences) can be described by multiple different grammars. An 'extensional' model of language makes no claim of psychological reality. An 'intensional' model of language purports to be a model of what speakers actually know.

Wasow (1978) and Soames (1984) argue that the methodology of generative grammar can only provide evidence for extensional models. If linguists draw their evidence solely from the set of sentences accepted (grammaticality judgments), then grammars

[1] The exceptions to this are technical terms that people create and provide definitions for.

written on the basis of that evidence can only be equivalent in extension to actual internalized grammars of real speakers. While the language to be described places some degree of constraint on possible grammars describing that, it is clear that this constraint still far underdetermines the grammar. Any similarity between linguists' grammars built on this kind of evidence and actual internalized grammars is a matter of chance.

There is, of course, nothing wrong with an extensional model of language. It is a legitimate object of study. However, when it comes to providing explanations, and not just accounts, of linguistic patterns, extensional and intensional models have different potential sources of explanatory force. The customary sources in mainstream syntax—language acquisition/UG/the LAD in GB and the interfaces with other cognitive systems at PF and LF in Minimalist theory—belong to the domain of intensional models. Extensional models, on the other hand, draw explanatory force from functional and structural considerations.

However, it is not just the methodology of mainstream syntax that removes it from its claimed sources of explanatory force. Seuren writes:

> In the transformational literature, it is customary to claim psychological reality for competence models on the one hand, but to deny, on the other, that they could be taken as process models. (Seuren 1982:4)

That is, mainstream competence models are supposed to be models of actual linguistic knowledge, but not of the knowledge that gets used in language processing. This separation entails that psycholinguistic results and theoretical considerations about language processing cannot inform the design of grammar, on such a theory.

In contrast, Bresnan (1978), Pollard and Sag (1994) and Sag and Wasow (1999) (among others) argue for what Sag and Wasow call 'performance-plausible competence grammar', that is, models of linguistic competence that are embeddable in models of language processing (performance). On such a model, psycholinguistic results can inform the design of grammar as well as specific linguistic analyses. To the extent that this happens, the model of grammar is indeed a model of the knowledge of actual speakers. Bresnan puts it quite succinctly:

> If a given model of grammar cannot be successfully realized within a model of language use, it may be because it is psychologically unrealistic in significant respects and therefore inadequate in those respects as an empirical theory of the human faculty of language. (1978:2)

Pollard and Sag (1994) motivate their declarative, constraint-based framework by citing psycholinguistic evidence that language processing is highly incremental and highly integrated with information about the world, the context, and shared knowledge between speakers, and that linguistic information functions to like effect in different processing activities (notably, comprehension and production). Similarly, as noted in Section 2.3.4 of Chapter 2, there is psycholinguistic evidence to support a traceless theory of extraction (Pickering and Barry 1991, Sag 1998).

One thing that performance-plausible competence models of HPSG and LFG share with mainstream syntax (GB and Minimalism) is their general conception of what belongs in the grammar. Chomsky (1955) applied the notion, developed by himself and others in Formal Language Theory, of language as a set of sentences and laid out a research program of building grammars that generate the appropriate strings. In this early work, the cognitivist position is not evident. Unconcerned with issues of linguistic knowledge, Chomsky is free to define the object of study in whatever way is convenient. Theories of linguistic competence, including the performance-plausible competence models discussed here, inherited this definition of the object modeled by a grammar. However, theories concerned with actual speakers' knowledge of language do not have the freedom to arbitrarily define the object of study. To my knowledge, there is no evidence that speakers possess a separate module of knowledge limited to that which is usually included in competence grammar.

To put it somewhat differently, I see performance-plausible competence models as modeling the knowledge of language that gets used in language processing. The studies in psycholinguistics and computational linguistics cited below as well as the results reported in Chapter 4 of this dissertation suggest that people have knowledge of language that is of a type not usually included in grammaticality-focused competence grammars. This leaves two possibilities: either people's knowledge of language is partitioned into two parts (grammar and other) or it forms a unified whole. If knowledge of language is partitioned, there are several lines along which that partition could occur. In the following three sections, I will discuss three different kinds of knowledge of language that I suggest belong on the grammar side of any partition. In Section 6.5, I will discuss a few possible heuristics for determining what is in the grammar and what is not.

6.2 What's in the grammar I: Social meaning

The first additional kind of linguistic knowledge I would like to consider is 'social meaning'.[2] This was discussed to some extent in Section 5.1.2 of Chapter 5, where I argued that social meaning could reside in both individual forms and in patterns of usage. Here, I would like to discuss the content of social meaning—that is, what do socially meaningful linguistic forms *mean?*—and review some previous discussion in the literature about including social meaning in the grammar.

6.2.1 Kinds of social meaning

In order to focus this brief discussion of social meaning, I will organize it around four types of socially meaningful forms:

- Forms associated with some property
- Forms associated with some stance (polite/impolite)
- Forms associated with some situation type
- Forms like *Please*, *Good morning*, *uh huh*

These types are not necessarily clearly distinct from each other, and are not intended as an exhaustive list of socially meaningful forms. Nonetheless, they should serve to organize what follows.

[2]Some would object to this term on the grounds that all meanings are social. Although I can hardly disagree with this point, for lack of a better term, I will continue to use 'social meaning' and contrast it with 'denotational meaning', for ease of exposition.

A potential example of the first type is the fronting of (o) and (oh) in Eckert's (2000) study of the speech of high school students in the suburbs of Detroit. Eckert found that burnouts tend to lead in variables involved in changes that are more advanced the closer one gets to the urban center. Regarding this pattern, she writes:

> This could be taken as an indication of greater access to these changes stemming from the burnouts' urban-oriented life style. It can also be taken as an indication that these urban-led changes have urban-related social meaning that gives them positive symbolic value for burnouts and negative symbolic value for jocks. (2000:136)

Thus these variables have a social value that involves properties associated with 'urbanness'. Using such variables in the construction of one's own style amounts to attempting to sound like one has those properties. Another example of this same type is Ochs's (1992) example of the sentence final particles *ze* and *wa* in Japanese (discussed in Section 5.1.2 of Chapter 5). Both particles express 'intensity' (with regard to the denotational content of the utterance) but they contrast in that *ze* is associated with the property of being coarse, while *wa* is associated with the property of being gentle. Note that in this case, because of the process of iconization, a sincere use of *ze* counts as being coarse, not just as sounding coarse. A similar effect may obtain with the urban variables in Eckert's study.

The association of forms with properties of individuals may be mediated by iconic individuals in the speaker's social landscape. Eckert describes Judy, a "burned-out burnout" as "a social and linguistic icon—a local personage whose extreme embodiment of burnout practice and style serves as a benchmark of social meaning for her cohort" (2000:2).

I suspect that the social value of copula presence/absence is of this first type: it is associated with properties of individuals that speakers may then use in the construction of their own personal styles, etc.

The second type of social meaning involves forms associated with a stance, typically a stance vis à vis the addressee. Thus addressee-oriented honorifics as well as solidarity forms fall into this class. The boundary between this type and the last is not clear cut. For example, the use of Japanese addressee-oriented honorifics not only expresses a certain relationship between speaker and hearer (i.e., indicates a stance), but may also be associated with certain (non-relational) properties of the speaker. For example, extensive use of all kinds of honorifics is associated with the speech style of middle-aged, upper middle-class women. Further, since addressee honorifics construe the addressee as socially superior, they are also potentially associated with a kind of powerlessness on the part of the speaker. In these latter functions, Japanese honorifics fall into the first type. However, one or more of these 'social meanings' may be indirectly indexed (cf. Ochs 1992). That is, if the meaning of honorifics is a stance of humbleness, their association with powerlessness might be a matter of inference from the repeated expression of a humble stance.

The third type of social meaning involves forms associated with particular situation types. For example, in Japanese, honorifics are used in certain stylized, formal situations, such as the tea ceremony. Their use in this case is independent of the actual social relationships of the participants. Or consider the speech style of livestock auctioneers in the US. This style is associated with the role auctioneer and therefore with the situation in which that role exists (the auction). Similarly, null objects in English are associated with the giving of instructions, as in recipes. I have argued elsewhere (Bender 1999) that stylistic markers like null objects are part of what makes a recipe a *recipe* and not merely a description of the making of something. Thus socially meaningful forms are involved in the social construction of these situation types (tea ceremony, auction, recipe) in at least two ways: On the one hand, they contribute to the texture of the situation. The tea ceremony is an extremely formal event, partially because of the formal language used. On the other hand, they can become a part of the definition of the situation. One hasn't properly performed the tea ceremony if one has used ordinary, non-honorific language during the event. Likewise, a livestock auction certainly isn't prototypical unless the auctioneer spoke in that particular rapid style.

Here again, the boundary between this and the other categories is not sharp. One use of honorifics in Japanese is for 'beautification' (*bikago*). In this use, the honorifics do not serve (primarily) to express politeness or formality, but rather to express a property of the speaker (see Hendry 1992 and the references cited there). Their meaning in this use is, of course, textured and complex, but part of the meaning may be achieved by linking the speaker to the elite, refined situations in which honorifics are used.

I have been considering here individual forms associated with particular situations, but of course more than one form may be associated with the same situation. Indeed, the auctioneer style mentioned above is not a matter of one feature (say, speed of delivery) but rather a whole complex including prosodic and segmental features as well as formulaic phrases (*going once...*). Such complexes constitute registers. In a theory that recognizes socially meaningful forms, one can say that the sum of the socially meaningful linguistic choices in a text (particularly those that relate to situation type) together constitute the register of that text. To the extent that the speaker's linguistic choices are influenced by the situation in the same way, there will be a certain consistency to those choices. As such, in recurrent situation types, similar patterns of linguistic choices may tend to recur, and in time the those patterns may be reified as recognizable, conventional registers.[3] In other cases, the register (in the first sense, as the sum of the linguistic choices) may be more free-form, as, for example, when a politician interrupts a stump speech with a joke, or a side comment.

The fourth type of social meaning is that attached to forms like *please*, *thank you*, *good-bye* and *uh-huh* and similar forms used to provide feedback during another speaker's turn. These forms are also associated with situations, but of a different order.[4] For example, one says *good-bye* upon parting. What is most interesting about these forms is that, while the previous types mostly concern alternations in the form of otherwise semantically contentful signs, this type involves words that have little or no other meaning. The existence of such forms provides the first argument that social meaning is a matter of grammar. As Hudson (1996) argues, to exclude social meaning from the grammar would be to deny these forms the status of signs.

[3]I leave the question of whether knowledge of these patterns should be integrated with other linguistic knowledge, and if so, how, to future work.

[4]The recipe as a situation may fall somewhere between these and the livestock auction/tea ceremony type of situation.

6.2.2 Some complications

The preceding remarks only scratch the surface of the issues around social meaning. A thorough discussion is beyond of the scope of this study. Rather, my purpose here has been to sketch the kinds of phenomena that I believe require some reflection in competence grammar. But what, exactly, should go into the grammar? Clearly, the meaning of Japanese honorifics does not include complete information about the tea ceremony, or about the particulars of the various relationships that the speaker finds him/herself in. Rather, what's in the grammar is probably something more abstract, some collection of relatively underspecified social meanings that can be used to various effects in situated speech.

Determining what those meanings are for AAVE copula absence/presence is beyond the scope of this study. The ethnographic methodology that Okushi (1997) developed and applied to the study of Japanese honorifics would seem to be a good starting point, however. Extrapolating from her study of the use of honorifics, one might posit a set of social meanings for these forms which includes 'formality', 'politeness' and 'distance'.[5] Thus the honorifics are used in the tea ceremony for their value in expressing the formality of the occasion, and the representation of their social meaning in the grammar need not include any reference to the tea ceremony. In Morgan's (1978) terms, the fact that honorifics get used in the tea ceremony is a convention of usage (and an extragrammatical fact about the tea ceremony), while the fact that honorifics can express formality is a convention of language. However, this does not mean that the use of honorifics in the tea ceremony is irrelevant to their social value. Their use in formal situations such as the tea ceremony reinforces (reproduces) their social value as formal. Another way in which the meaning-in-use of Japanese honorifics goes beyond their basic meaning is in what Okushi terms 'playful' or 'sarcastic' uses of honorifics in situations that don't warrant them. That is, speakers can achieve a playful or sarcastic tone by playing off the normal expectations for the use of honorifics.

Note that in this respect social meanings are similar to denotational or referential

[5] These are clearly related stances, and may be reducible to one underlying meaning, a question I leave to future research.

meanings. Whether or not one includes encyclopedic information in semantics (and therefore in the grammar) (cf. Langacker 1987), the meanings of ordinary words get vivified in everyday use (Sag 1992). For example, consider the pair *hear* and *listen*, which mean something like 'perceive aurally' and 'pay attention to what one perceives aurally'. In Mandarin, there is the pair *tīng* and *tīng jiàn* with similar meanings, but in this case the contrast has to do with successful perception. *Tīng* means something like 'listen for' and *tīngjiàn* something like 'succeed in hearing'. Now consider the following short discourses in English:

(22) A. Did you hear what I just said?
B. Sorry, I heard you but I wasn't listening.

(23) A. Are you listening to me?
B. Yes, but I didn't hear you until just now.

In (22), the notion of 'successful perception' appears to be associated with *listening*. In (23), it appears to be associated with *hear*. In other words, 'successful perception' is not really associated with either word in English, and depending on how their meanings are vivified in context, either can include it.

Sag (1992) provides the example in (24).

(24) Craig cut the lawn/hair/cocaine/record/rookie.

Here the meaning of *cut* is quite different depending on the object. In at least the first three cases, the sense of *cut* involves separating a single thing or mass into two parts. However, the actual actions that Craig engages in in each case differ. On hearing that Craig cut the lawn, we don't imagine that he got down on hands and knees with a pair of scissors and trimmed the blades of grass so as to make the lawn more stylish. Thus denotational meanings, like social meanings, are subject to vivification in actual contexts of use. The fact that the latter senses of *cut* (with *record* and *rookie*) may be seen as metaphorical only shows how powerful this process is.

To summarize the remarks on this first complication, social meaning, like denotational meaning, is subject to vivification in its actual contexts of use. This complicates considerably the task of deciding what exactly about the social meaning of an item belongs in the grammar, but does not detract from the point that some social meaning does belong.

Another complication that arises with social meaning is the possible specificity to particular groups of interlocutors. That is, it is plausible that copula absence/presence has some social meaning for AAVE speakers that it does not have for other groups, and that AAVE speakers are aware of this. Here again, there is a similarity with denotational meaning. Take any ordinary word that also has a use as a technical term, for example, *daughter* (of a node in a tree), *mouse* (attached to a computer), *government* (of a dependent by a head), etc. Speakers who know the technical meanings of these words probably also know which interlocutors are likely to share that meaning.

At first glance, this information would seem to be an implausible candidate for inclusion in the grammar. However, Harder (1996, 2000) presents a view of linguistic meaning as distilled out of communicative practice, where sentences do not encode messages so much as provide hints for the interlocutor to arrive at that message. Harder writes:

> In this picture, the meaning of the word *horse* is distinct from the concept HORSE per se: the word *horse* is a means of interaction, which invokes the concept HORSE for purposes of interaction—and the meaning consists in [the] act of invoking the concept. (2000:9)

If this view is right, and meaning involves a representation of the effect that using a word can have on an addressee, then such a representation could also specify the type of addressee one should expect to be so affected.

One way in which social meanings do differ from denotational meanings is that one would expect the former to be subject to more disagreement between speakers. That is, while most English speakers probably have fairly similar meanings associated with the word *horse* (or in Harder's terms, similar beliefs about what concept its use evokes), certain kinds of social meaning appear to be much more fluid, more easily molded for local stylistic purposes, and more likely to be misunderstood. This is not to say that denotational meanings are not subject to (re)negotiation, for they patently are. One need only look as far as Clinton's famous "that depends on what the meaning of *is* is." Likewise, words that have undergone semantic change provide evidence of past renegotiation: *will* (once meaning 'wish'), *mistress* (once parallel to *master*), *awesome* (once meaning 'awe-inspiring', now similar to *cool*)... Thus what we have here is not a difference in kind, but rather a difference in degree.[6]

One might think of this difference in terms of degree of conventionality. But what is convention? For Lewis (1969), conventions are norms or agreements to which each participant prefers to conform (as long as everyone else does) rather than being the sole dissenter, and which could have been otherwise. Thus, using the word *horse* to invoke the concept HORSE is a convention because any individual English speaker should prefer to maintain that association as long as everyone else does: dissenting alone would lead to being misunderstood and failing to understand other people, but there is no point in sticking to it if no one else does. Further, the convention could just as well have been otherwise, as indeed it is in other languages. Searle (1995) sees facts about language on a par with other 'institutional' facts about the world. A piece of paper with a certain kind of printing on it counts as a \$5 bill, because we collectively assign that function to it. A \$5 bill counts a money, because we collectively assign that function to it. Likewise the word *horse* can serve to evoke the concept HORSE because we collectively assign that function to it.

Lewis (1969) (see also Clark (1996)) describes the mutual knowledge involved as a recursive sequence of "you know that I know that you know...". Searle, on the other hand, argues that it is a matter of irreducibly collective intentions. This doesn't require any kind of community mind. Rather, individual speakers hold beliefs that start with "We intend that..." This notion of collective intentions is important because, as Harder (2000) points out, this is what accounts for the fact that, to speakers, the meaning of a word is not a matter of their choice, but rather a fact about the world.

Harder (2000) argues that there is still more to it, however. Conventions of language rely not only on collective intentions but also on communicative practice. It is indeed communicative practice that makes facts about language facts about the

[6]Changes in denotational meaning are often highly tied up with aspects of social meaning, and changes in both kinds of meaning only take place in socially situated interaction. This is particularly clear in the case of words like *mistress* (McConnell-Ginet 1984, 1989).

world. Harder offers the analogy of traffic lights:

> [T]he function of traffic lights would not change merely because an observer decides to withdraw his allegiance—his only chance would be to try to disrupt the actual pattern of road interaction in such a way that the causal feedback cycle itself was threatened. (2000:11)

For Harder, the causal feedback cycle is the defining feature of functions. Something has a function if it forms a part of that larger whole and it is causally implicated in the preservation or reproduction of the larger whole. In the case of language, the larger whole is the internalized systems of individual speakers. The usefulness or function of any part of that system is dependent on the speakers' experience with communicative practice up to that point. Likewise, any given use of language provides input to the next stage of the interlocutors systems. That *horse* can be used to evoke the concept HORSE among English speakers is a fact about the world. Humpty Dumpty deciding otherwise won't change that. However, as with traffic lights, Humpty Dumpty could try to disrupt or change the feedback cycle by consistently using *horse* to refer to something else. This doesn't guarantee a change, but could lead to one if Humpty Dumpty has the right social status, etc.[7,8]

So how does this apply to the difference in malleability between social and denotational meaning? One thing to note is that it is possible for speakers from different English-speaking communities to have a conversation in which the denotational semantics functions as intended without understanding the social meaning that may have also been expressed. For example, I could go to Detroit (where the vowel shift has been extensively studied by sociolinguists) and ask for and understand directions without understanding the social meanings associated with my interlocutor's rendition of the vocalic variables. The effect of this interaction on my grammar might

[7] To get *horse* to stop meaning HORSE, Humpty Dumpty would also have to introduce a new form for HORSE and hope that that catches on as well.

[8] This view of convention as rooted in practice is echoed by Eckert:

> Convention is not a thing but a process, and the possibility of convention resides in speakers' ability to hypothesize about others' behavior and to take interpretable action, along with a commitment to doing so within a particular social unit. (2000:45)

include improved recognition for words spoken with a Detroit accent, as well as an association between that accent and whatever salient properties I perceived in my interlocutor (perhaps just the fact of being from Detroit, perhaps something more related to personality characteristics). However, I would have to get to know my interlocutor and her community much better before I would learn anything about the social meaning attached to those variables for her.[9]

This is just another way of saying that we expect the broader English speaking community to share a meaning for e.g., *turn right*, but not the social meaning of variable vowel pronunciations. Where might this expectation come from? I think in this regard one cannot ignore the influence of language standardization—schooling, dictionaries, widely available written sources—in shaping those expectations. But there may well be more to it, in particular, the kind of concepts that the different kinds of meaning refer to. Why should Belten High students expect an outsider to understand the social meaning of (ay) raising, if that social meaning for them is based in a shared social landscape populated by individuals like Judy (the burned-out burnout, the social icon) whom the outsider doesn't know? On the other hand, I can expect any adult to be familiar with the notion of turning right. The in-between cases also fall into line here. Politeness is a kind of social meaning that I expect most adult speakers to know about, and indeed Brown and Levinson (1987) are able to isolate universal aspects of politeness systems. I would suspect, then, that associations between form and politeness-based social meaning would remain constant across larger groups of speakers than other kinds of social meaning. Similarly, I don't expect all adult speakers to be familiar with the concept evoked by *c-command*, and technical terms have currency only in smaller communities of use.[10]

Following this line of thought, fluidity is the flip side of locality. It is easier to change local meanings than more global ones because local meanings have a smaller mass of expectations supporting them. One way to look at it is that, with a local meaning, each individual speaker accounts for more of her interlocutors' experience

[9]Having heard my out-of-town accent, my interlocutor might have surmised all this and it may have had an effect on her performance.

[10]The phenomenon of slang would be particularly interesting to investigate from this perspective, as part of its function is to deliberately limit comprehensibility to members of the in-group.

with the item in question. Because of the kinds of concepts they involve, social meanings tend to be more local and thus more fluid. However, social meanings can be global and denotational meanings can be local, and I would predict that their relative fluidity would depend more on their locality than on the type of meaning involved.

To summarize, this section has made the following suggestions about social meaning: Social meanings boil down to expectations about the reaction of others to certain linguistic forms, and these expectations are grounded in actual practice and individual speakers' experience with that practice. Social meanings in the grammar are somewhat abstract or underspecified and can get vivified differently depending on the situation of use. The social meaning recorded for a given item can be relativized to different types of addressee. The first property is likely to be equally true for denotational meanings. As for the other two, social meanings differ from denotational meanings only in degree: social meanings may be more fluid and more subject to vivification than denotational meanings. However, it might well be that there is no clear distinction between social and denotational meanings along these dimensions but that rather they merge into each other (see again McConnell-Ginet 1984, 1989.).

6.2.3 Arguments for inclusion in the grammar

I am not the first to suggest that social meanings constitute an aspect of signs that should be included in the grammar. In this section, I will briefly review proposals and arguments in the literature.

The first argument, mentioned already above, concerns words that appear to only have social meaning (e.g., *please*, *good-bye*, *uh-huh*) (Hudson 1996:257). If signs are pairings of form and meaning, the only meaning available here is the social meaning. Exclude it, and the status of these words becomes obscure indeed.

Likewise, certain contrasts in form correspond only to contrasts in social meaning. A clear case of this is verb-form honorifics in Korean and Japanese. On Pollard and Sag's (1994) analysis of Korean referent honorifics, the contrast in form corresponds to a contrast in the presuppositions. That is, a verb with a honorific ending introduces an *owe-honor* relation into the BACKGROUND. This last is a feature within CONTEXT that, among other things, encodes the presuppositions associated with using a given sign. The manner in which such presuppositions introduced by lexemes combine to give the presuppositions of the sentence remains to be worked out, but as properties of signs, they are a matter of grammar.

I believe that the results of the experiment reported in Chapter 4 suggest that the social meaning of copula absence/presence in AAVE is an entity similar to the social meaning of honorifics in Korean or Japanese. That is, the contrast in form corresponds to a contrast in social meaning. The obvious difference is that, like much denotational meaning, the proper use of honorifics is a matter of formal study in these cultures. In this respect, the meaning of honorifics can be expected to behave more like a denotational meaning in its currency among large groups of people and in its inertia with respect to local renegotiation. However, I conjecture that a corpus-study of honorifics in either language would turn up internal constraints that, in turn, interact with the social meaning much like what I found in AAVE.

A second argument comes from the idea that the conventions of language are rooted in communicative practice. Given that social meanings are part of our experience of language, why should we expect them to be stripped away? On this point, Langacker writes:

> [U]nits are acquired through a process of **decontextualization** ... If a property (e.g., the relative social status of speaker and hearer) is constant to the context whenever an expression is used, the property may survive the decontextualization process and remain a semantic specification of the resultant unit. (1987:63)

This argument presupposes a view of language acquisition that is at odds with that of Principles and Parameters theory. Rather than seeing language acquisition as a matter of discrete parameter setting in response to triggers in the language learner's experience, the view presupposed here emphasizes gradual accumulation of knowledge of language through experience and generalization over that experience. However, regardless of the theory of acquisition of syntax, acquisition of lexical knowledge—detailed and language-specific as it is—must follow more of a gradual accumulation model. Thus this argument should go through at least for social meaning attached to individual lexical items, even if syntax is acquired more discretely.

Finally, Hudson (1996) notes that facts about who uses a form refer to the same thing (the same *sign*) as other facts about that form (e.g., its meaning, its part of speech) and wonders why the former type of fact should be treated separately. Of course, it is possible for different kinds of knowledge to refer to the same thing and yet remain distinct types of knowledge (e.g., encyclopedic and semantic knowledge both make reference to some of the same concepts). However, I believe that social meaning is linguistic in the relevant sense, and that heuristics for determining what is and isn't in the grammar will class it with other facts about signs. Some possible heuristics are discussed in Section 6.5 below.

6.3 What's in the grammar II: Minimalist v. maximalist syntax

In the previous section I argued that social meaning can attach to signs and therefore should be considered a matter of grammar. In the case of AAVE copula absence, social meaning appears to be attaching to a lexical sign—(overt) forms of the copula—but with differential intensity depending on certain properties of the complement of the copula. Now, as argued in Section 2.5 of Chapter 2, the distributional evidence supports positing entries for the copula that are underspecified with respect to the part of speech of the complement. This gives us two options: a general entry with an intensity of social value that is a function of the part of speech of the predicate, or more specific entities that specify the part of speech of the predicate and the intensity of the social value. Because of the relationship of the intensity to frequency (discussed in Section 6.4 below) the second choice appears to be more promising. This section briefly reviews evidence in the literature regarding other phenomena that seem to require the addition of extra, overspecified types to the grammar.

6.3.1 Usage-based model

Langacker (1987, 1990, 2000) proposed a model of linguistic knowledge that he calls the usage-based model (see also Kemmer and Barlow 2000). This is a sign-based model in that all linguistic units are pairings of form and meaning. On this model, such units are extracted by generalizating over experience and then recursively generalizing over the units extracted. Langacker writes:

> The grammar lists the full set of particular statements representing a speaker's grasp of linguistic convention, including those subsumed by general statements. ... Speakers do not necessarily forget the forms they already know once the rule is extracted, nor does the rule preclude their learning additional forms as established units. Consequently, particular statements (specific forms) coexist with general statements (rules accounting for those forms) in a speaker's representation of linguistic convention, which incorporates a huge inventory of specific forms learned as units (conventional expressions). Out of this sea of particularity speakers extract whatever generalizations they can. (1987:46)

This is what might be called a maximalist model of grammar. In HPSG terms, a minimalist grammar would include only those types that are necessary to generate all and only the sentences of the language. A maximalist grammar would include further subtypes that are more specialized. Such a conception of grammar goes against the common conception of theoretical parsimony in linguistics, which favors 'rich deductive structure', i.e., maximally simple underlying systems from which all of the data be can produced, although sometimes only with long chains of reasoning. However, as Henderson (1989), Jurafsky (1996) and others have noted, psychological models tend to view storage as relatively cheap and processing as relatively costly.[11] Now, sign-based grammars (including HPSG) already tend towards greater numbers of stored units (compare constructions in sign-based frameworks to the simple, uniform, phrase structure backbone of GB/Minimalism). The usage-based model takes the next step, including (partially) prefabricated pieces as it were: chunks of linguistic structure that could be built up out of smaller pieces already existing in the grammar, but that

[11] See also Bresnan 1978:14.

are frequent enough and useful enough to have around in addition to the minimal types.

Again in HPSG terms, the general picture is as illustrated in Figure 6.1. In this schematic picture, the top part represents the minimal grammar—those types that are absolutely required for descriptive, generative adequacy. However, the more specified types can also fit into the same hierarchy as further specializations of the existing types. Examples are frequently used phrases like *I don't know* (discussed more below), the canonical forms of idioms and collocations (see Riehemann forthcoming), and the complement-specified subtypes of the copula proposed here.

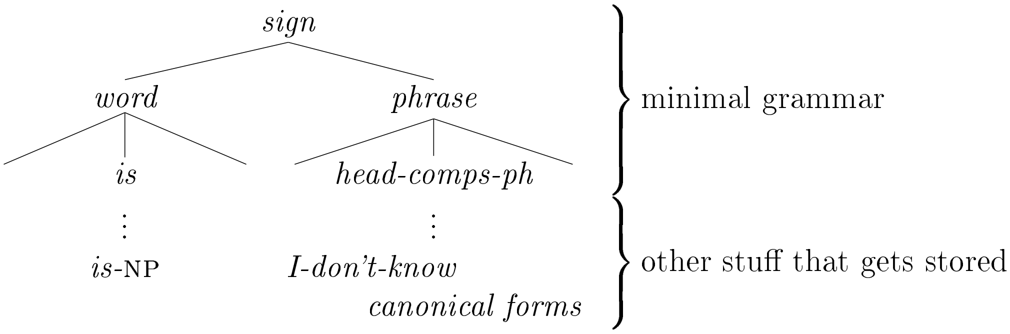

Figure 6.1: Extended grammar

Now, it should be possible in principle to "prune" the hierarchy shown in Figure 6.1 to produce the minimal grammar. However, since all of the information in this hierarchy is represented in a uniform format, it is difficult to see why speakers would partition their knowledge of language along that pruning line. Thus if speakers can be shown to have knowledge of more specific types, as is predicted by the usage-based model, then there is reason to believe that that knowledge is a matter of grammar. The next subsection explores some evidence that speakers have this kind of knowledge.

6.3.2 Linguistic evidence

The evidence for storage of extra types falls into three general classes: the existence of collocations, distributional evidence, and evidence from variation across speakers. Each will be treated briefly in turn here.

To begin with the question of collocations, it would seem that their very existence is evidence of storage of redundant types in the grammar. Examples include *avoid* x *like the plague*, *everything you always wanted to know about* x *but were afraid to ask*, and *get up on the wrong side of the bed*. These phrases all sound familiar, and indeed are listed in collocation dictionaries.[12] Further, phrases that are similar to but not the same as these collocations sound odd, or at least distinctly less typical. Compare, for example, (25) and (26).

(25) get out of bed on the wrong side

(26) avoid *x* like a tax audit

(25) is actually the form that Kirkpatrick lists for this collocation, but it sounds distinctly wrong to me. Perhaps this collocation is subject to regional variation in form (Kirkpatrick lives in Scotland), but the important point is that while (25) certainly means the same thing as *get up on the wrong side of the bed*, it is possible to recognize one and not the other as the usual way of saying it. Similarly, (26)[13] contrasts with *avoid* x *like the plague* in that the former sounds fresh while the latter sounds clichéd. Anyone making such a judgment must have a representation of these phrasal chunks. But such a representation would be redundant in a minimal grammar because the phrases can certainly be built up out of smaller pieces.

An example of distributional evidence for extra types comes from Bybee and Scheibman's (1999) production study of the reduction of *don't* to /ɾõ/ or /ə̃/. They found that this reduction occurred by far the most frequently in common, fixed phrases containing *don't*, such as *I don't know* or *why don't you*. From a strictly distributional point of view, this indicates that these phrases must be stored as linguistic units in order to be the domain of application of the *don't* reduction rule. Further, if the common phrases are stored as units, the fact that *don't* reduction occurs here and not elsewhere (or, eventually, here more than elsewhere) acquires a natural explanation: If these units constitute single production routines that are recognizable as single chunks by addressees, there is less pressure to articulate each

[12] In particular, Betty Kirkpatrick's *Clichés*, New York: St. Martin's Griffin, 1996.

[13] An authentic example collected by Susanne Riehemann (p.c.).

piece clearly. Further, frequent use of these units on the part of speakers could result in phonetic change that doesn't spread to other instances of the word *don't*.[14]

The evidence from variation across speakers comes from Zwicky's (1982) study of stranded *to*—instances of complementizer *to* separated from its VP complement by a parenthetical or by ellipsis of that complement:[15]

(27) a. I made a decision whether to — and he made a decision, too — have him in the group.
(John Lennon, *Rolling Stone Interviews*, p.150)

b. ... I can put my dukes up now if I have to in life.
(Joni Mitchell, *Rolling Stone Interviews*, p.389)

Zwicky presents the results of a grammaticality study involving 74 native speakers in which he finds that there is considerable variation in the constraints on stranding *to* in this manner. While all speakers tend to agree on the core examples (i.e., examples of patterns with high frequency in text data), they disagree on the less frequent cases. For example, some, but not all speakers, require the constituent before stranded *to* (the host it leans on phonologically to when its VP is absent), to be either a VP or a predicator within a VP. The examples in (28)[16] don't meet this constraint, and are accepted by only some of the participants in Zwicky's study.

(28) a. %I really need someone to fix the radiator, so it was awfully nice of Helen to.

b. %I realized that you sometimes stay out past midnight, but it seems excessive *always* to.

c. %When they wanted someone to make dinner for 30, they persuaded Quentin to.

Zwicky observes that this 'dialectal' variation does not correlate with social or geographical factors. Instead, he suggests that the analysis of this phenomenon is "underdetermined by the data available to the child acquiring English" (p. 51), and

[14]Further evidence for the existence of such over-specified types comes from the fact that social meaning can attach to them. For example, Penelope Eckert (p.c.) suggests that /aõno/ is a socially meaningful way of pronouncing *I don't know*.

[15]The examples in (27) were collected by Zwicky 1982:9.

[16]From Zwicky 1982:39.

that different speakers generalize from the available data differently. Some end up with very conservative generalizations, in effect requiring all of the properties shared by the frequent, core examples in an example they will judge acceptable. Some generalize further, and in divergent ways.[17] This state of affairs is exactly what one would expect given a model in which speakers slowly abstract generalizations from linguistic experience.

Further, Zwicky notes even speakers who reject the sentences in (28) above accept one class of exceptions to the constraint that rules out (28a–c), namely, sentences in which *to* attaches to monosyllabic function words (29).[18]

(29) a. I want to calculate the bill, but I don't know HOW to.
b. You must write a thank-you note, because not to would be impolite.
c. George didn't do it, until Kim persuaded him to.

By way of explanation for how this state of affairs could have come about, Zwicky writes:

> If a nonlexical item [= function word–EB] frequently occurs in combination with some word, especially a word like *to* with very idiosyncratic distribution, then the combination can easily be interpreted as a unit on its own. Then, when unquestionable combinations of word-plus-*to* become available via the application of [VP ellipsis] after verb-*to* sequences, other closely knit word-*to* sequences are natural targets in an extension of the domain of [*to* stranding]. (1982:45)

These 'word-*to*' sequences are just the kind of prefabricated chunks that are included in a maximalist view of grammar.

6.3.3 Summary

This section has briefly reviewed some theoretical and empirical arguments for including extra, overspecified types in the grammar. One consequence of this is that

[17]There are, however, boundaries to how permissive a grammar speakers will construct. Zwicky argues that these are provided by universal grammar.

[18](29a–b) are from Zwicky 1982:14,26.

non-local trees might get stored, as in the case of *I don't know* and *why don't you*.[19] This raises certain issues that I will return to in Section 6.5.5 below.

6.4 What's in the grammar III: Probabilities

In this section I will discuss two different kinds of probabilistic or frequentistic information that speakers appear to have knowledge of, postponing arguments about whether one or both belong in the grammar until Section 6.5 below. The first kind corresponds to what might be seen as the resting activation of various linguistic elements (such as alternative word senses, valence patterns, or constructions). This kind of information plays a role in processing, but can be overridden by contextual constraints. The second kind of probabilistic information corresponds to the non-categorical constraints on sociolinguistic variation studied here. It is not immediately clear whether these should be treated in the same way, or if they have differing status.

6.4.1 Resting activation

MacDonald et al. (1994) and Jurafsky (1996) argue that a number of processing phenomena can be best captured by models in which the processor makes use of frequentistic information associated with signs. For example, MacDonald et al. (1994) argue that the relative frequency of different argument structures of verbs can influence the processing of main verb/reduced relative ambiguities. They report the results of an experiment by MacDonald (1994) in which reading times of sentences like (30a–c) were compared.

(30) a. The rancher knew that the nervous cattle *moved* into the crowded pen were afraid of the cowboys.

b. The rancher knew that the nervous cattle *pushed* into the crowded pen were afraid of the cowboys.

c. The rancher knew that the nervous cattle *driven* into the crowded pen were afraid of the cowboys.

[19] Such non-local trees are not, however, necessary for the treatment of copula absence/presence proposed here.

In (30a) and (30b) there is an ambiguity at the italicized word. That word could either be a main verb, taking the NP *nervous cattle* as its subject, or the head of a reduced relative modifying *nervous cattle*. The rest of the sentence provides disambiguating information in favor of the reduced relative reading. Note that in the reduced relative, these forms are passive participles. (30c) is unambiguous because *driven* can only serve as a participle and not as a past tense main verb. (30a) and (30b) differ in that *move* occurs more frequently as an intransitive verb, while *push* occurs more frequently as a transitive. This difference is relevant here because the passive, reduced relative reading requires a transitive argument structure. Thus MacDonald (1994) predicts that (30a) should have more of a garden-path effect than (30b). This prediction is borne out: reading times at the point of disambiguation (*were afraid*) were reliably longer with the *move* type verbs than the unambiguous control sentences (like (30c)), while reading times at that same point with the *push* type verbs do not differ from the control sentences. This result shows an effect of relative frequency of argument structures on processing; that is, it shows that speakers have knowledge of this frequentistic effect and bring it to bear in comprehending sentences. MacDonald et al. (1994) make similar points regarding PP attachment ambiguities and NP/sentential complement ambiguities.

Jurafsky (1996) reports on a computational model of human natural language processing. This model incorporates probabilities attached to constructions and valence patterns. The probabilities associated with each construction and valence pattern used in a parse can be combined to calculate the probability of the entire parse. The parser entertains several possible parses in parallel, but prunes any that fall below some threshold of probability. This pruning is motivated as a means of keeping the computation tractable in the face of the massive ambiguity of human language. With this model, Jurafsky provides a uniform account of frequency effects in the access of lexical items, idioms, and constructions, differing access points in different idioms, attachment preferences (e.g., PP attachment), and garden path effects related to frequency of constructions, valence patterns, and both together.

Now, these frequency effects are such that they can be overridden by context. For example, the strong garden path effect in (31) is partially due to a preference for *raced* to appear as in intransitive.[20]

(31) The horse raced past the barn fell.

However, this sentence can be rendered interpretable, and thus the preference can be overridden by an appropriate context:[21]

(32) The horse that was raced down the hill got there just fine. The horse that was raced along the river had some trouble. And the horse raced past the barn fell.

This property of defeasibility or interaction with context suggests a possible distinction between grammatical and other knowledge of language. According to this distinction, grammatical knowledge is that which remains relatively constant, while frequentistic information that can be so radically modified by context belongs to a separate ('performance') component of knowledge. However, here again the distinction is not so clear cut. For example, in the discussion of vivification above, it was noted that the meaning of words also interacts rather dramatically with context. In that case, it seems fairly clear that the underlying, more abstract meaning is a matter of grammar, and the more specific meanings might not even be a matter of knowledge at all; rather they are context-specific effects computed on the fly. If this frequentistic knowledge is analogous, then the resting probability or activation would be a matter of grammar and the actual activation in some context of use computed on the fly. My main purpose in this section, however, has been to introduce this kind of knowledge so that I may ask whether non-categorical constraints belong to the same type.

6.4.2 A tentative formal proposal

In order to consider the status of non-categorical constraints, it will be helpful to spell out a tentative model. This section will do so, building on the work of Jurafsky (1996) and Kemmer and Israel (1994).

Kemmer and Israel (1994) present an analysis of grammatical and social constraints on t/d deletion within the usage-based model. The essential properties of

[20] Other factors include the plausibility of *horse* as the subject of intransitive *race*.

[21] This type of example is due to Crain and Steedman (1985), who are in fact arguing against a probabilistic model.

their analysis (for present purposes) are as follows: All common words subject to variable t/d deletion have two representations in speakers' grammars. The existence of many such pairs leads to the abstraction of a general schema of t/d deletion. However, here we will concentrate on the individual words. In any given pair, say /læs/ and /læst/ for *last* the more frequent alternant will be more entrenched, that is, have higher resting activation. Both forms compete for selection when the speaker goes to pronounce the word, but the more entrenched form tends to win out, in proportion to how entrenched it is. However, both forms may be associated with information about the social situation, etc. If the characteristics of the social situation 'match' the specification on the less frequent item,[22] then the less frequent item will be promoted with respect to the more frequent item, and vice versa.

Note that this approach involves the same kind of random application of probabilities as the OT and Variable Rules models considered in Chapter 5. As such, it is not possible for it to model social value associated with individual instances of linguistic variables. Any given instance of the less likely alternative could either be the result of the random application coming up with the less likely alternative, or of the speaker tilting the scales in favor of that alternative by a desire to express the associated social value. The listener would have no way to tell which it was. Nonetheless, the idea of associating social value with signs, and letting that social value play a role in the selection is promising.

An interesting feature of Jurafsky's (1996) model of parsing is that it uses probabilities in a more deterministic way. Essentially, the probabilities of all of the elements of a parse, resting and context-dependent, are combined to give the overall probability for that parse. The same is done for all of the active parses under consideration, and the most probable parse is chosen. Although the system described in Jurafsky 1996 is a parser, Jurafsky has suggested (p.c.) that the same ideas could apply in production. In this case, given a communicative intent, the speaker selects the means of conveying that intent with the highest probability. On the face of it, this model

[22]Or, more accurately in my opinion, if the speaker wishes to do his/her part in constructing the social situation to match that specification.

would countenance no variation; only the most frequent variant would ever get produced. However if we let the variants be associated with social value, as in Kemmer and Israel's (1994) model and as suggested in Section 6.2 above, then that social value could be part of the communicative intent.

What I have in mind is a system in which the desire to express the social value associated with the disfavored variant can override the bias against that variant. However, the stronger the bias against the element, the more the speaker has to want to express the social value in order to overcome it. Since copula presence is more disfavored with V+*ing* complements than with NP complements, it takes more desire to express the social value of copula presence to use it in the V+*ing* environment. Hearers, knowing this, would react to copula presence accordingly, as they did in the experiment.

There are several complications here. First, it is clear that the probabilities involved are not a once-and-for-all matter, but can be relativized to the interlocutor. The probabilities a speaker uses in production are not necessarily the ones s/he uses in comprehension. A case in point are the non-AAVE speaking African Americans in my study. These speakers showed knowledge of the probabilities of AAVE speakers in their evaluations of the stimuli, but presumably do not apply those frequencies in their own speech. A second complication has to do with the possibility that the same speaker could have different resting probabilities based on who they're talking to, i.e., based on their expectations about how their speech will be heard by that listener.

One prediction of this model is that only the disfavored variant can be used to express social value, a following *ing*-form verb favors copula presence. Interestingly, the results of the experiment seem to indicate that only one variant was meaningful for each group showing an effect (i.e., AAVE speakers and other African Americans). At least, the effect of the following grammatical environment was only apparent for copula presence for the AAVE speakers and copula absence for the other African Americans. However, I don't expect it to be the case that only one variant of any given variable is ever meaningful. I see two other possibilities within this model. The first is that the favored variant is indirectly socially meaningful in that its use indicates a lack of sufficient desire to express the social meaning associated with the disfavored variant. As such, use of the favored variant expresses a weak negation of that social meaning. However, the grammatical environment should have no effect, as it would not be possible to tell how far the speaker's attachment to the social meaning fell short. The second possibility is that other constraints (such as information about the context and the interlocutor) could interact to change which variant is favored, for example, copula presence might be the favored variant when talking to a priest, teacher or other community leader. The grammatical constraints could remain constant in this situation. That is, a following NP favors copula presence more strongly than a following *ing*-form verb. Conversely, a following *ing*-form verb favors copula absence more strongly than a following NP. However, the social value would switch to the newly disfavored variant (copula absence, in this example). Note that these two possibilities (social meaning through negation of the meaning of the other variant, and context-based changes in which is the favored variant) are not incompatible with each other.

The preceding paragraph presumed that there is one, fixed social meaning associated with any variable that speakers can either express with one variant or refrain from expressing by using the other. However, this assumption is not a necessary one for this model of non-categorical constraints. All that is required is that sociolinguistic variables have some social meaning, which may be continually renegotiated. What remains (relatively) constant is the effect of the grammatical environment on the social meaning.

There is a further complication in that non-categorical constraints can in principle be so different or so strong in their effect, that one variant is preferred in one grammatical environment while the other is preferred in another. Indeed, in most of the data sets represented in Table 4.1 of Chapter 4 (see page 144), the V+*ing* environment has a Varbrul weight of over .5, favoring copula absence, while the NP environment has a Varbrul weight of under .5, disfavoring copula absence. However, these weights interact with other factors (e.g., whether or not the subject is a pronoun) and with the input probability, so that it is not possible to conclude from these numbers alone which variant is preferred, and whether that differs across grammatical environments.

For concreteness, I will present a tentative formal representation of my proposal.

However, I regard the preceding discussion as speculative, and offer this formalization only for purposes of illustration. Throughout the discussion, I have been considering four sentence types: copula presence with nominal and verbal predicates, and copula absence with nominal and verbal predicates. However, the grammar doesn't usually provide the choice of nominal or verbal predicates with (roughly) the same meaning,[23] so it is actually forms that differ only in presence/absence of the copula that are competing. This means that the two entries in Figure 6.2 are in competition with each other, as are the two entries in Figure 6.3. In these entries, social value is associated only with copula presence, and indicated by the word 'educated' as a short hand for something that is undoubtedly more complicated. The numbers to the left of the avms are meant to represent differing degrees of resting activation (varying between 0 and 1). It is these values that are used in computing the probability of a parse. The actual values shown in the figure are there only for purposes of illustration, and not based on any particular data set.

$$\mathbf{.4}\begin{bmatrix} \textit{copula-be} & \\ \textsc{comps} & \langle\ \text{NP}\ \rangle \\ \textsc{ctxt} \mid \textsc{social} & \textit{`educated'} \end{bmatrix} \quad \mathbf{.6}\begin{bmatrix} \textit{silent-copula-ph} & \\ \textsc{args} & \langle\ \text{NP}\ \rangle \end{bmatrix}$$

Figure 6.2: Competing entries for copular sentences with NP predicates

$$\mathbf{.2}\begin{bmatrix} \textit{copula-be} & \\ \textsc{comps} & \langle\ \text{VP}[\textit{prp}]\ \rangle \\ \textsc{ctxt} \mid \textsc{social} & \textit{`educated'} \end{bmatrix} \quad \mathbf{.8}\begin{bmatrix} \textit{silent-copula-ph} & \\ \textsc{args} & \langle\ \text{VP}[\textit{prp}]\ \rangle \end{bmatrix}$$

Figure 6.3: Competing entries for copular sentences with V+*ing* predicates

According to the model sketched here, all else being equal, copula absence would prevail in both cases. However, should the speaker desire to express the social value of copula presence (e.g., should the speaker want to sound educated), there is the option of using an overt form of the copula. In the case of an NP predicate, the two

[23] Indeed, the test stimuli used in the experiment had to be carefully constructed.

options are relatively close in their activation, and so it doesn't take much to override the resting probabilities. In the case of the V+*ing* predicate, however, there is a much stronger bias towards copula absence and therefore the speaker must be much more emphatic about the social value of copula presence to override this bias. Listeners who have experience with this pattern (and have reason to believe that the speaker is working with the relevant set of probabilities) can interpret this as copula presence with V+*ing* sounding more educated than copula presence with an NP complement.

6.4.3 Interacting factors and context effects

Ivan Sag (p.c.) suggests that one way of partitioning knowledge of language would be into persistent, relatively context-independent knowledge (i.e., grammar) and a component of more context-sensitive, volatile information including the resting activation of various signs. The purpose of this subsection is to consider whether non-categorical constraints are on a par with the kind of resting activation behind garden path sentences. Now, I've just sketched a model of non-categorical constraints in terms of resting activation. However, it is still possible that these constraints interact differently with the context than do other types of bias.[24]

It is first important to note that the experiment reported in Chapter 4 and experiments such as that reported in MacDonald 1994 are not parallel in at least one important respect. MacDonald's (1994) experiment is aimed at investigating how knowledge of probabilities is brought to bear in ambiguity resolution. The experiment reported here, on the other hand, is aimed at investigating how probabilities (or frequency or markedness) affect the social meaning of utterances. That is, the former is a matter of choosing between two alternative structures with their attendant meanings while the latter concerns the meaning of a single structure. When I talk about the speaker's desire to express the social value overriding the grammatical constraint, the latter does not simply disappear but instead has an enhancing effect on the social value.

What about other contextual factors that could interact with the non-categorical

[24] If this is so, a better model of non-categorical constraints might distinguish their implementation from that of other statistical knowledge.

constraints studied here? There are probably numerous factors that enter into the choice between variants in actual production. So far, I have been concerned only with social value and the constraining effect of the grammatical environment (non-categorical constraints). But surely there are also factors, such as speech planning issues, where one might expect to find overt forms of the copula potentially used as fillers in case of disfluency, and recency effects, where a previous use of one form or the other could prime that form for use again. The question is, then, if the non-categorical constraints retain their effect on the social value regardless of these other contextual factors. If they do, then they appear to be distinct from other kinds of statistical knowledge—that is, more persistent, more like other familiar (categorical) grammatical constraints. The resolution of this question will require a more sophisticated model of production and more sensitive experimental techniques.

6.5 The boundaries of grammar: Some heuristics and examples

In the preceding three sections, I have suggested that speakers have knowledge of three kinds of linguistic information not usually included in competence grammars—social meaning, overspecified types, and probabilities of a certain kind—and I explored possible evidence for including this knowledge in grammar. However, by opening up the boundaries of competence grammar to include these kinds of information, I do not mean to imply that grammar is essentially boundless. Clearly, there are kinds of knowledge that are not grammatical. I would like to propose that grammar includes just that knowledge of language that is used in linguistic processing.[25] Presumably, one could devise psycholinguistic tests to determine in any given case whether or not people have knowledge of the information in question and whether they use it in language processing. Here, however, I would like to discuss some general heuristics and in the process sketch a general picture of what I take to be the domain of grammar.

[25] The possibility of partitioning out context-sensitive information will be returned to below.

6.5.1 Arbitrariness/language specificity

One fairly standard criterion for including something in a competence grammar is that of arbitrariness or language specificity (in the sense of specific to a given language). That is, linguistic facts that are conventional in Lewis's sense—they could have been otherwise—must therefore be learned and in the grammar. This criterion is meant to distinguish grammatical facts from natural consequences of functional pressures. It leaks, however, in several respects.

The first problem is it really only works as a sufficient (not necessary) condition for inclusion in the grammar. That is, linguistic facts such as that the word for HORSE in French is *cheval* can be shown by this criterion to be in the grammar, but the non-arbitrariness of other facts does not preclude their grammaticization. That is, just because a pattern could follow from independent principles doesn't mean that speakers don't learn it as a pattern. Indeed, this process is probably what is behind many cases of grammaticization, for example, the development of tones from the different effects of voiced versus voiceless consonants on the formants of vowels (Maddieson 1984). Another example is the tendency to name men before women in when naming couples (Wright and Hay 2000). While this tendency may come from a cultural fact about men being more salient, and thus named first, it is grammaticized in *Mr. and Mrs.* (cf. #*Mrs. and Mr.*) and partially grammaticized in *he or she* (compared to *she or he* which is used with intent to make a point about sexism in language).

The second problem, related to the first, is that language is not as arbitrary as linguists often make it out to be (Langacker 1987). Examples of non-arbitrariness include sound symbolism of the type evidenced by word-sets like *slime, slippery, slink, ...* and morphologically complex words that have drifted just a little from their original semantically compositional meanings. Thus although the full meaning of a word like *fire engine* can not be predicted from its parts, given the meanings of *fire* and *engine*, it is not entirely arbitrary that they should be used together with the meaning of *fire engine*. The same applies for fully compositional words. Langacker writes:

> An obvious but seldom-made observation is that any polymorphemic linguistic sign (this includes the vast majority of expressions) is nonarbitrary to the extent that it is analyzable. For example, given that *staple* means what it does, and that *-er* means what it does, it is anything but arbitrary that *stapler* is the form used in English for a stapling device. (1987:12)

A third problem arises from the process of iconization. If Irvine and Gal (2000) are right, then the connection between form and social meaning is non-arbitrary FOR SPEAKERS CONCERNED, even though it could well have been otherwise. They write:

> *Iconization* involves a transformation of the sign relationship between linguistic features (or varieties) and the social images with which they are linked. Linguistic features that index social groups or activities appear to be iconic representations of them, as if a linguistic feature somehow depicted or displayed a social group's inherent nature or essence. This process entails the attribution of cause and immediate necessity to a connection (between linguistic features and social groups) that may be only historical, contingent, or conventional. (2000:37)

For example, for many English speakers, copula absence and other features of AAVE are iconic of lack of education (or worse, low intelligence) in that they are believed to show or symbolize it directly. Indeed, this kind of social value probably arises as an extension of the stigmatization (by outsiders) of AAVE speakers to the stigmatization AAVE features (again, in the minds of those outsiders). Thus, this kind of social meaning is non-arbitrary in its inception, although it is still arbitrary in the sense that it could have been otherwise. (For example, if AAVE had a different collection of constructions, or if AAVE was not stigmatized but rather exalted.)

To summarize, arbitrariness is at best a one-way criterion for inclusion in the grammar. Many grammatical facts are either not completely arbitrary or not experienced as arbitrary by speakers.

6.5.2 Structure: A system of contrasts

Some authors (e.g., Langacker 1987, Langacker 1990, Hudson 1996) ask, Why should grammar be a separate module of knowledge? They suggest instead a view in which linguistic knowledge is embedded in other knowledge. I think that the motivation for distinguishing (some part of) knowledge of language (the grammar) comes from the fact that the system of contrasts gives internal coherence to grammar, even if it is embedded in a larger knowledge system. One might think of a mosaic, with a white background and the picture picked out in black as an analogy. Although both background and picture are made of tiles, the tiles in the picture together form one coherent system.

This criterion can certainly be interpreted as placing social meaning in the grammar. Social meaning, like denotational meaning, exploits contrasts in form. Further, Irvine (in press) argues that style is a matter of distinctiveness, i.e., contrasts between different groups of people in the social landscape. As building blocks of style, socially meaningful linguistic elements thus pair contrasts in form with contrasts in social space.

One might also be tempted to talk about certain probabilities as a matter of contrasts. That is, copula presence and copula absence are necessarily complementary in their frequency of occurrence because the variable is defined in such a way that choosing one precludes the other. A similar thing could be said for frequencies of valence patterns and the like. However, in this case, it seems that the contrast in frequency merely follows from the contrast in form or in valences. As far as I know, nothing else builds on the contrast in the way that contrasts in social space build on or attach to contrasts in form.

6.5.3 Creativity of use

Another possible criterion is generativity in the sense of creativity of use. Knowledge of syntax and semantics allows speakers to create and understand novel utterances. If other knowledge about language functions in a similar way, by this criterion it should be in the grammar. The California Style Collective (1993) refer to individuals creating their styles as *bricoleurs*, a French word which evokes a picture of building something out of whatever resources and raw materials that one can find. On this view, individuals create their (linguistic) styles by borrowing and combining symbolic resources made meaningful by their use by and association with others. Socially meaningful forms can thus combine to create novel styles in a manner that is analogous to the way that lexemes and constructions combine to create novel utterances.

It is somewhat less clear how to apply this criterion to non-categorical constraints and overspecified types, but it might be seen to suggest that both belong in the grammar. As for non-categorical constraints, inasmuch as their effect on social meaning allows them to be a tool in the construction of style, they should be considered grammatical for the same reason that social meaning is. With over-specified types, possible evidence for including them by this criterion could come from evidence of people making reference to collocations in the creation of new utterances. Merely using collocations doesn't provide any evidence with respect to this criterion as, by definition, collocations are semantically compositional and can be built up with smaller pieces. However, making reference to collocations means involving them in some sense in the production of novel utterances. An example would be Kraft's marketing of a macaroni and cheese product as *cheese and macaroni*. The unusual placement of *cheese* as the first conjunct is meant to suggest especially cheesy macaroni and cheese. This effect of this suggestion comes partially from the contrast with the usual way of saying it, that is it relies on the knowledge that one usually says *macaroni and cheese*.

6.5.4 Heideggerian *thrownness* of use/acquisition.

Winograd and Flores provide another possible heuristic:

> [T]he essential feature of language activity (the processes of saying and listening) is the thrownness of a person within language. When we are engaged in successful language activity, the conversation is not present-at-hand, as something observed. We are immersed in its unfolding. Its structure becomes visible only when there is some kind of breakdown. (1986:68)

Winograd and Flores define Heideggerian *thrownness* as "the condition of understanding in which our actions find some resonance or effectiveness in the world"

(1986:32). Thrownness contrasts with detached contemplation. Isolated and categorized representations of objects are creatures of detached contemplation. Most language use, however, is entirely transparent. In conversation people pay attention to the messages they are exchanging and not the details of how they are said. Indeed, it takes some training for linguists to learn to pick out examples of phenomena of interest to us from the flow of language use that we encounter in our everyday lives. A "breakdown", in this context, is some kind of problem in the functioning of a tool, such as language. Experiencing a misunderstanding can cause even non-linguists to focus on the words and structures used, to produce and categorize (conscious) representations of the problematic linguistic elements.

By this criterion, the three types of knowledge described in Sections 6.2 to 6.4 above are similar to (other) grammatical knowledge. As Ochs (1992) notes, we learn, comprehend and deploy socially meaningful contrasts in language without paying much attention. That is, it is enough to decide to be informal, or to be polite, or be distant. As competent speakers, we need not pay attention to the individual linguistic choices that implement such decisions. Likewise, we acquired the relevant knowledge with the same seeming effortlessness as the rest of our grammars. Similarly, whatever frequentistic information we've stored is clearly not the result of formal study or conscious tabulation (cf. Haser and Chromiak 1977, discussed in §5.2.1 of Chapter 5). Indeed, part of the reason some authors have doubted that people store frequentistic information is the sheer difficulty of learning it by these other means. Finally, although sayings and proverbs are studied in school in some cultures (notably in Chinese-speaking cultures), this is not the case for the extra types representing such collocations as *avoid like the plague*.

What this criterion does rule out is knowledge of language that is not used in language processing. For example, take the fact that *honcho* in *head honcho* is a borrowing from Japanese. Speakers of English who know this fact about English know it because they were told so explicitly, or because they also speak Japanese and noticed the connection. Both cases involve focusing attention on the word in question, and thus do not constitute *thrown* acquisition. Likewise, it does not seem likely that this knowledge could naturally figure in language processing, and certainly not without treating it with the kind of detached contemplation that is inconsistent with thrownness. I picked an example from Japanese and not, say, French, because something like the knowledge that *je ne sais quoi* or *raison d'être* are borrowings from French does seem to figure more subtly into their use. French has a certain *cachet*, i.e., air of sophistication, among many English speakers, and a speaker going for this sociolinguistic effect might quite naturally reach for a French phrase. Whether this process is as *thrown* as the rest of language use is open to investigation.

6.5.5 Context sensitivity

The heuristics considered so far seem to converge on the conclusion that social meanings are a matter of grammar. To the extent that they were applicable to the question of extra types, the heuristics also support including the extra types in the grammar. The status of probabilistic information is less clear. In this subsection, I would like to return to the distinction considered in Section 6.4.3, between persistent knowledge that functions to similar effect across contexts and more 'volatile' knowledge that is more subject to modification in context. Modulo the reservations given above about semantics being more sensitive to context than one might suppose, this criterion serves to distinguish knowledge of signs (units of grammar) from probabilistic information about those signs. This subsection briefly describes a potential empirical test that could substantiate this distinction and show that in the minds of actual speakers, these kinds of knowledge constitute separate systems or modules. The test involves finding a situation in which grammar appears to be restricted and then looking for evidence that people have knowledge of language which they use in processing which is not constrained in the same way.

One potential restriction to investigate is that embodied in the HPSG theory of locality of selection (e.g., Pollard and Sag 1994). This theory holds that heads should not be able to select for arbitrary properties of words or phrases contained within their dependents. For example, it predicts that the verb labeled V1 in Figure 6.4 should not be able to constrain the case of the noun phrase labeled NP1. This constraint is enforced by two aspects of the theory: the cancellation of elements of valence lists and the selection of *synsems*. That is, since the complement requirement of the head verb of the embedded sentence in Figure 6.4 is satisfied within the embedded sentence, the embedded S bears the feature specification [COMPS ⟨⟩]. As the embedding verb (V1) selects only the *synsem* of its complement S, it has no access to the DAUGHTERS feature(s) of its complement, and thus no access to information about the object of the embedded sentence. Without this information, it can enforce no constraint on the case or other properties of that NP.

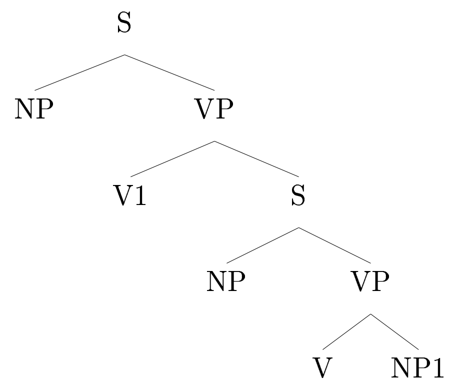

Figure 6.4: Tree illustrating locality of selection

Now, Bod (1998) and Neumann and Flickinger (1999) describe computational systems that make use of statistical information about previously encountered parses to improve disambiguation and parsing performance, respectively. The information recorded by these systems crucially involves non-local trees. By recording such trees, these systems are able to record dependencies that do not obey the hypothesized locality constraint. If people could be shown to record and use the same information, and if the locality hypothesis could be confirmed, then one would have to conclude that grammar forms a separate subsystem of linguistic knowledge.

However, there are several reasons to believe that it might not turn out that way. On the one hand, the hypothesis of locality is maintained in HPSG by using other mechanisms to make available information that is blocked by this constraint. For example, English verbs select for PPs headed by specific prepositions. Given the locality hypothesis, the identity of the preposition within the PP should be unavailable. To circumvent this, information about the head of the PP can be passed up through semantic features (identifying the preposition by the relation it introduces). However, such prepositions are often semantically vacuous, and so purely formal features are introduced (Pollard and Sag 1987). Similarly, Bender and Flickinger (1999) argue that tag questions attach to sentences (and not VPs). It follows from locality that the information about the subject of the sentence should not be available to the tag modifier. But tags patently agree with the subject, and therefore Bender and Flickinger (1999) take this as evidence that verbs should bear an agreement feature. This feature is within the SYNSEM, and therefore the requisite information about the subject is available at the level of the sentence. Other examples include the problem of including anaphors in the complement position of certain prepositions in the binding domain of the head selecting the preposition (Pollard and Sag 1994) and recent proposals by Przepiórkowski (2000) and others to propagate the ARG-ST feature to phrases. In most cases, the authors cited provide some independent motivation for considering the information that is passed up a legitimate property of the embedded head. However, the hypothesis of locality of selection appears to be weakened to the extent that it is circumvented with purely formal (or diacritic) feature passing.

On the other hand, it is not clear that people do store statistical information about non-local subtrees. Gibson and Pearlmutter (1998) claim that people need only store frequentistic information about lexical items. According to Gibson and Pearlmutter, the strongest evidence to date for the need to keep track of frequentistic information about phrase structure is the different biases towards interpreting *that* as a complementizer or demonstrative article depending on its position in the phrase structure (33).

(33) a. That cheap hotel was clean and comfortable to our surprise.
b. That cheap hotels were clean and comfortable surprised us.
c. The lawyer insisted that cheap hotels were clean and comfortable.
d. The lawyer insisted that cheap hotel was clean and comfortable.

In sentence-initial position, *that* is more likely to be a demonstrative, making (33a) easier to process than (33b). After a verb, however, *that* is more likely to be a complementizer, making (33c) easier to process than (33d). Note that these biases involve non-local trees, as shown in Figure 6.5.

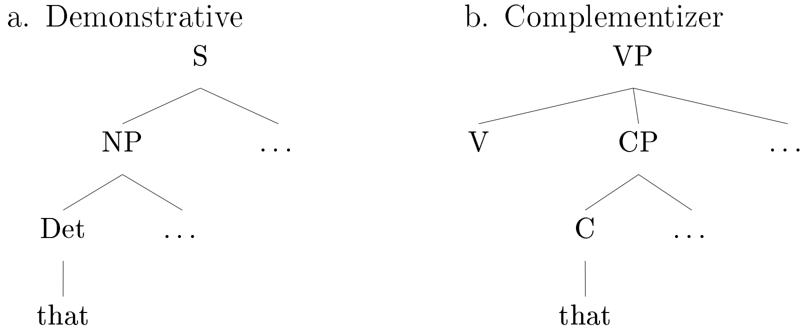

Figure 6.5: Trees encoding the dependency between interpretation of *that* and its syntactic context

However, Gibson and Pearlmutter (1998) suggest an alternative account, where memory resource factors (see Gibson 1998) would favor the article interpretation in sentence initial position. This is presumably because sentential subjects generally require the parser to posit more constituents before the main verb is encountered. The second part of their proposed account is that memory factors would not differentiate the two possibilities in the post-verbal context, and so straight lexical frequencies could apply, favoring the complementizer interpretation.

Further, it is not the case that allowing some frequency tabulation on non-local trees would necessarily violate locality irreparably. Frequency information about verbs and the prepositions they co-occur with, for example, would simply mirror one of the 'exceptions' to locality that is already countenanced. Note also that this proposal applies not only to frequency information about non-local trees, but also the suggestion of storing certain overspecified types in the grammar. For example, the phrase *I don't know*, which phonological effects suggest is a part of the grammar, spans a tree two levels deep. Likewise, idioms like *kick the bucket*, which must be in the grammar by the arbitrariness criterion, can also include multiple levels of phrase structure. The fact that some idioms are discontinuous (*put* x_i's *foot in* x_i's *mouth*) and the fact that idioms generally conform to the syntactic patterns of the language argue against giving them unusual flat structures. Thus it could be that there is some principle to the storage of non-local trees that nonetheless preserves the core predictions of locality. Perhaps such trees must be fully specified, or if they do preserve only some information about lexical items, those items must be heads.[26] This would rule out trees encoding, say, a dependency between a verb and the case of an NP buried down inside its complement.

I would like to note in this regard that Bod's (1998) arguments for preserving all of the subtrees of previously encountered parses are based on the performance of a computational system. This system does a better job of disambiguation when it has access to all of the subtrees of the training corpus. Any systematic reduction in trees (including a reduction to trees lexified only with heads such as that mentioned above) hurts its performance. What is not clear from Bod's experiments is to what extent saving all of the trees serves to compensate for machines' lack of world knowledge and other contextual information. It is well known that humans rely on both for disambiguation and they therefore may need to store less than the computer (see, e.g., Pearlmutter and MacDonald 1992). Further, even if storing more would improve human systems, it doesn't follow that we do so.

The bulk of this section has been spent on exploring how the proposed test is complicated by circumstances, but I believe that this test or another (focusing on a different kind of constraint on grammars) could in principle be used to determine whether knowledge of language of the statistical sort must be kept separate from the (rest of) grammar.

6.6 Grammar as practice

In this concluding section, I would like to suggest that the model of grammar sketched in this chapter (and similarly, the usage-based model of Langacker 1987) is a promising candidate for a theory of knowledge of language within the social theory of language that Eckert (2000) calls for. After describing the connection that I see, I will end with a brief note about the status of grammaticality judgments on this view of grammar.

[26] This won't cover all cases of idioms, however (Susanne Riehemann, p.c.).

6.6.1 Linguistic knowledge in a social theory of language

Eckert articulates some of the requirements on a theory of grammar as practice:

> It is impossible for a social theory of language to view *langue* as a pre-existing convention, for a social theory of language must be about the process of conventionalization. By the same token, it is impossible for a social theory of language to view the individual speaker's competence as a simple internalization of convention. Convention and individual competence are mutually produced and reproduced in practice, thus linguistic practice is not simply the consensual use of a common system. Convention is not a thing but a process, and the possibility of convention resides in speakers' ability to hypothesize about others' behavior and to take interpretable action, along with a commitment to doing so within a particular social unit. Our speaker, or speaking subject, can not be a clone but must be an agent in a process of convention-making. (2000:45)

All of the suggestions for changes to the conception of grammar presented in this chapter can be seen as following from the view that convention is grounded in practice. If speakers' knowledge of language derives from (and is constantly fed by) their experience of situated language use, then the social aspects of those situations should be able to attach to particular signs. Further, the claim that convention is a process includes the idea that language is always changing. Each utterance potentially redefines or reinforces the system of the interlocutors attending to it. If Eckert and others are right (and they seem to be) that change in language structure is driven by the social actions that speakers carry out through their choice of forms, then the social meaning of linguistic forms must be represented somehow. Further, if individual utterances are what slowly reinforce or redefine the grammar, then the possibility arises that speakers store both overspecified types (e.g., common utterance types) and frequentistic information.

Note however, that this view of grammar is not meant to be in itself a social theory of language. Nor is it a theory of the process of conventionalization. Rather, it is a model of what linguistic knowledge should look like, given that "convention is not a thing but a process."

6.6.2 Problematizing grammaticality judgments

In shifting the definition of grammar away from concerns about grammaticality, it seems appropriate to reconsider grammaticality judgments as well. Langacker writes:

> The notion of syntax as an autonomous formal system has encouraged the expectation that speakers should be capable of simple categorical judgments (grammatical/ungrammatical) on the well-formedness of sentences, out of context and without regard for semantic considerations: either a sentence meets all formal specifications, or it does not. (1987:36)

Grammaticality judgments have been conceptualized/naturalized as a property of the brain, partially hidden from observation by performance factors. Linguists and, to some extent, people we train as consultants share a set of beliefs that allow us to ask and answer these questions. It does not follow that there is some underlying system which generates judgments that are then obscured. To the extent that the effect of asking for judgments on decontextualized examples has been acknowledged as a problem, it has generally been considered as simply a confounding factor, obscuring the otherwise pristine data of grammaticality judgments. From a grammar as practice viewpoint, however, the fact that speakers need to be given or generate some context in order to make sense of examples submitted for judgment indicates that the activity of giving grammaticality judgments is derivative. The most natural use of our linguistic systems is in situated conversation.

From a grammar as practice viewpoint, one could say that speakers have beliefs like "we say it thus when we mean this," based on experience. These facts are naturalized for speakers in the sense that they come to believe that it is an essential property of speech (or, if they reflect on it, of their language), that things are said in a certain way. This process of naturalization gives the beliefs some inertia or persistence. However, enough experience (especially *thrown* interaction involving language) can provide new models. This is not to deny patterns in language, but the judgment that something never said is out or bad is a different matter. Grammaticality judgments reflect the linguistic system, but are not the essence of it.

6.7 Conclusion

In this chapter, I have considered the boundaries of linguistic competence in light of the results presented in Chapters 4 and 5. My goal in this chapter has not been to definitively establish where exactly the boundaries of competence lie. Rather, I wanted to raise some possibilities not usually considered in generative theory and therefore highlight the fact that the question remains unsettled. The mainstream position, that competence is merely a matter of grammaticality, has been accepted but not proven, and these issues merit further investigation. I hope that this dissertation has shown that sociolinguistic variation, the social value of variables and the non-categorical grammatical constraints that apply to them provide an interesting locus for the study of the boundaries of linguistic competence.