Chapter 3

Attention and Linguistics

**The job of the linguist**

I begin this chapter with some remarks made by Richard Hudson about the duty of a linguist with regard to the problem of meaning. He writes,

> Since I am a linguist, I shall not be able to penetrate [semantic] structure much further than the point at which the structure is supplied more or less directly by the bits of the cognitive network which are closely connected to the words themselves, so in this sense the semantic structure is relatively "linguistic,"compared with some other information which would have to be derived by more complicated routes.
>
> For example, take the sentence *Three couples came to dinner, but we only had enough meat for seven people*. The semantic structure for *three* and *couples* should show that three pairs of people came, and that *but* should show (I assume) that there is some kind of incompatibility between the two clauses; but we must leave to the other parts of the cognitive structure associated with this sentence to show just what the nature of this incompatibility is. . . I shall assume my job as a linguist is done once I have given satisfactory definitions for the meanings of the words themselves, and my responsibility does not extend to covering rules of multiplication and the etiquette of entertaining. (1984: 131-132)

Hudson makes a sound case for limiting the role of the linguist, but even he might agree that it is a limitation that has passed its usefulness, and that it is time to theorize linguistic meaning as an offshoot of cognitive semiotics. Since, strictly speaking, I am not a linguist, I will assume my job is done once I have given a satisfactory account of how linguistic structures interact with these other cognitive and social

processes, for the linguistic theory that shapes a grammar of attention proposes no such thing as *meanings of words themselves*. Words are simply formal recipes for building nonverbal readings of the world under a culture's guiding hand. We can appreciate the subtleties and complexities of meaning construction only by examining as thoroughly as possible the specific situations governing language production and interpretation (the aim of this and the next chapter).

I now set forth to describe what a linguistic theory looks like once we reckon it with a theory of attention. Doing so will require the following expository tasks: 1) providing a clear definition of language as a tool for thinking, acting, and interacting, 2) describing the peripheral status of attention in the generative linguistics tradition, while presenting an alternative tradition of linguistic theory, known collectively as cognitive-functional linguistics, where attentional phenomena play a defining role, 3) articulating in broad outline a theory of language based on the components of awareness and attention, and 4) applying of this theory to attested data using mental space networks as the preferred modeling device.

**The Semiotic and Interactive Functions**

Language has two characteristic functions: semiotic and interactive. Language functions as a semiotic system for initiating and maintaining symbolization of thought by means, sound, gesture, and inscriptions induced across situations. It takes conventional signs and uses them to perform representative and interpretive functions. Language constitutes our most powerful tool for generating public mental simulations removed from the here-and-now—simulations that help maintain our own sense of self as beings who harbor intentions and who live both in and through time.

Language functions as an invitation to interact with other minds (rhetorical) in that it permits the public expression of thought and promotes and social cohesion. Interaction is a fundamental component because speakers and writers, hearers and readers always enact a language within a physical space

accessible to others. Even if the physical space and the people and things filling it bears only a tangential relation to the message itself, their irascible presence influences the production and interpretation of signs.

The semiotic and interactive dimensions of language are only separable as a methodological convenience. As Sinha (1999) argues, language shares a dual grounding in embodiment and discourse. The "dual grounding hypothesis", as he calls it, stipulates that the cognitive dimensions of coherence (which issue from and return to the body) provide the mechanisms for the speaker to symbolize relevant features, aspects, and dimension of the world. These embodied mechanisms, however, tell only half the story, since that body is situated in a larger ecology. Fluid language requires discursive grounding insofar as the speaker is always aware of her own position vis-à-vis an addressee (even when talking to herself). The discursive ground of language includes the "primordial scene" of shared attention. Linguistic abilities like turn-taking, deixis of place, person, and time, perspective taking, cohesion and coherence demand an ecological stance (Sinha 1999: 247). Discursive grounding is always implicated in the embodied grounding, as was made apparent to me by a true story about my youngest son's own linguistic ingenuity.

The setting is a hot summer day in the middle of July. Simon comes into the kitchen and asks me for a popsicle (frozen juice on a stick). He then takes it outside to eat it as he plays in the sandbox. Twenty minutes later, he comes inside with the stick and sticky cherry-flavored syrup running between his fingers and down the palms of his hands. The next day, he comes into the kitchen and asks me for a *lick-it-quick*. Initially nonplused, I soon realized that he was pointing to the freezer and I determine that he is referring to the same things he had called *popsicle* the day before. Why this inventive naming?

Recall that surfaces and substances in our immediate environment produce sensations in the individual that may play a significant role in the kinds of mental simulations an individual is likely to construct about real and imagined occurrences. This instance illustrates the role these sensations play in shaping, at least momentarily, the structure and use of language. Analysis of my son's locution suggests that the affordance relation (see chapter 1) between hand, popsicle, and the weather now shapes

significantly his notion of eating a popsicle, so much so that the new name reflects that new relation, a verbal instruction for consuming it on a hot day. For analysis to be complete, however, one must also note its immediate social context. His new locution would not have been as successful without the pointing gesture, eye contact, and some knowledge of turn-taking procedures.

Affordance relations are not just ephemeral, *ad hoc* inventions by children, however. Sometimes they filter into the permanent lexicon of a language, particularly in English nominal compounds. Anyone experienced with woodworking will be familiar with the tool known as a *palm sander*. This tool is not tailor made for sanding only wood from palm trees, nor is it a tool for exfoliating dead skin from one's hand, rather, it is an electric sander whose handle fits within the palm of one hand (as opposed to a belt sander or orbital sander, each of which requires two hands to operate). Naming it thus appeared to result from the need to single this tool out (selective attention) based on the distinctive manner by which the woodworker interacts with it, rather than by a part-whole relation (as is the case with belt sanders). Carpenters seem to be attending to the manner of use when selecting this tool.

The semiotic and interactive functions of language rely on flexible repertoires of form-meaning pairs that structure information. The guiding principle of this study is that linguistic knowledge *emerges* from and subsequently *exploits* other cognitive functions such as sensation and perception, categorization, memory and attention. The emergence of linguistic constructions as conventional "recipes" for meaning reflects a cumulative cultural evolution of human cognition that can reveal, as Michael Tomasello reminds us, "the entire collective wisdom of the entire social group throughout its entire cultural history" (1999:7). I will concentrate primarily on the exploiting function of linguistic knowledge (i.e. performance), saving substantiation of the prior emergence hypothesis for a later project.

The assumption that knowledge of language is built on prior and more general cognitive processes marks a great divide among linguists (to be discussed later). This study assumes that language and attention are inextricably related and that the components of awareness and attention influence

language structure and use in the same way they influence perception and sensation. Language, like perception, is a way of organizing what someone wants herself or others to pay attention to. Linguistic constructions are not just empty syntactic vessels, but instructions for making something stand out as figure against a less differentiated ground.

Consider once again my excursion to Kelvin Smith Library, as I am momentarily distracted by the architectural features of an adjacent building. In a previous chapter, I recounted a fairly circuitous semiotic 'route' through which my attention shifts back on task as I perceived the topological fit of the concave and convex facades. But there was in fact a more direct semiotic presentation that could have reclaimed my attention had the words printed in bold black letters on a large aluminum box in front of the library not escaped notice. The sign on that box reads: ***LIBRARY BOOK RETURN***. The word *Library* would be a sufficient memory jogging cue to shift attention back to my initial goal, since buildings themselves promote the ongoing activities that occur within them. University campuses, like any other planned habitat, function as means of alerting, orienting, selecting, sustaining, and controlling thought, speech, and action. As an inhabitant, I have ready at my disposal many accessible, prefabricated mental spaces for representing multiple facets of the Case Western Reserve University campus.

## Attention and linguistic theory

Current thinking among linguists about the role attention plays in "linguistic competence" is anything but settled. In this section, I will present a selective review of linguistic theories and the role of attention.

### *Chomsky and universal grammar*

In an often-quoted passage from *Aspects of a Theory of Syntax,* Chomsky (1965) advances a definition of linguistic theory and method based on the assumption that knowledge of language is an autonomous

faculty encapsulated from all other mental functions. The role of attention plays no discernable role in language structure:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogenous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual situations. (1965:3)

Chomsky's point is that a speaker-listener's performance (a.k.a., E-language)—what she actually says and hears—is a mere manifestation of an *a priori* competence (a.k.a., I-language), a yet to be discovered "algorithm for determining the description of an arbitrary sentence given an arbitrary grammar" (Hockett 1968:41). As such, other cognitive functions like those listed above pertain only to a theory of performance, not a theory of competence (or Universal Grammar), and theories of performance provide little more than interesting observations, not explanations. Therefore, linguistic theory is concerned only with competence, which concerns itself only with unearthing the principles and parameters of UG while avoiding any concern for the bothersome and "untidy" performance data. This is the so-called "minimalist program".

The principal assumptions enabling Chomsky to simplify linguistic inquiry are 1) that knowledge of language is a well-defined and closed formal system (see Hockett 1968), and 2) that this well-defined system comes "preinstalled" in the newborn infants brain via genetic inheritance. The subject of linguistic theory is the initial state of the language faculty. Therefore, UG is the biologically endowed device that makes learning language possible. Since the features of any given language are so abstract and complex that it is *prima facie* impossible to learn it by simple induction (see Chomsky 1957, 1965, 1972, 1980, 1995), the child must already possess innate knowledge of language in order to constrain the "search space" of all possible grammars. Children come with adult-like abstract linguistic categories, such as

"noun" and "verb," syntactic pattern options, such as "head first" and "head last", which switch on or off during maturation. Learning a language depends on this *a priori* grammar which itself is not learned. Metaphysically, Chomsky's theory can be characterized as *abductive idealism*: human beings are born with a restricted class of admissible hypotheses about grammatical form (idealism) and a guessing instinct (abduction) for making the correct match.

Abductive idealist notions of linguistic competence regard differences between languages (geographical and social dispersion) and changes within a language's structure are byproducts of changes in parameter settings during language acquisition. According to Lightfoot, languages change because the child "selects relevant stimuli according to criteria that are already present internally" (1991: 2); that is to say, the data that children can attend to are already predetermined. In this respect, the selective attention is understood as a data extraction mechanism at the presentative condition of the sign only (i.e., the formal distribution of signifiers in syntactic space). It seems that shared attention, sustained attention, and shifting attention play no role in determining the nature of linguistic structure. Abductive idealism also makes it clear that language change only occurs in children, resulting from ever so slight variations among the initial parameter settings: linguistic knowledge is a closed circuit of forms; hence language change comes about only through inconspicuous changes in the parameters of those forms from one generation to the next.

I rehearse the above theory only to point out that cognitive processes like attention come into play only tangentially. The child must attend to the external linguistic triggers of adult performance in order to link UG to English, Japanese, Farsi, or Danish, for instance. Under the program initiated by Chomsky, linguistic research proceeds by studying I-language in isolation. Since I-language is cognitively encapsulated, other cognitive processes, like attention, merely help to implement I-language, which, one presumes, must be understood in relation to these other mental processes. Only after linguists have fully explained I-language can they then proceed to discuss with any cogency E-language. I am not particularly

sympathetic to the direction of explanation Chomsky has chosen. Instead, I favor the approach taken by Hopper & Traugott that language change issues from pragmatics (not syntax), specifically from the "tendency of speakers and hearers to organize information in accessible ways" (1993: 208). Disagreement with the Chomskyan programme does not mean I fail to appreciate that his two big ideas--that all languages are essentially the same and that these principles are innate--have altered the scope and focus of linguistics away from relating sound and meaning to uncovering universal properties. His influence being so broad and consequential, I would not be surprised that many linguists would consider what I am doing well beyond the purview of their discipline.

Be that as it may, I contend that whatever we might gain from Chomsky and his followers (and I concede that they may be significant), we have also paid a great price. Generative linguistics has abdicated any responsibility for developing a general theory of semiotics (indeed that is a result of the central, nativist assumption), and hence, has avoided developing satisfactory accounts of meaningful behavior.[1]

At the same time, many of Chomsky's theoretical abstractions, such as *trace theory*, inadvertently suggest that attentional mechanisms such as selection, sustain, and control provide a significant property of UG. For instance, Chomsky writes that embedded clauses in a sentence like (1),

(1) John, a New York police officer, is in love with Mary and has resigned from the force and moved to Anchorage to be near her,

are surface manifestations of four independent activations of the subject *John*:

(1a) John is in love with Mary,

[1] In one of his most recent lectures, Chomsky takes a squinty-eyed view of semiotics. He equates it (by way of reference to a comment made by Dan Sperber) to the study of traffic lights, and thus opining, "A lot is known about . . . language; there's not much to say about traffic lights. There are other big topics like cinema, art, human relations, and those are terribly important topics but I don't think you learn much about them from semiotics" (2000: 48). Chomsky's comment reveals the precise nature of our disagreement: he thinks knowledge of language has nothing to do with human relations (hence nothing to do with general semiotics), I think the more we learn about human relations the more we will learn about knowledge of language. We approach the problem from opposite directions.

(1b) John is [t] a New York Police officer,

(1c) John [t] has resigned from the force,

(1d) John [t] moved to Anchorage to be near her.

The surface manifestation hides the fact that John is the head of four different clauses, and that UG as manifest in English has a built-in mechanism for tracking the referent. It would seem, then, that trace theory implicitly relies on a covert form of sustained attention, for in order to bring these deep structures to the surface requires holding the referent in working memory long enough for these argument structures to come online. Of this example, Chomsky would ask the obvious question, "why do we hear the name *John* only once?" His answer is that we hear the name *John* only once but that our mind 'hears' it four times. Chomsky then would elaborate his answer by appealing to the "displacement property" as a principle explaining why we physically hear it only once, claiming that "the phonetic output interface erases everything but one and it does so in a very general fashion" (2000:25). In other words, knowledge of language requires sustaining attention on the trace element during the syntactic processing, and Chomsky's view of this sustained attention is the covert process by which the mind generates clauses (1a-d) but that the articulatory-auditory circuit compresses it to one. This operation is a simple "copy operation", where sustained attention operates only on the surface.

Let us assume for a second that trace theory is correct and works exactly as Chomsky describes it. The next logical question then would be, "Why should language have a displacement property?" Aside from offering the argument that it satisfies an "externally-imposed legibility condition," Chomsky simply asserts that there is no good answer to that question.

I propose that one of the reasons why there is no good answer is that the program forbids looking beyond its own formal principles and parameters for an answer. What if we violate this edict and look beyond the formal systems for an answer. What if this "externally-imposed legibility condition" is not so much an abstract rule but a general semiotic notion that the topic of a compound-complex sentence

appears once and at the beginning to select attention in the representative condition of the sign, thereby activating working memory? Once selection occurs, the attentional budget is spent on entities associated with the selected entity corresponding at that moment when presentation slips to representation, which, automatically necessitates continuous interpretation if a stable mental scene is to develop. Thus, a discourse participant selects an entity, a particular "male human being" (a designation) and builds a series of conceptualizations (arguments) around him. The selection process invites attentional budgeting because symbols for human beings bring to the fore additional representations of intentions, actions, and reactions, so important to human existence. The speaker builds a network of representations around a single, selected focal entity, *John*: his mental state and the object of his desire (Mary), his occupation, two actions (quitting the force and moving to Alaska), and his motivations for acting. Under this view, Chomsky's notion of trace only applies to the presentative condition, but that representative and interpretive conditions also need to be folded into the knowledge of language. The trace element, then, is not merely a phonetic marker in the mind and mouth. In fact, the "erased" names from each clause (if we want to characterize them thus) may bind with the same referent but surely referent is not the exact same entity in subsequent clauses. At first mention he is a New York police officer; at second mention he is a lover of Mary; at third mention he is someone who has made a choice; and at forth mention, this now ex-police officer and desiring lover of Mary has left New York for Alaska. In essence, part of the answer to why speakers of English can "gap" these references to *John* has as much to do with the distribution of attention across a network of mental spaces as it does with the any "phonetic-output interface". Or, to put it another way, the formal linguistic constraint Chomsky cites emerges from nonlinguistic semiotic systems, the "other" of language. It seems that any appeal to externally-imposed conditions shifts the ground of explanation to more generalized cognitive processes.

Ray Jackendoff can be regarded as the generative linguist most interested in understanding how knowledge of language relates to other cognitive processes. It is to his most recent theoretical statements

on attention, consciousness, and language that I now turn.

*Jackendoff on attention and the language faculty*

In his recent book, *The Architecture of the Language Faculty* (1997), Jackendoff makes several statements in his chapter titled "How Language Helps Us Think" that accord in principle with a grammar of attention. First, that language enhances significantly the range of thought and the means of sharing those thoughts. (No linguist would quarrel with this assertion.) Second, that language affords human beings the ability to "experience" their own thought in a way heretofore impossible for nonhuman primates, even great apes. We are conscious of our own reasoning processes. We are not only aware of them, we are attentive to them. Third, language affords abstractions. With language, argues Jackendoff, we can make present in consciousness abstract thoughts like predicate and argument, kinship relations, counterfactual and hypothetical situations, to name a few (1997:200-201). The presence of language in consciousness "allows us to pay attention to them". And paying attention to them means applying "extra processing power" to the things they represent. "Attention to some details," he writes, "may lead to elaboration of further details"—selective attention slips to sustained attention. With respect to language, attention to phonetic form (i.e., the presentative condition of the sign) "anchors attention" onto a pattern allowing long term memory to enact it as a repeatable symbolic unit.

As isolated phonetic forms, language focuses attention for continued mental processing, often of retrievable memories not present in the immediate environment. Attention to items further permits valuation of those items. Language "holds thought still" so that we can do new things with it. Since these signs are public, attention to combinations of phonetic forms and their representations permits collective knowledge. Still, Jackendoff endorses (at the same time he concedes as unproven) that core linguistic knowledge is syntax and the study of its properties should be isolated from these other concerns. I naturally disagree with that assumption and approach.

*Origins: a"mute verbal modeling device"?*

Chomsky and Jackendoff both assume that language evolved as a means of representing thought first and as a means of communication second. They echo Sebeok's (1998:125) argument that language evolved as a "mute verbal modeling device" that only later gets pressed into use ("pre-adaptionism" in the evolutionary literature) for "message swapping". The direction of language and speech evolution goes from the internal world to the external world to the social world. In this respect, language existed as a formal modeling device with innately specified properties. If linguistic competence has anything to do with attention, it is with attention as a purely private, subjective phenomenon.

Sebeok never explains exactly what a 'mute verbal modeling device' is, other than that our early hominid ancestors (*Homo erectus*) make internally conventional signs perform representational functions they could not produce publicly. It is certainly true that comprehension precedes production ontogenetically (see Deacon 1997, Lamb 1998), and it stands to reason that some form of internal comprehension must have proceeded smooth, articulate production phylogenetically. However, I question whether the meaning/communication distinction is as discrete as Sebeok would appear to have it. He assumes that modeling behavior is entirely internal, or that a fairly elaborate internal production line for bundling symbols with meanings exists. I would argue that modeling devices of all types—especially those that become "quintessential taxonomic markers" of a species—depend on production results outlasting working or short term memory limits of the organism, as consistent with the materiality principle outlined in the previous chapter. The evolutionary trajectory does not move from inside to outside, but rather moves dialectically from outside to inside to outside again; this developmental trajectory resembles liberally Vygotsky's (1962) psychological views and, more broadly, Clark's contention that most cognitive strategies "involve the use of some type of external structure or "scaffolding" to mold and orchestrate behavior" (1997: 23).

Deacon (1997) presents a similar evolutionary picture of language. Rather than the brain changing to fit language, language changes to fit the brain. Morgan (1995) reminds us that human neonates are unlike neonates of any other species in their total, abject helplessness. Chimpanzee newborns can walk postpartum and begin to fend for themselves; human newborns cannot do anything for themselves, except be socially precocious. In order to compensate for this helplessness, Morgan argues, neonates are born in to a richly intersubjective world. By all accounts, this precarious position of the human neonate predates language by a few hundred thousand years. By this logic, language fits a brain already primed for shared attention, imitative learning, and modeling devices manifest externally before they manifest internally.

Human beings inhabit a world of many potential signs. It would at this point be important to emphasize, as Peirce did (1:538) that "Every thought is a sign," and that language exemplifies human modeling behavior as it manifests our robust ability to exploit signs bearing only a conventional relationship to what it represents. This internal ability is semiotic and not strictly linguistic. The initial modeling device is internal to the extent that early ancestors developed this capacity. However, it is also apparent that nonhuman primates—especially those coming in constant contact with human beings—display this same kind of behavior using rudimentary forms of sign language, which would suggest that the human social milieu has a critical role to play in the emergence of this capacity. [2] So, why is it that

[2] I am not arguing that the rudimentary forms of sign language exhibited by the bonobo, Kanzi (and, in even more attenuated form, by his relatives, and other species of great ape) is some "evolutionary intermediate form of language". There is no such thing. Whatever linguistic capacities Kanzi has are due to the fact that he has spent nearly all of his life in a human culture. The only evidence counting as a truly evolutionary precursor to language would have to come from observation of feral great apes. There is no such evidence.
The fact that cultured bonobos and other great apes manipulate conventional signs is not disputed; however, the degree to which they do, and whether or not one can say that they 'have language' is hotly contested. I take it that the fact that great apes, when in captivity, press conventional signs into use for making requests is one (but not the only) kind of rudimentary modeling device behind the evolution of language, reflecting the expanding presentative condition of the sign necessary for the ever expanding representative and interpretive functions characterizing human thought. Even if Kanzi's behavior is pure mimicry (which I do not believe), it manifests not only this basic semiotic basis for the development of higher-order thought but also betrays a nascent understanding of the other. In my opinion, it would seem that many nonhuman primates also possess a modeling device that, as near as I can tell, resembles Sebeok's "mute verbal modeling device". Unlike human beings, however, Kanzi does not come close to the power a human three-year-old has to perform conceptual integration across diffuse domains. Without that

our semiotic behavior far outstrips even our closely living relatives?

To answer this question, we need first enumerate what the best evidence suggests about what all mammals, nonhuman primates, and humans have in common. For this I cite the work of Tomasello & Call (1997) and Tomasello (1999).

First, all mammals live in a physical world of permanent objects (we detect both form and motion, for instance). Consequently, all primates can represent categorical and quantitative relationships among objects; they can remember where objects are located in their habitats (e.g., "what" bugs live "where," "what" fruit hangs on "which" trees, and so forth); they can deal with contingencies by making novel paths and detours through that space; they pass object permanence tasks (they can visually track both visible and occluded objects); they create categories based on perceptual similarity; they can match objects and small numbers of objects; they display problem solving capacities. Second, all primates inhabit a similar social world in which they recognize other members of the same species (i.e., conspecifics) and they can recognize and affirm through elaborate behavior relations of dominance, submission, and affiliation. Consequently, they can recognize individuals within social groups; they can form direct relationships based on rank, friendship, and kinship; they can predict the behavior of another by the way others move; they exploit social and communicative strategies in order to obtain food and other valuable resources; conversely, they cooperate with others in order to obtain valuable, but limited resources; and they learn from each other (Tomasello 1999: 16).

What primates in their natural habitats do not do is quite telling: they do not point or gesture toward objects for others; they do not display objects to others; they do not call attention to objects or events in locations for others to witness; they do not offer objects to one another; they do not seem to teach other individuals new behaviors. In short, they lack shared attentional strategies. Although they show tremendous individual creativity and inventiveness, feral apes do not try to pass on or share their

---

capacity to efficiently integrate disperse categories into coherent and sometimes novel scenes, nonhuman primates cannot move beyond symbolizing over a very narrow range of intersubjective situations (all of which seem to refer only to the immediate present).

inventions. They cannot stabilize inventive practices to "ratchet up culture". From this, Tomasello concludes that

> . . .for many animal species it is not the creative component, but rather the stabilizing ratchet component, that is the difficult feat. Thus many nonhuman primate individuals regularly produce intelligent behavioral innovations and novelties, but then their groupmates do not engage in the kinds of social learning that would enable, over time, the cultural ratchet to do its work. (1999: 5)

Chomsky, Jackendoff, and Sebeok (each for different reasons, I suspect) base their theories on individual innovation rather than collective replication and learning. A theory of the sign is incomplete without taking into account this ratchet component, or what Tomasello calls "joint attentional scenes" (1999:96). The problem with the mute verbal modeling device hypothesis as outlined by Sebeok is its inability to provide a persuasive account of how such a device would ever develop the necessary ratchet effect without shared attention as a prior condition and as an intimate partner in the process (not just a "trigger").

Shared attentional scenes are the middle ground between perception and representation generally and perceptual events and linguistic symbols specifically. A generative approach (at least the approaches with which I am most familiar) propose a link between perceptual events and linguistic symbols to be yet another layer of symbols comprising Universal Grammar. But this approach seems to ignore a fundamental principle that linguistic reference is a social invention, whereby a producer of signs attempts to focus the attention of an interpreter of signs on something else (and vice versa). This capacity, I argue, depends on the principles outlined in the previous chapter—principles of attention which themselves precede linguistic symbolization but which are greatly enhanced by linguistic symbols.

For linguistic symbolization to take root, a human being must possess a nascent sense of self to an extent that she can conceptualize herself as a participant in an interactive scene; she must conceptualize

the other person as an intentional, volitional subject within the same spatial temporal environment who cares about or can be made to care about similar happenings therein. The ability to do this, which I believe separates human primates from nonhuman primates, is not a linguistic capacity, but a capacity of mental simulation implicating other minds. Language sits on top and refines these capacities, but language cannot exist without them.

Social semiotics depends on internal selection, sustain, and control of attention at the same time that it depends on the ratchet effect of shared attention. I would argue that the reason I have a lexicon for architecture and prepared and novel ways of relating representations of them is not because I have innately specified categories of noun and verb, subject and object, but because I have a brain that draws on a much larger palette of signs (the presentative condition of the sign far outstrips other species); and the reason I have a brain that does so is that I was born into a culture that selects for social precocity and makes use of this wide palette of signs. It is the social precocity that precedes linguistic symbolism. Thus, I can spy something, focus my attention in it long enough (over 200 milliseconds) to perceptually categorize it as a *window* (which is not the same as translating my visual signs into phonemic form), use that token of the category as a reference point for perceiving and conceptualizing a facade, which, in turn, functions as a conceptual reference points for *architectural styles*, and so on. At that moment, I am selecting, sustaining, and controlling attention in such a way that it brings the perceptual environment in line with long term memory, categorization, and valuations that are not perceptually present, but represented and interpreted. This use of attention seems to be in agreement with Jackendoff's approach.

Suppose, however, that I am walking along the pedestrian avenue with a colleague. I then use the linguistic symbolization (along with gestures) to call his attention to these features. Now both of us have focused our cognitive resources on a narrower scope within the perceptual world. Many more objects, events, and actions occur around us, and we are aware of some of them, but these things are not a part of the same scene. I may be aware of the trees that flank the walkway, but I do not think of them as bearing

specific relevance to my mental simulation, unless a colleague points them out to me. Linguistic symbolization is perhaps our most flexible means of engrossing other minds in the same scenes for specific purposes, and perhaps the most powerful and flexible means of revising or modifying such scenes, however trivial or significant.

I am sure that Jackendoff, Sebeok, and Chomsky would not quarrel much with the above account. However, I would argue that these shared attentional scenes, which adults engage in so expertly and fluidly, serve as the very basis for acquiring linguistic symbolization in the first place. I am sure Chomsky and Jackendoff would disagree with me on this point.

One can acquire and use linguistic symbols without an intervening layer of UG (at least as specified by generative linguistic theory) but not without the "memory limitations, distractions, shifts of attention and interest" that Chomsky deems irrelevant. I think it desirable in proceeding that we invoke a very different kind of ideal speaker-listener than the one Chomsky invokes at the beginning of *Aspects of a Theory of Syntax*. Let us invoke an idealized speaker-listener who has acquired language in the manner discussed in the next section.

*Legacy: How is language acquired?*

At the heart of generative linguistics is the hypothesis that children possess adult-like grammars from the very beginning.[3] More specifically, children come to the language learning environment with the categories noun and verb already activated (as a kind of empty set in the initial state). Supporting arguments for this "continuity hypothesis," as Tomasello (in press) calls it, has been largely negative; that is, "knowledge of language is so abstract and formal and the possible variations so indefinite that children could not possibly learn it without the aid of some internal acquisition device for constraining all the logically possible interpretations of the triggering symbol". In short, evolution has "bundled" knowledge

[3] Many linguist, including Chomsky himself, claim that UG is expressed directly in the genome, while others, like Jackendoff, hint that UG is expressed indirectly through epigenesis.

of language with the latest version of *Homo sapiens sapiens*.

Recent research in child language in the form of systematic analysis of children's spontaneous speech as well as from controlled experiments directly contradicts the continuity hypothesis. Of special note is Tomasello's body of research on language production and controlled experiments with / using children aged two to three-and-a-half. If children did indeed possess adult-like syntactic and semantic categories from the start, the evidence should show that they can produce speech in which both nouns and verbs function as more-or-less open slots. They should, for instance, use certain verbs in a wide array of constructions, both transitive and intransitive. While two-year-olds do seem to have a rich open nominal system, the same does not hold for verbs. According to Tomasello (1992), acquisition of language in general and verbs in particular suggest a more item-based and uneven account of language acquisition that is a genuine alternative to the continuity hypothesis that has positive empirical support. Tomasello (in press) summarizes this alternative called the "verb island hypothesis". The hypothesis states that

> children's early language is organized and structured totally around individual verbs and other predicative terms; that is, the 2-year-old child's syntactic competence is comprised totally of verb-specific constructions with open nominal slots. Other than the categorization of nominals, nascent language learners possess no other linguistic abstractions of syntactic organization. This means that the syntactic categories with which children are working are not such verb-general things as 'subject'or 'object'or even 'agent' and 'patient',but rather such verb-specific things as 'hitter' and 'hittee', 'sitter' and 'thing sat upon'.

If Tomasello is right, then the child brings to language learning an ability to recognize objects, a precocious desire for human interaction, a nascent ability to conceptualize interactive scenes with both objects and people, and an ability to construct an interpretation of these scenes, and a desire to repeat them and represent them. The child learns individual items or small, pre-compiled symbolic assemblies

that are patterns discerned from adult usage, she then makes abstractions that further allow her to make her own structured inventory of constructions that are at least recognizable to the adult as expressive, informative, and functional symbols. In short, the child's burgeoning linguistic competence, and the abstractions that characterize it are byproducts of a social-cognitive learning interaction with adults and other, not causes of this. In this respect, Tomasello's account sees grammatical abstractions as emergent not immanent.[4]

Shared attention between subject and other form the foundations of learning—the child has a stable sense of self that she understands to be different from but similar to others, and as such she is keenly interested in the intentions, actions, and thoughts of others (especially her caretakers), and that she understands that both she and these others harbor intentions toward and about objects, events, and actions in the immediate environment. When the child is in the kitchen with her caretaker, for instance, it is typical that she will attend to food and other kitchen related objects. It is through the shared attentional to "the kitchen" that the range of possible topics and foci that are to be simulated and eventually symbolized are narrowed. Such activities depend on three processes: *imitative learning*, *analogy* and *structure mapping*, and *structure combining* (in press).

Imitation is not mimicry. The child is not simply repeating verbatim what the adult has just said. Imitation means that the child reproduces the form precisely because she has matched the form with the meaning (i.e., purpose and function) of the pragmatic situation. For example, a child might hear her mother say *See that playground*. In order for that form to become an item in the child's inventory of symbolic forms, the child must understand that her mother intends that she share attention with her about a particular object or set of objects. The child need not have an abstract notion of verb or noun. What she

[4] On this point, Tomasello's approach to linguistic symbolization has much in common with Hopper's (1987, 1998) theory of emergent grammar. Both Tomasello and Hopper see grammatical form as a posteriori and depended on prior social and cognitive forces. It would probably be fair to say that these approaches are complimentary insofar as Tomasello focuses on individual cognitive development within social settings, whereas Hopper focuses on historically derived grammaticalizated structures and the social/political forces that favor one form or set of forms over another.

needs to have is a nascent understanding of subjectivity and intentionality as it relates to a developing sense of communicative intentions (publicity), to wit she needs to understand that her mother intends to manipulate her intentions. This intentional reading gives way to imitation once the child begins to understand that she too can manipulate her mother's attention using the same symbols.

Structure mapping begins to settle in once the child begins to develop a structured inventory of (usually quite small) symbolic assemblies that "attach" or "index" specific kinds of attentional scenes. She begins to find similarities and differences in these symbolic assembles as her comparative knowledge of interactive scenes grows. Since human beings are able to discern similarities based not only on perceptual and functional features, but similarities among events and relations abstracted across the particular objects involved, then analogy becomes a viable mode of inductive learning. That is, the child may come to understand and use verbs, such as *stir, mix, beat, roll, pat*, etc. when in the kitchen, she may begin to see a verb-general pattern emerge among these forms, namely that they all involve manipulation of hands or tools and/or some other (usually pliable) object. Thus these verbs begin to cluster around a kind of material situation that has many common perceptual, functional, relational, and actant features. The child who understands and can reproduce all or part of *see that playground* gradually develops a more productive but more schematic form of this epistemic construction as it applies to other sense perception verbs. What started out as a *see that* ___ verb island construction, now becomes a *hear that* ___, *touch that* ___, *smell that*___, *taste that* ___constructional schema that can broaden out to a number of situations calling for the manipulation of another's attention.

Once the child begins to induce similarities among the representational resources of her language, she begins to combine these assemblies (components) with other assemblies to form composite assemblies of various shapes, sizes, and abstractness. Data from Tomasello (1992) reveals that by 19-22 months T (the female subject of his study) was producing three word utterances that the data history suggest strongly to be composites of two simpler assemblies. She would say *See Daddy's car*, a derivative

of the previous verb island construction *See*___ and the possessive construction *Daddy's* ___. This more complex "see possessive construction" blends representations for focusing attention and conceiving ownership.

The three components of language learning—imitation, mapping, combination—relate well to the components of attention. Imitation suggests that the child can orient herself and focus attention on another's actions within a given physical and social setting. Mapping or generalizing form means that the child can sustain attention on something long enough to establish relational inferences from long term memory, a process entailing the narrowing of incoming perceptual stimuli so that she can work with that which already has spent the majority of her attentional budget. Combination suggests sustain and control of attention, she sustains attention when working with the same material and relating it with other material means creating new, often novel inferences; she controls attention insofar as she switches from noticing something to getting someone else to notice the same thing and attributing an abstraction to that thing that captures another's attention—all with just three little words.

Additional cross-linguistic research in first language acquisition epitomized by Sinha and Jensen De López (2000) calls for additional emphasis on sociocultural context, particularly with respect to the embodiment hypothesis attributed to Lakoff and Johnson (1980, 1999). Sinha and Jensen De López argue that the strong version of the embodiment hypothesis (i.e., that bodily experience structures most if not all psychological and interpersonal domains through metaphoric projection) may not be completely true. Their study of English-acquiring, Danish-acquiring, and Zapotec-acquiring children as they acquire and use locatives, and as they test them using language comprehension and action-imitation tasks, suggests an equally influential role played by sociocultural context in cognitive development. At issue is the structural fact that the Zapotec language exhibits no morphological distinction between the nominal whose English equivalent is *stomach* and locatives meaning *in* or *under*; whereas the English and Danish languages distinguish *in* from *under*, and acquisition patterns and experiments suggest a definite "in-bias"

among English-acquiring and Danish-acquiring children as a "good" example of "containment" and with a bias for regarding "under" as a special case of containment (one that constrains an object's or entity's movement, but without completely encapsulating it). In contrast to their English and Danish counterparts, an "in-bias" when using locatives is in evidence for Zapotec-acquiring children.

As the history of the Zapotec language attests, the role of the human body is a salient source for linguistic concepts, as is evidenced by the fact that body-part terms acquire locative functions. But this is an historical effect and not a directly cognitive effect, for this study suggests that bodily experience is not the driving force for linguistic constructions of space and for the acquisition of spatial terms. What is the driving force?

Sinha and Jensen De López argue that we have to look at the sociocultural context and artifactual composition of cultural settings for an answer. Unlike Danish-acquiring and English-acquiring children, who, for the most part, are born into a world of richly diverse sets of artifacts, each of which perform highly specific functions,Zapotec-acquiring children grow up in material cultures with few artifacts, and, therefore, make use of them in more flexible ways. One salient artifact of containment in Zapotec cultures of Southern Mexico is the basket. The child enters a world in which baskets are used as often to *cover something up* (e.g., tortillas, storage, catching chickens) as they are used to *place objects in*. The inverted orientation of the basket is a defining part of their material culture. In Zapotec culture, containment vis-à-vis baskets counts equally in its inverted orientation (*under*) as it does in its canonical orientation (*in*). The same is not true for English or Danish speaking cultures.

Sinha and Jensen De López offer the tentative explanation that Zapotec-acquiring children did not evince the same "in-bias" in their responses as English-acquiring and Danish-acquiring children, because baskets (the artifact used in all the experiments) are not used in the same canonical way. From this we have evidence that :

a *nonlinguistic* sociocultural difference regarding canonical artifact use, *embodied* in the

> material cultures and *exemplified* in nonlinguistic cultural practices, gives rise to. . . significantly different conceptualizations of "containment" in the different cultures. (author?2000: 20)

The dual grounding of language in embodied cognition and in mediated social practices seems at least plausible to assume. What this means for a semiotic theory of attention is that the immediate perceptual environment, studded with artifacts and observable practices that are repeated and later imitated and, inevitably, modified by its neophytes, plays a role in selecting, sustaining, and controlling attention in such a way as to bias conceptualizations in one direction or another.

*Psycholinguistic experimentation: attention detection and conceptualization*

To what extent is an attentional approach to linguistic knowledge more psychologically plausible—and hence a better account and possibly a real explanation—than mainstream generative theory accounts? On the face of it, the attentional approach produces a more satisfying descriptive account because meaningfulness is always in the forefront of explanation; additionally, there is no need to posit innate, preinstalled meaning structures. As recounted in the previous section, an alternative hypothesis about language acquisition is, in fact, well supported observationally and experimentally.

A further line of research exists based on psycholinguistic experiments conducted by Tomlin (1995, 1997) suggesting other cognitive processes, especially attention, bear significantly on the relationship between form and meaning. Tomlin, therefore, proposes that we replace the classical notion of clause-level theme or topic with the notion of "focal attention," a cognitive notion well-grounded in psychology literature and amenable to experimentation. The goal is to construct a theory of discourse, syntax, and grammar that links production directly with cognition.

Tomlin proposes the following hypothesis for the function of the subject in English:

> At the moment of utterance formation, the speaker assigns the referent in the current

representation which is currently attentionally detected as the syntactic subject of the utterance. (1995:167)

He argues, I think rightly, that when speakers report on observed visual events "some sort of representation of activity witnessed exists prior to attempts to describe the witnessed activity" (168). The selection and sustain of attention may be fundamental in the transition from non-linguistic representations to linguistic representations. The basic unit is an event, composed of a set of parameters standing in some relation to one another in a field.

To test this hypothesis, Tomlin developed the following experiment. Two kinds of visual experience were viewed by 12 native speakers of English. (Tomlin also conducted these experiments cross-linguistically, using Polish, Russian, and Bulgarian speakers among others.) The first experiment is with dynamic events, in which multiple concrete objects interact for a brief time. For instance, Tomlin has his subjects look at a screen saver program of a iterated scene in which two fish, one light one dark, approach each other until, in an instant, one of the fish swallows the other and continues on swimming. Tomlin asks, "how is that brief scene represented conceptually, and on what sort of conceptual representation does the language-production system operate?"

To answer this question Tomlin devised an experiment where he asked subjects to view the fish-swallowing event and verbally report what they had seen. Attention is cued by a flashing + sign or a flashing arrow in the place where one of the fish (predator or prey) will appear 150 millisecond before the presentation of the swallowing action—too brief an interval for attention to "wander". At 500 milliseconds into the trial a mask is presented on the screen that covers the cue, after which subjects produce a report. If the cue was on the prey 75 milliseconds before event onset, the prediction was that speakers would produce passive constructions (e.g., *The yellow fish was eaten by the blue fish*); if the cue was on the predator 75 milliseconds before event onset, the prediction was that speakers would produce active constructions (e.g., *The blue fish swallowed the yellow fish*). Ten of the twelve subjects performed

as predicted. Overall, argues Tomlin, "on a speaker-by-speaker basis, this experiment successfully predicts subject assignment and concomitant voice structures on an utterance-by-utterance basis"(1995: 178). Two subjects did not perform as predicted, producing active constructions exclusively.[5]

Let us take this empirical support and apply it to a reading of this fictional example that we can imagine being uttered by a detective to reporters standing just outside a crime scene:

(2) All we know at this time is that the victim was murdered with a scalpel.

Why did the detective use a passive construction at this moment?

When a given location becomes a murder scene, it does so by virtue of the presence of a dead human body. It is typically the most salient feature of the immediate environment. Remember that we orient visual attention toward the novel. Since murder scenes are not habitually so (descriptions of places within inner cities in which much killing occurs are called "war zones" not "crime scenes" or "murder scenes"), the presence of a dead body is a novel stimulus. While an experienced detective will likely look at several features of a scene before carefully examining the victim's body, it is undeniable that the body gets the lion share of attention. In this situation, the patient role and its condition are salient in attention, so it turns out to be grammatically salient as well. Furthermore, the detective(s) has no other information to go on than a few signs, wounds that index a particular type of weapon (or the weapon itself).

Once placed in the foreground, new information can be symbolized. Unlike many instance of agentless passives, this expression carries a salient implicature, through abduction, regarding the class of persons who might have committed the murder, perhaps a surgeon or medical practitioner. The passive construction is employed because it gives first mention to the most salient focal participant in this kind of situation and secondary prominence to those features bearing direct relevance to the resultant state of the focal participant. This expression focuses reporters' attention on the same thing that the detective is

[5] There are a number of possible explanations for their anomalous performance. They may not have followed the instructions to keep their eye on the character; subjects might have avoided the passive outright; there might be an age effect since both subjects were under twenty; they may have paid attention to the fish that opened its mouth at the moment of eating and this may have competed with the patient for their attention

focused on for purposes of representing the investigators' current knowledge of the situation.. The nonverbal signs of the crime scene itself focus the detective's attention on specific features within it; his verbal reporting in effect guide's a third party's attention (presumably with no immediate access to the scene) on the same features.

*Cognitive linguistics and attention*

Tomlin's and Tomasello's respective research programs exemplify a different research paradigm, known as Cognitive Linguistics, that began to emerge in the middle eighties with the work of Fauconnier (1985), Lakoff (1997), Langacker (1987, 1990, 1991), and Talmy (2000), among others. Although different in many particulars, cognitive linguistic approaches, it can be surmised, rally around the common theme that knowledge of language depends on other cognitive processes, attention included. Shifts of attention, memory limitations, categorization are not irrelevant performance limitations, because the ideal speaker-listener is first and foremost a being who remembers, who categorizes, and who attends in all complex ways described in the previous chapters. Linguistic symbolization is not a matter of linking consciousness with universal grammar; rather, linguistic structure itself emerges from other cognitive capacities dependent on the components of attention, individual and shared.

In this view, linguistic competence is not a core set of abstract, *a priori* rules that, with maturation, eventually link up with normal but peripheral processes of learning and memory. Rather, mastery of linguistic symbols means linking them up with constructional schemas (i.e., doing, having, existing, happening, transferring, experiencing, and moving) which themselves emerge from perceptual and interpersonal experience. Cognitive linguists think it an unwarranted assumption that conscious cognitive processes can be characterized only as verbal just because we report hearing an inner voice when we speak. Instead, the mental processes are primarily the construction of mental images, as argued in the previous chapters. Language processing is a matter of generating verbal recipes for "cooking up"

mental representations and internalized models of reality. Only after this kind of linking takes place does symbolization take on special characteristics specific only to linguistic communication. What kinds of processes figure in the linguistic symbolization?

Human beings possess innate capacities for certain kinds of experiences that form the initial conditions of meaning construction. In this respect, linguistic symbolization satisfies a semiotic imperative.

Normal human endowment ensures that we experience a species typical range of visual, auditory, tactile, olfactory, and gustatory sensations; it ensures that we possess common notions of our body moving in space and of the feel of extended space; it ensures that we can project instrumental functions onto artifacts and objects; it ensures that we experience a range of shared emotional states; it ensures that we can uses sensory stimuli as a scaffold for comparing two or more experiences; it ensures that we can abstract or schematize various situations as the same, or similar, or dissimilar from one another for purposes of categorization and interpretation; it ensures that we establish relationships by grouping sets of entities and thereby manipulating them as single units or unpacking the group by singling a subset out for special attention; it ensures that we can engage in mental scanning, either sequentially or in summary fashion; and it ensures that we can represent perspective taking of ourselves and others. All of these mental capacities depend upon the capacity to alert, orient, focus, sustain, and control attention.

Symbolization occurs as an interactive imperative, for there is pressure to manipulate signals to effect a change in mental states of others. Linguistic symbols do this nicely because they can, better than other symbol systems, represent "a distillation of shared human experience" (Langacker 1999:1) The human endowment that confers the ability to abstract form-meaning pairs (sometimes called in linguistics literature *constructions* or *constructional schemas*) permits a common currency of symbolic assemblies. A form can have a schematized meaning that becomes useful for a wide variety of semiotic purposes. Under this view, the lexicon and grammar of a language, far from distinct in kind, reflect degrees of

conventionality and schematization, with lexical items tending to retain their specific points of fine detail and with grammatical items tending to lose their specific detail in favor of schematic features portable across contexts. For example, what is the status of a lexical verb *return*? Does it stand alone as an item *sui generis* or does it stand as part of a larger constructions, such as [[NP][*return*] [[NP][PP]]] (e.g., *She returns the book to the library*)? Langacker's (1998:124) and my own answers would be that it is both. With respect to the larger construction, it organizes a more general grammatical pattern and grounds that pattern in the semantic and pragmatic exigencies of the discourse. With respect to the lexicon, it specifies a kind of action worthy of repeated representation. On this view, the same linguistic element functions in multiple and intersecting networks. Both the lexicon and the grammar are different means of orienting, focusing, sustaining, and controlling attention (a point I shall develop later in the chapter).

Taken together, the lexicon and grammar constitute a representational system of four domains: *content*, *epistemic*, *speech act*, *metalinguistic* (Sweetser & Dancygier 1999). What makes language such a powerful semiotic tool is it can represent the "contents" of our thoughts; it allows me to selectively attend to something "out there" and package it for someone else, as with example 3.

(3) There's a playground.

Relatedly, language is perhaps our most powerful means of expressing "negative content," as with sentence 4.

(4) There's no playground here anymore.[6]

Another thing that makes language such a powerful semiotic tool is it can represent our own powers of reason and package them for someone else, as exemplified in sentence 5.

[6] It is tempting to think that language actually causes us to 'invent the negative' (as Burke 1966 once put it). I would argue that the negative is indeed a semiotic but not a purely linguistic invention. It seems inventing the negative depends on counterfactual mental simulation. Upon seeing an empty field, I can match the present condition of immediate environment with past mental "maps" of it from long-term memory. Since pattern matching (mental simulation) operates within temporal spans much shorter than the time it takes to turn out a linguistic production (there does not seem to be much if any difference in production time between "internal monologue speech" and "public speech"), I assume I could invent the negative before I could symbolize the negative. I suspect, however, my power to symbolize the negative makes reinventing the negative faster and more complex; it takes up a greater portion of mental resources.

(5) There must be a playground around here somewhere, I hear the distant sounds of children playing.

Yet linguistic representation does not stop there; it also permits me to represent interactive situations, as with a speech act that, once uttered, places the speaker under a specific obligation. Sentence 6 exemplifies the speech act domain of language use.

(6) If elected I will make sure this park has a playground.

Perhaps most exotic of all, language permits me to represent the adequacy or inadequacy of instances of linguistic communication, as with sentence 7.

(7) Playground is not exactly the right word to characterize this place; it is more like a "recreational space".

*Cognitive grammar: construal and profiling; components and composites*

One of the most well-developed of cognitive linguistic theories is Ronald Langacker's *Cognitive Grammar*, a linguistic theory based on the notion that a language is a structured inventory of form-meaning parings, such that grammar and meaning are indissoluble, and that language has three components or "poles": the semantic pole, the phonological pole, and links between the two.

This view of attention outlined above applies to Langacker's notion of "construal," the relationship between the speaker/writer and the situation she represents. Since there is no single way to represent a situation, the grammar of a language provides users with a range of constructions, many of which differ subtly in meaning and satisfy varying semiotic and interactive goals (as was demonstrated previously with the detective's use of the agentless passive). Construal, then, reflects her ability to adjust the focus of attention by altering the mental imagery representing a situation. Cognitive Grammar also depends on the notion of "profiling," which refers to the effect of construal; for any linguistic act, some entity stands out in profile against a background. Contrasting figure/ground relations (Langacker calls

them trajectors and landmarks, respectively) influence meaning in profound ways. Varying profiles reflects our basic ability to conceive and portray the same objective situation in alternate ways for specific expressive purposes. “The symbolic resources of a single language,” writes Langacker,

> provide an enormous range of options for construing any given situation, and speakers show great dexterity in shifting from one to the other. (1998: 5)

Cognitive Grammar also claims to be usage-based; that is, conventional symbolic assemblies become entrenched by means of use in specific contexts. Instead of positing a schematic meaning of a lexical item *a priori*, its schematic meaning emerges from usage events. Take Langacker’s own example of the lexical verb *send*. For adult speakers, *send* has three conventional variants. The first variant occurs in a [[*send*] [NP] [NP]], as illustrated by example 8,

(8) Send Capital Hill a message,

where the indirect object and benefactor (Capital Hill) receives secondary prominence, because the focus of attention is most salient with the clause final direct object and mover (message). In this respect, the first noun phrase functions as conceptual ground for the second noun phrase, the conceptual figure. The second variant [[*send*] [NP] [[to] [NP]]], as illustrated by example 9,

(9) Send a message to Capital Hill,

highlights the goal (Capital Hill)as the most salient element of attention. In this construal, the indirect object stands out as a figure with the direct object functioning as the conceptual ground. The third variant [[*send*] [[for] [NP]]], as illustrated by the command in sentence 10,

(10) Send for the school nurse,

focuses most attention on the object or person (patient) the sender wishes to obtain. The verb *send* and its implicit agent function as conceptual ground for the entity in profile (nurse). All these variants of *send* exploit context-neutral schematization of transfer, a constant representation. But the crucial point is that these schema arise bottom up, as a result of using the same symbolic assembly in different situations. As

Langacker puts it, "a variant enters into a kind of "ecological system" with its structural context and does not necessarily exist outside that habitat" (1999: 91). The reason we have these different constructions for the same linguistic unit is due to the common sense condition that for any objective situation, speakers will invariably need or want to focus attention on different elements within it. Grammar is maximally redundant in the sense that it provides an indefinite range of productive and fixed (i.e., idiomatic) constructions that use the same representational material but configure the components differently for different purposes.

Construal and profiling are both tools of production and comprehension. In fact, the basis of variable interpretation is a matter of construal, an often *ad hoc*, opportunistic, and context sensitive operation. Consider my interpretation of the sign on the aluminum box in front of Kelvin Smith Library. The interpretation of the sign on the aluminum box LIBRARY BOOK RETURN depends on my ability to construe the conventional imagery of the English language in specific (often novel) ways. In this case, an interpretation refocusing my attention back to my original task depends on a somewhat novel interpretation of this syntactic group of nouns and nominals—an interpretation, I am sure, not foreseen by the sign's producers. My novel interpretation rests on the apparent fact that English nominal compounds are highly analyzable; that is, the contribution of component structures to a composite structure is easily recognizable; the greater its salience, the greater an influence it has on the interpretation of the whole. To aid explication, I examine this sign's default interpretation, then examine how this same pattern accommodates the peculiarities of my trip to Kelvin Smith Library.

As a linguistic construction, this form-meaning pairing consists of a *composite structure*, symbolic structures of any potential size assembled from fixed or novel *component structures* (of smaller size) according to regular grammatical patterns (see Langacker 1987, 1991, 1999).[7] One would likely analyze the composite as a blend of two component structures—the noun compound [library book]—

[7] Tomasello's verb island structures are component structures with open nominal slots that later combine into larger composite structures.

signifying an entity, and a second component—the compound [book return]—as two distinct but related facets of a domain whose composite structure relies on the shared sub-component [BOOK/book], with capital letters representing the semantic pole and lowercase letters representing the phonological pole. These two composite structures blend into a single composite to represent a specific kind of human action (returning something) involving a specific kind of artifact (a book belonging to an institution). The coded event, by implication, contrasts with a similar event involving a different kind of artifact (a book I own). This analysis seems appropriate when thinking about the fixed expression within a default social and historical context. Within the PUBLIC LIBRARIES schema, the compound construes not only an entity but ownership, too. And the notion of public ownership is a salient part of the circulation of library books: the person borrowing the book does not own it but merely possesses it for a specific period of time. Thus, the frame of library includes the rich cultural background knowledge epitomized in physical circulation of books to and from the library. The circulation of books seems a logical derivative of the institutional ownership of those books. The component structure "return" construes the completion of the physical loop of the circulation event. A default context analysis of this form would resemble this notation [[LIBRARY BOOK] [BOOK RETURN]] at the semantic pole; their integrated structure at the phonological pole would follow the notation [[library [book] return]] with the inner brackets marking the component structures and outer brackets marking the composite structures.

Signs, however, are always interpreted in very specific circumstances and are therefore open to *reanalysis* (Langacker 1977) or "change in the structure of an expression that does not involve any immediate or intrinsic modification of its surface manifestation" (Langacker 1977; quoted in Hopper & Traugott 1993:40). Since I am a faculty member at a research university who has taken out thousands of books from its libraries, I have internalized the library schema in such a way that I do not need to be told that an aluminum box (resembling a large mailbox) placed in front of a library is for library books only. In fact, the aluminum boxes at other library sites use the clipped form *book return*, assuming any passer-

by will need not be told that it applies only to library books. In the day-to-day behavior of faculty life, I doubt I even need the linguistic prompt "book return" to know what to do when I come to the library with Thomas Sebeok's *The Sign and Its Masters* in hand.

I argue that breaking down the composite into a different set of component structures is essential for effecting the shift of attention back to Kelvin Smith Library. In this instance, I see the whole structure but focus attention primarily on *library* as a separate component referring not to the entity that I would normally dispense into the aluminum box, but, metonymically, to the institution itself (after all, I did not come to return the book but to borrow it). In this scenario, the components are [library] and [book return] with [library] opening a mental space for the relevant institution and [book return] builds an embedded mental space for circulation of books, a representation of library at a finer level of granularity. Following Langacker's notation and terminology, one would say that [library] functions as the formal cue for [BOOK LENDING INSTITUTION], yielding the notation [BOOK LENDING INSTITUTION/library]. The fixed form [[library [book] return]] becomes as [[library] [book return]] at the level of phonology, and the default interpretation [[LIBRARY BOOK] [BOOK RETURN]] is reinterpreted as [[BOOK LENDING INSTITUTION][BOOK RETURN]] at the semantic level. I set *library* apart from the rest of the construction. The very typography of the letters (cf. materiality principle) facilitates reinterpretation of this compound because 'library' appears above 'book return' on the aluminum return box. In the default interpretation, the phonemic form [library] is not nearly as salient as [book return], because the focus of attention is on the function the box promotes; in the new interpretation, however, [library] achieves primary focus because the building and institutional status form the conceptual profile. Figure 3.1 represents the contrasting profiles of the default and contextual interpretations of the sign (see appendix).

Translated into mental space terms, two distinct mental spaces, one embedded in the other, emerge from the same organizing frame, but differ with differing degrees of focused attention (Fauconnier & Turner 1998:198) The second space for returning books is composed but not attended: it

fades into subsidiary awareness. It is the first space for the larger institution of libraries that is the focus of attention, and as the focus it comes to means something like *the place where I should be going, and a place to which I will eventually return, because whatever I borrow needs to be returned*. The initial word is important because it focuses attention back to my initial task. I can now admire the architectural style of Severance Hall and still complete the task I originally set for myself. The initial word "reminds" me of my task, while the second term serves as a reference point for reconstructing the scene in which I get Boswell's biography. A novel construal of the conventional imagery gives rise to a new profile relation.

I have presented some theoretical approaches to linguistic representation that place attentional phenomena at the core (rather than at the periphery), thus establishing a common vocabulary for this study. My next task is to connect these insights about language to mental space theory.

*Mental space grammar*

Mental spaces theory provides a general semiotic model to give a real-time feel to the analysis that meshes nicely with defining elements of cognitive grammar, particularly the notion that mental spaces are models we build to interpret past, present, future states of affairs, and, as such, each mental space construes these worlds in distinct ways, profiling some information over others. Additionally, we can assume that component structures can sometimes function as mental spaces in their own right and that their combination form composite blends (as suggested by Mandelblit 2000). Generally speaking, mental space organization calls attention to the "invisible" (Fauconnier calls 'backstage') semantic constructs that guide meaning construction. Fauconnier's mental spaces model, though not a linguistic theory per se, has shown that many of the classic problems of reference, logic, and formal semantics find natural solutions if mental spaces are posited as a basic feature of human cognition. In fact, mental spaces was initially devised to answer questions about indirect reference and referential opacity, only later did it prove to be useful for describing all sorts of semantic and pragmatic phenomena. In mental space theory,

linguistic structures are cues that prompt discourse participants to set up elements in referential structure, elements which may or may not refer to objects in the world, as mental spaces contain representations of entities and relations of a given scenario as perceived, imagined, or remembered by the participants. Metaphorically, mental spaces are temporary "containers" for relevant information about a particular domain or schema.

A typical conceptualization, linguistic or otherwise, comprises multiple mental spaces linked in particular ways and defining natural paths of access. If the variants of *send* enact slightly different scenarios, they reflect our ability to construct different mental spaces for representing a given state of affairs, real or imagined. Commentary on example 2, *All we know is that the victim was murdered with a scalpel*, will illustrate the usefulness of mental space theory for modeling discourse.

Grammatical structures like adverbial and prepositional phrases, deixis of time, place, and person, and determiners function as *space-builders* for new mental spaces or as *refreshers* for previously established but unattended mental spaces. The cognitive function of these grammatical forms is something analogous to the orienting and selection of attention; they, in effect, place interpreters in the proper position to select certain elements for additional processing. The adverbial clause, *All we know is that* builds a mental space that orients the audience's attention toward the speaker and his status as part of a detective corps. The pronoun "we" establishes the speaker's identity as spokesperson for a team (itself a conventional blend of input spaces for Speaker and Group) and the verb *know* both represents the speaker as someone who knows something and orients the audience's attention to discourse topic: the murder. The audience learns, for instance, that the collective knowledge about the event is small, as indicated by the negative polarity item *all*. This mental space, which I call the Investigation space, provides the present viewpoint for accessing another mental space constructed by the passive construction that follows the relative pronoun *that*. From a linguistic perspective, the adverbial phrase, *all we know,* and the relative clause marker, *that,* work into the attentional budget two different kinds of representations: speakers and

topics. The first, focusing attention on speaker's knowledge, the second switching attention back to the investigation. The second mental space, which I call the Murder space, is now in focus and provides information about a past event. It is past relative to speech event, the audience's attention to the speaker as an authority worth listening to is accessed through the Investigation space. The fact that there was a murder provides proper exigence for the detective's address.

While cognitive grammar provides a rich model for studying the link between semantic structure and phonological structures, mental spaces provides a rich model for understanding their real-time use in specific situations. In the current discourse space that comprises walking across North Campus to Kelvin Smith Library, the black letters on the aluminum box comprise sparse and efficient cues for rebuilding the initial mental space network I described in the previous chapter. The semiotic difference is that I use a different material marker as a cognitive reference point: a linguistic symbol, instead of the facade of a building.

**The linguistics of attention**

This partial survey of current linguistic theory should provide a sufficient starting point for exploring the theoretical premise of this chapter: the components of attention provide the sensational, perceptual, and conceptual anchor of language. I do not claim originality on this point, for as far as I have shown that attention has played a sometimes subtle, sometimes prominent role in linguistic theorizing. My goal is only to theorize globally what has already been done piecemeal by Deane (1991) on island constraints, Gries (2000) on particle movement, Tomlin (1997) on voice, Ariel (1990, 1999) and Givón (1985, 1994) on structures indicating high or low attentional salience (so-called Accessibility Theory), Chafe (1994: 53-70) on intonation units, and Talmy's (2000: 257-309) path breaking work the distribution of attention

as one of the three "schematic systems" of his cognitive semantics. [8]

My plan for shoring up this foundation is to discuss how the five components of awareness and attention can characterize a linguistics of attention. The principal difference I see between these other efforts and my own is that my approach does not begin with structure but begins with models of interpersonal use (i.e., the sixth component of attention) or what Benveniste called "*énonciation*" (1966: xx). To be more precise, human attention evolved in three dimensional spaces with other human beings; therefore, the theoretical modeling has to be explicit about the primordial settings of vocal communication—face-to-face, one-to-many, many-to-one, writer-to-reader, an so forth. The reason my approach uses mental spaces is precisely because mental space networks are intended to incorporate into their analysis specific conditions of enunciation. More specifically, the base space in mental space networks can be used to specify the precise conditions of the discourse. In any mental space analysis, for instance, the base space functions as the space to which the analyst can add or subtract specific information about the immediate conditions of the discourse—the who, what, where, when, and how of speaking or writing, hearing or reading.

Incorporating the components of subsidiary awareness into a linguistics of attention reveals the necessity of the base space.

*Alertness*, you recall, refers to an individual's general readiness to process incoming or new information. Recall as well the axiom: *not all incoming information is equally important*, and that *different organisms are alerted to different kinds of stimuli*. Human beings are highly attuned to human speech of any kind (i.e., phonetic recognition). In any given situation, the auditory cortices of my brain are primed to recognize incoming sensations of human voices (regardless of the language). When a

[8] In essence, Talmy argues that a schematic system for attention is necessary for effectively "distributing attention over a given structure from a given perspective point" (2000: 76). The attentional system has four dimensions: strength of attention (from faint to intense); patterns of attention (entities at the center are salient, entities along the periphery are faint); windows of attention (entities of a referent scene in attention are "windowed" while entities not in attention are "gapped"); and the mapping of attention (like Langacker's notion of "construal" refers to the possible ways a speaker can window or gap entities within a given referent scene).

human voice fills a silent space, I am automatically alerted to pay attention and process its message. The mere presence of voice "disturbs" present consciousness. Language is a powerful tool because the awake and alert brain will nearly always mind it; audible or scriptable voice is doubly powerful because it can alert us to something not present, operating as a *virtual alerting system* portable from situation to situation.

*Orienting*, you recall, refers to an individual's disposition to select particular kinds of incoming information over other kinds of information. When many voices fill a space, I am still alerted to voice but now have to select one and filter out the others while remaining peripherally aware of those other voices. Phonemic recognition is largely a function of the alerting and orienting insofar as I am more likely to be alerted to the sounds of English than any other language. Therefore, in a room filled with foreign voices, I will be specifically attuned to the sounds of English, so much so that should I hear someone speaking English I will turn toward those sounds in preparation to engage in face-to-face conversation. Phonemic recognition is the most basic linguistic process that relies on orienting awareness, but it is by no means the only one. Vocal intonation and syntax are two others. A shouting voice places everyone within earshot on high alert such that everyone will disengage attention and turn their bodies to face the direction of the sound. More subtly, the way a speaker pronounces a word or the way a writer manipulates the punctuation, typography, spelling, and layout of a text can affect how a hearer or reader regards or interprets the message, for in those instances she is disposed to do so in a particular way.

Recall that the sign for the library book return has a material instantiation that plays perhaps a larger role than usually thought in meaning construction (see materiality principle in Chapter 2). The fact that the sign was in large print and in English is important, but, for my purposes, what is more important is the arrangement of the individual words themselves:

(11) Library

Book Return

The fact that the word library is spatially isolated from the other words influences what I am going to pay attention to over other similar information. The isolation of the word *library* from the rest of the phrase adds salience, so that, at that moment, I walk by it I am primed to pay attention to that word over the others. Under these conditions, my paying attention to this sign could produce the serendipitous effect of reminding me of my primary task.

The fact that speakers of a language are predisposed to process certain structures over others has important implications for theorizing language change as well. Fennell (2001: 6) argues that one language-internal factor in English becoming an analytic language is that speakers developed a fixed stress on the nuclear syllable, drawing attention away from the final syllable and ultimately bringing about the loss of inflectional endings. In contrast, languages within the Indo-European family that allowed major stress on any syllable preserved their inflectional characteristics. Over generations, English-acquiring speakers automatically began to orient attention on syntactic placement when surveying the linguistic landscape, because word order became for them the primary means of determining grammatical relations. That is to say, orienting attention to the nuclear syllable of a word meant being alerted to changes in syntax rather than changes in morphology.

A polysynthetic language like Siberian Yupik (Eskimo) contrasts markedly with English in that attention to word order does not appear to be a viable disambiguating strategy, because one lexical morpheme often incorporates a complete English sentence. Consider example 11 from Comrie (1989:45)

(12) Angya-ghlla-ng-yug-tuq
Boat-AUGMENTIVE-ACQUISITIVE-DESIDERATIVE-3PERS SING
'He wants to acquire a big boat.'

The expression contains only one lexical morpheme, *angy* (boat), followed by a series of grammatical suffixes: *ghlla* (an augmentive); *ng* (an acquisitive); *yug* (a desiderative); and *tuq* (a third person singular pronoun). The Eskimo-acquiring speaker is alerted and oriented not to word order, per se, but to word-

internal components that reflect pragmatic order rather than grammatical order

When cognitive linguists talk about linguistic structure, they refer to stable repertoires of signs operating in subsidiary awareness, with phases of alerting and orienting operating reciprocally and in mutually defining ways: phonetic awareness means we are alerted to human speech; phonemic, morphemic, and syntactic awareness means we will orient ourselves to particular kinds of human speech; and prosody means we possess a linguistic orientation which, in turn, determines language specific altering operations. Such "pre-attentive" phenomena serve as the perceptual ground for explicit attentional processes.

All the material elements comprising a "referent scene" can be incorporated into a general semiotic theory that emphasizes the qualitative, coincidental, and conventional characteristics of signifiers (see chapter 2). With respect to mental space grammar, these concerns often fit within descriptive apparatus of the Base space.

*Selection*, you recall, refers to initial assignment of an item or items from perception or from short-term and working memory into the attentional budget. Selection refers to incoming information that becomes the content of a mental space, a participant, a role, an object, artifact, event, action, or abstract idea. Proper nouns, common nouns, indefinite articles, deixis of place and time, verbs and other linguistic foci (e.g. adverbial and prepositional phrases) are typical elements of linguistic construction designed to select entities, objects, and relations for further processing. Selective attention explains Talmy's (1995, 2000) notion of "windowing" and "gapping". Given a kind of referent scene, for example, a motion event with a conceivable initial, medial, and final image schematic organization, the options available to discourse participants are three-fold: windowing all features, windowing two and gapping one, or windowing one and gapping two. Sentence 13,

(13) I went to the library,

windows the final point and gaps the initial and medial points of the referent scene, whereas sentence 14,

(14) I left my office and went to the library,

windows the initial and final points but gaps the medial point of the referent scene, and finally sentence 15,

(15) I was walking along Ireland way when I was overcome by the smell of fresh-brewed coffee,

gaps the initial and final points in a previously selected referent scene. To generalize even more, the introduction of a distinct lexical form implies a high degree of salience in the mind of the speaker; however, by introducing it she presumes a low degree of attentional salience in the mind of the hearer. Pragmatic notions of old information and new information, Grice's (1975) cooperation principle and his maxims of conversation (particularly relevance) are deeply embedded in the structure of language, if that structure comes from instances of use where the objective is to say something you think someone else does not already know.

The prototypical linguistic forms of selection--first person pronouns, nouns, noun phrases, and verbs--function as *stable attractors* in the linguistics of attention: their function is to share attention about something. For example, the act of calling out a person's name at a crowed party will alert the person so named and will likely cause him to orient his body toward the sound source. From the speaker's perspective, however, the utterance reflects an already well-developed micro-event, whereby she has scanned the crowd, selected attention on an entity, recognized it, classified it, and processed it. From the addressee's perspective, hearing the sound of his own name is a reliable cue to shift his attention. The shifting of attention is a *catastrophe* in the grammar of attention, meaning that the micro-event of hearing his own name being called out perturbs present time-space, bringing about a sudden change in the hearer's conscious present. Stable attractors, then, are reliable structures for bringing about such catastrophes. In this respect, the selection has much in common with catastrophe theoretic semantics as described by Wildgen (1992: ch. 2) insofar as linguistic structure reflects verbal routines for "disturbing" the semiotic

landscape in ways that will attract the attention of those who share the same space.[9]

With respect to mental space grammar, selection of attention applies to all three space types: Base, Viewpoint, and Focus. When the act of speaking gets explicitly marked, then a momentary selection for the speaker as the speaker takes places. Often it is the case that such evocations also function as a viewpoint for selective attention to another topic.

*Sustain*, you recall, refers to the continued processing of an entity, event, action, or relation as it pertains to remembering, reasoning, planning, deciding, or acting. Sustain or focus of attention means adding new but closely related information to the attentional budget. Pronouns, reflexive pronouns, appositives, restricted relative clauses, prepositional phrases, definite articles (among other devices for achieving cohesion and coherence) are elements of linguistic structure designed to sustain attention by focusing in and elaborating on the selected entity or topic. If selection of attention can be said to disturb space-time, sustain of attention can be said to globally *stabilize* a mental space network. I say globally stabilize, because sustain of attention also entails many local "disturbances" of mental spaces, since the speaker is adding new conceptual structure. For example, the discourse developing after the speaker calls out the hearer's name stabilizes the present space-time in that it defines who the discourse participant will be at the same time that subsequent speech acts disturb the conscious present by use of additional stable attractors, thereby adding structure to existing mental spaces or thereby building entirely new mental spaces (a process we can liken to "micro-catastrophes") within an otherwise stabile base space. For instance, the speaker introduces the addressee to a third party, thereby attracting his attention to a new being in the conscious present.

[9] I risk doing great injustice to catastrophe theoretical linguistics by incorporating using their own terminology for my own theoretical purposes. I want to make it clear that I find these terms useful for describing the dynamic relation between the form and meaning in specific situations; but my appropriation is by someone who has yet to familiarize himself with many of the intricacies of this approach. Suffice it to say that I regard their general approach, namely that spatial relations in perceived scenes form a basic set of semantic archetypes for language and that perceptual processes themselves provide individuals with a repertoire of cognitive routines that formally instantiate abstract semantic archetypes into the structure of language, as very promising (see Petitot 1995; Wildgen 1982).

Together, selection and sustain of attention constitute the attentional budget as it composes, completes, and elaborates a network of mental spaces (many of which are blends) developed to represent and interpret real or imaginary thoughts and actions. The presentative conditions for representing and interpreting and interacting with them is the special domain of the linguistic construction.[10] When our fictitious detective uses the definite article as part of the clause subject, *the victim*, he assumes his audience is already attentive to the present situation (i.e., that there is a victim) and is ready to process new information about it.

In English, selection of attention functions from within clauses, whereas sustain of attention can function from within clauses (i.e., argument-predicate relations in a simple declarative sentence, for example) and among different clauses (i.e., co-reference and anaphora in complex sentences, for example). In the Siberian Yupik, selection of attention functions (*ceteris paribus*) from the nuclear lexical morpheme (i.e., *angya*) with sustained attention functioning along the outlier morphemes (i.e., *yug-tuq*). Hence, one difference between these two languages is that in English the experiencer role ('he') functions as the initial attractor from which the rest of the referent scene coalesces, whereas in Siberian Yupik the patient role (*angya*) functions as the initial attractor from which the rest of the referent scene coalesces. While each example tells the same story, each example chooses a different starting point for telling it.

*Control*, you recall, refers to the ability to perform two tasks simultaneously (divided), or to start one task, put it on hold for something else, attend to that something else, and return to the primary task (switching). In other words, control of attention means coordinating more than one mental space network in present consciousness, as when I switch between going to the library and admiring architectural features of a building. Since complex semiosis depends on felicitously controlling the attentional budget, language is no different. In fact, close examination of spoken and written discourse reveals successful and

[10] The present description of selection and sustain agrees with the principle hypothesis of Accessibility Theory (Ariel 1999, Givón 1994, van Hoek 1997), which states that the appearance of full noun phrases (names, common nouns, descriptive phrases, and nominalizations) indicates low accessibility of the current referent in the hearer's consciousness. In contrast, the appearance of grammatical reference points (pronouns, reflexives, and definite articles) indicates high accessibility of the referent in the hearer's consciousness.

unsuccessful attempts to control information flow. Discourse markers (e.g., *now*, *anyway*), adverbial phrases, nonrestrictive relative clauses, complement clauses following epistemic and speech act complement- taking verbs (e.g., *I presume that. . ., I propose that*. . .), register shifts, deictic gestures, parenthetical remarks, rhetorical figures like *ploche*, *anaphora*, and *antimetabole,* and footnotes are among some of the linguistic and graphic elements designed to control the flow of information when speaking and listening, writing and reading. If selection and sustain govern topic and focus in language at the prosodic, lexical, phrasal, clausal, and sentential layers, control of attention governs the switching of topic and focus at the phrasal, clausal, sentential and discourse layers.

For example, suppose our *bon vivant* interrupted a conversation between the addressee and his wife. The addressee's ability to divide attention and negotiate both conversations in a time-locked fashion entails holding two different stable situations within the larger situation (a party) in working memory. A phrase like *excuse me* addressed to his wife is a good examples of an attention shifter insofar as it instructs the addressee to hold the now suspended topic in working memory; a question like *Now what were we talking about?* is a good example of an attention refresher inasmuch as it refocuses attention back on the suspended topic; and a declarative sentence like *We were just talking about Japanese gardens a moment ago* is a good example of an attention sharing device, inasmuch as its indirect illocutionary force and perlocutionary effect invites the interlopers (i.e., the person who called out his name) to join the conversation.

Consider once again sentence 2.

(2) All we know at this time is that the victim was murdered with a scalpel.

The mere fact that our fictitious detective arrives on scene, stands in a particular spot, and appears about to speak is enough to alert and orient the attention of the fictitious reporters. Utterance of the initial adverbial phrase, *All we know at this time*, is enough to select attention back to the Base and the Investigation spaces by alerting and orienting the reporters' attention to the immediate conditions of the

discourse. The pronoun *we* focuses attention on the fact that the speaker is a member of the investigating team, further selects attention on the speaker as holding the floor (i.e., "you reporters need to pay attention to my voice"). The quantifier *all* and verb *know* has the effect of refocusing attention to the Investigation space; and the prepositional phrase *at this time* establishes a temporal connection between the Base and Investigation spaces such "all that we know I am about to share with you". Once the speaker introduces the complement clause with *is that*, the focus of attention shifts from matters of enunciation to matters of the investigation. This example demonstrates that control of attention has a fundamental place in linguistic theory, because the implicit (metalinguistic and speech act) conditions of the discourse and the exigent matters (preestablished topic) motivating the act speaking must be coordinated in and over small intervals of time. It is possible for the speaker to come out and say *The victim was murdered with a scalpel* with no rhetorical preparation, but such occurrences are rare. It does seem that discourse coherence demands that we refresh the conditions upon which speaking and writing take place, hence the need for discourse participants to control attention among the cooccurring content, metalinguistic, speech act domains of language.

As a summary statement on these matters, I have condensed this theoretical sketch into table 2.1.

@@Insert Table 2.1 Here

Formal linguists like Chomsky like to "bracket" phenomena from the contexts of its use. Understanding what it means to engage in discourse in the first place seems quite irrelevant. A linguistics of attention takes the opposite view that levels of linguistics, though distinguishable, interpenetrate because linguistic structure is an *emergent property* of utterances. Hence, distinctions between phonetics & phonology, morphology, syntax, and semantics make sense as a byproduct of an entrenched pragmatics, itself a byproduct of shared attention. A linguistics of attention takes the broad view, articulated most clearly and forcefully by Hopper (1998), that a grammar is forged during actual discourse out of utterances accumulated from previous situations, leading to a notion fundamentally alien to *a priori* theorists that

grammar is intrinsically “unstable” and “indeterminate”. I then concur with Hopper that it is best to think of speaking as “remembering procedures” than “following rules” (1998:167).

The format of table 2.1 reflects this set of priorities. In order to decipher table 2.1, consider the top row (with its thick borders) to specify the grammar of attention with shared attention as the primordial phylogenetic and ontogenetic activity in constant “dialogue” with the remaining five components of awareness and attention. Consider as well, the left-most column (with its double-line borders) to specify conceptual domains associated with verbal behavior: human semiosis itself, verbalization, eye contact and other face-to-face phenomena, gesticulation, rhetorical modes of address, forms of talk, print technology, and, finally, mental space grammar.

The corresponding cells of each row (left-to-right) then specify specific metalinguistic, paralinguistic, and linguistic phenomena as it relates components of attention. For instance, verbalization at the presentative condition of the sign categorizes phonetics and signals under the alerting of attention, phonology and phonemic recognition under orientation (as outlined above), whereas morphology, syntax, and lexical semantics comprise the representative and interpretive conditions of the sign when the attentional load is the heaviest for the discourse participants. With respect to mental space grammar, table 2.1 emphasizes my point that shared attention presupposes that discourse participants construct sufficiently similar mental space networks, in part because the attentional processes of two or more individuals operate in like manner because they share similar sensory and perceptual categories as well as similar material conditions of living in the world. The moment-by-moment delegation of mental spaces as Base, viewpoint, and Focus plays out in similar fashion along the alerting-controlling cline. However, one should not conclude from this presentation that these categories are impenetrable; rather, quite the contrary, they interpenetrate, probably in ways neither I nor anyone else can fully appreciate. But it does suggest a cognitive justification for these domains and levels: that with great semiotic complexity comes greater investments in attention and time. In order for something or someone to be marked for attention,

the individual attending to it must be alerted and oriented to it; once an individual makes the initial verbal investment in an object of attention, her or his cognitive resources will go toward maintaining that object of attention, relating it to other objects of attention, and so on. As one moves from left-to-right, one moves from lesser to greater commitment of verbal and conceptual resources.

Table 2.1 suggests that some linguistic phenomena only provisionally line up with a specific component of attention, for, given the right circumstances, anything can be marked for attention. For example, it is typical that when I read for content I apperceive the layout. I do not ruminate on its meaning. In such a situation, text layout stays at the presentative condition of the sign, inhabits subsidiary awareness, and can only be described as an element in the base space that is accessible but not active or salient. The discipline of graphology, however, attests to the fact that such phenomena can be the object of selective and sustained attention in special conditions. For instance, I can focus attention on the specific typographical properties of long-s as telling me something about early-modern print technology, rather than as representing a voiceless frictive. Similarly, lexical items may take a backseat to grammatical items, should a speaker shift emphasis, as exemplified in sentences 2b- 2d.

(2b) The victim was murdered ˈwith a scalpel.

(2c) The victim was murdered with ˈa scalpel.

(2d) The victim was murdered. . with a scalpel.[11]

In (2b), the speaker places primary accent on the preposition calling attention to the instrumental status of the noun; in (2c), the speaker places primary emphasis on the article preceding the noun, perhaps implying that the investigators do not actually possess the weapon, but rather that identification of the weapon was made indirectly (through an examination of the wounds, for instance); in (2d), the speaker pauses briefly between the main verb and the prepositional phrase, suggesting that the fact of its being a

[11] An apostrophe appearing before the word indicates primary accent with noticeable pitch deviation; it can be accompanied by loudness or lengthening of the vowel. A two-dot ellipsis (. .) appearing after a word indicates a brief pause or break in the flow of information. A three-dot ellipsis (. . .) appearing after a word indicates a lengthy pause of up to a second.

homicide is not established in the addressee's minds, making *murder* the linguistic cue for selecting "means of death" as the new topic (as opposed to some other cause) and making *scalpel* the linguistic cue for sustaining attention on the "means of death" topic.

The linguistics of attention points to the notion that linguistics is a subset of cognitive semiotics, and thus, descriptions of linguistic phenomena need to satisfy principles of subjectivity, intentionality, temporality, simulation, publicity, and materiality. Subjectivity and Intentionality principles force linguists to treat issues of subject traits and motives (their respective "footings" as Goffman (1978) terms it) as a constitutional component to the description of language structure and use, rather than incidental desiderata. One cannot account for the different points of emphasis illustrated in sentences 1b-d without accounting minimally for the implicit conditions governing the discourse. The components of attention predict that these differences are motivated by the speaker's tacit understanding of what the addressees already know, what they want to know, and what the speaker himself thinks they need to know, to name a few variables. These specific questions go directly to issue of attention.

The temporality principle is an important variable for theorizing the flow of information. The short pause in the flow of discourse cues the addressee to attribute importance to the last thing said, because the punctuated silence selects that last element for continued processing. The next three second interval is likely to focus on the murder weapon now that the speaker's brief silence affords the addressee time to process information. In marking the mid-clause verb as salient, the speaker ensures that the proceeding prepositional phrase will be embedded within that topic.

The simulation principle predicts that elements occupying selective and sustained attention will gain prominence in the mental images and mental models addressees construct. Emphasizing the preposition *with,* for instance, may cue the addressee construct a mental image of a hand wielding a scalpel, whereas emphasizing scalpel instead may attenuate the prominence of hand imagery in mental

simulation tasks.[12]

The publicity principle predicts that these shifts in emphasis can reflect issues of rhetorical footing as alluded above. The materiality principle predicts that most presentations operate below the threshold of conscious introspection, but on analysis show themselves as a great influence of language development, structure, and use.[13]

If it does nothing else, this theoretical sketch maps future lines of investigation.

**Descriptive applications**

I now proceed to examine a selection of English language examples from a mental spaces and blending framework, the purpose of which is to provide readers with a preliminary view of a linguistics of attention at work. The constructions examined--verbs and auxiliaries, verb collocations, idioms, reflexive pronouns, morphological affixes, and a passage from Boswell's biography of Samuel Johnson--make up an impressionistic and piecemeal sampling of attentional phenomena. Although impressionistic, the cumulative effect I hope to achieve is that the job of the linguist is not complete until she has given a plausible account of how language interacts with the immediate conditions of production and interpretation (as best they can be reconstructed), and that doing her job requires a theory of attention.

*Go*

In this section, I will consider variants of *go*, with special reference to diachronic changes in its meaning and grammatical function.[14]

[12] I know of no experiments testing this hypothesis, but one can could be devised. Subject her and report on the content of their mental imagery after hearing one version of the same sentence with different prosodic features.

[13] See the previous chapter for a discussion of the role material culture plays in language acquisition.

[14] Hopper and Traugott (1993) use *go* to open their discussion of grammaticalization; however, they do not specify

Grammaticaliztion processes are motivated by satisfying different local expressive functions. I claim (but can only speculate at this time) that metamorphosis of a linguistic form (i.e.changes in meaning (expansion or contraction), syntactic variation, morphological and phonemic changes) reflect a functional history of that form as a means of selecting, sustaining, and controlling attention.

*Go* entered the language as a lexical, intransitive verb coding directed motion, as with example

(16) I *go* to the library every Thursday.

Later, its meaning expanded to indicate intentionality and with it the expansion of form as an auxiliary verb, as with

(17) I am *gonna* go to the library tomorrow.

Later, the meaning of the verb has expanded to include a citational function, as with

(18) I told him to stop teasing her, and he *goes* 'mind your own business.'[15]

Historically, this verb displays a steady expansion of meanings of grammatical form beginning with the denotation of directed motion, proceeding to denote intentionality, and proceeding further to denote speech metaphorically understood in terms of intentionally directed motion. Is this diachronic shift in meaning and function arbitrary? If not, what might explain it?

An answer: our attentional system is on the look out for entities moving through space; motion is an unavoidable condition of human semiosis. We are alerted and oriented to pay attention to moving entities; if something nearby moves, our inclination is to track its by placing our bodies in position to survey it (hopefully at a safe distance). That it moves guarantees it will become an object of attention. Individuals will select it for further processing and evaluate it (usually as dangerous or benign, undesirable or desirable) by sustaining attention on it. Such immediate assessments permit deliberate action and planning. If the person sensing the motion evaluates the moving entity as harmless and irrelevant, she will likely return to the task she was performing before these sensations flooded her

---

its quotative function.

[15] The earliest recorded instance of citational *go* appears late in the eighteenth-century.

attentional field. If the entity in motion is considered significant, she might sustain attention onto it, or, if she is not alone, might use the repertoires of linguistic symbols to control the attention of another, as with the exclamatory expression

(19) Look, there goes a hawk diving after that squirrel!

As I claimed in the previous chapter, human beings not only pay attention to entities that move, they also interpret how and why they move. If an entity seems to move without any external force behind it, such as a person walking along a path, that entity seems to be harboring intentions and volition, and it is usually up to us to guess what those intentions might be. Language evolved, in large part, to externalize these intentions. Directed motion makes accessible intentional motion. It is reasonable to conclude that a verb symbolizing directed motion can easily come to symbolize intentionality. The evolution of *go* suggests that English speakers expanded the meaning and the grammatical form to allow for motion and intentionality to be separately but coordinately represented. Once alerted to the fact of motion and oriented to interpret that motion as intentional, an individual can select and sustain attention on a moving entity to consider what it is doing, how it is doing it, and why it is doing it. As an auxiliary verb, *going* represents intentionality.

Directional and intentional *go* reflect so-called primordial events, the representation and interpretations of which take up the lion's share of mental processing. Given their continued prominence in cognition, it is reasonable to assume that primordial events such as directed motion can serve as a analogical base for representing and interpreting complex verbal behavior.[16] Face to face conversation has many literal parallels to directed motion: voice emanates from a source and travels through space to a destination; the source appears to have voluntarily chosen to utter sounds, leading to the default

[16] One might consider verbal conversation a primordial event. Although it is a highly entrenched activity, one that defines our species, it would be a mistake to equate detection of directed motion in others and automatic production of directed motion with vocalization. Human beings survive without vocalizing, many (autistics, for instance) even survive (albeit more precariously) without even an understanding of what it means to converse verbally. No human being will survive on her own without some means of detecting, representing, interpreting, and producing directed motion.

assumption that the body producing the audible voice acted intentionally; sources and recipients of verbal messages try to orient their bodies into positions optimizing transmission and reception of utterances, a rough analogue to optimizing the three-dimensional path entities and objects travel. Just as people and things go from one place to another, so human vocalization can be understood as going from one place to another (see discussion of image schemata in the introduction). Just as human beings are always on high alert for organisms and objects coming their way, so are they on high alert for words and phrases coming their way. Just as human beings interpret the intentional meaning of an organism or object coming their way by trying to assess the source from which it emanates, so do they orient their attention to the source of the voice to interpret its intentional meaning by blending metalinguistic functions with body motion.

In characterizing directed motion as "coming their way," I bring up an important issue about the meaning of intentional and citational *go* that deserves comment. Its citational function allows the speaker to simulate the perspective of the represented speaker, since *go* means observed external behavior that indicates an internal intentional or volitional state on the part of the moving body. As a space builder for representing the words of another, *go* provides cues whereby the actual "speaker" in the Enunciation (base) space, takes the perspective of the fictive "speaker" in the Conversation space that, counterfactually in this case, represents the actual speaker as the "recipient" of the quoted message. The use of go in sentence (18) is a conventional blend in which the value of the "speaker" role in the base (Enunciation α in Figure 3.2) and the value of the "addressee" in the Conversation space combine into the role of "animator-addressee" in the blend (so named Enunciation β). In the blend, the speaker in the base enacts in real- time the scene imputed to have occurred in the recent past, bringing it into the conscious present. In the blend, but in neither input space, the voice of speaker α and speaker β collapse to identity (via metonymic tightening). The desired effect is to focus the addressee α's attention on the words of speaker β as if she were present herself (i.e. direct discourse). The blend serves as a staging ground for shifting attention from Base space to the Conversation space. The speaker can even shift the viewpoint as

well as focus to the Conversation space by mimicking the pitch, stress, and intensity of speaker β's words (see Figure 3.2 in the appendix).

In essence, the verb establishes the perspective as the Conversation space rather than the Base space. As a test, try substituting the verb *come* for *go*, as in sentence 18a,

(18a) * I told him to stop teasing her, and he comes "mind your own business".

This sentence does not make much sense because it conflates representations of a past conversation with the presentation of the present conversation: the speaker wants the recipient to pay attention to her represented self as the recipient of a verbal assault, she cannot do that in the base space, since, in this space, she is the sender of the message, not a recipient. It is possible for the actual speaker to represent herself in the base space as the recipient of the words, but only with respect to a special framing device, such as artistic or divine inspiration, where the external source is "offstage," as with sentence 20's prophetic attribution of authorship:

(20) These words come to me from the Holy Ghost.

Understanding the proper use of the citational *go* means simulating shifts in perspective in order to faithfully represent the back-and-forth dynamic of face-to-face conversation. Acquiring this verb as a reliable linguistic unit entails, I speculate, a prior ability to follow the flow of conversation in real-time (i.e., attend to the actions of its participants), abstract it as a kind of repeatable activity, and model that reality in different representational formats (language being perhaps the most conspicuous format).

As the history of *go* indicates, its deictic form and that of its companion, *come*, loom large in English as a means of drawing attention to relations stemming from primordial events of directed motion and intentional action.[17]

Consider this English idiom epitomizing karmic retribution,

(21) What goes around comes around,

[17] Add to the linguistic palette the verb *get* and one has the representational elements necessary to structure verbal accounts of my trip to Kelvin Smith Library, and indefinitely other events and actions.

and its commutated counterpart,

(22) What comes around goes around.

Suppose a speaker angrily says to her interlocutor, Sam:

(23) I know what you did to Ethel! You ought to be ashamed of yourself.
Remember, what goes around comes around!

Now, contrast the first version with this one:

(24) I know what you did to Ethel! You ought to be ashamed! Remember,
what comes around goes around!

I defer commenting on their differences until I comment on their common elements, namely the discourse marker *remember* and the relative pronoun *what*.

The verb, *remember*, connects the ensuing piece of wisdom with the previously established discourse topic, Sam's behavior toward Ethel. This linguistic form functions to control attention. Absent this marker, it is possible for representations of going and coming to distract Sam from the discourse topic. In short, the control of attention functions to sustain Sam's attention on the situation with Ethel and to ensure the separable tasks of coming and going blend with and do not distract from the established topic--the cause of this karmic coming and going of unspecified bad fortune. Both expressions place the relative pronoun *what* in the subject position. This null-subject cues us to place onstage the whole situation involving Ethel and Sam. The speaker assumes (cf. accessibility theory) the mere mention of Ethel's name will be a sufficient cue for Sam to reconstruct theses past events and sees no need to recount them verbally. Hence, the focus of attention can be placed on the speaker's generalizations about the effects of Sam's actions and how those actions, in turn, will determine his destiny. In this respect, the use of the null subject performs the dual function of selection and sustain of attention by refreshing an already established mental space and adding new structure to it. The speaker points out something she thinks Sam will already recognize, but she also thinks that sustained attention on what he already recognizes will

force him to reinterpret his actions.

Let us now consider the contrasting syntactic placement of *come* and *go*. According to Fillmore (1971:99) the prototypical function of *go* is to signal movement to a location not occupied by the speaker at reference time, whereas the prototypical function of *come* is to specify a destination attributed to the location of the addressee at reference time. In this case of coordinated coming and going, however, the deictic verbs do not apply directly to the speaker and listener in the Base space, but rather to the focal participants (Sam and Ethel) in the reference or Event space.

The first version of events generates a mental simulation of Sam's conduct that begins with Sam (as agent) and his intentional, volitional conduct toward Ethel (as patient) at time *T*, thereby implying an ensuing consequence for Ethel, the recipient of his action at time *T+1*. (Sam is represented implicitly as an agent moving toward Ethel.) With the verb phrase *goes around*, the speaker then construes a counteraction (by an unspecified agent working on Ethel's behalf) that will produce an effect on Sam at time *T+2*. Thus, attention to intentional *go* at the beginning of the clause shifts to unintentional *come* at the end of the clause, with the understanding that the implicit filler of the agent argument at time *T* fills the implicit patient argument at time *T+2*. Sentence (23) profiles Sam as a fixed reference point for the events corresponding to times intervals *T, T+1* and *T+2*.The initial perspective is with Sam and his action(s).

The second version of events generates a mental simulation of Sam's conduct from an opposing perspective. Ethel is now the cognitive reference point for the simulation, which begins with the consequences of Sam's actions on Ethel at time *T+1* and proceeds to represent a counter action visited upon Sam at subsequent time *T+2*, leaving offstage the initial actions in favor of profiling its initial effects on Ethel and subsequent effects on Sam. In other words, the initial perspective is with Ethel.

This example reflects two distinct ways of construing the conventional imagery of a language for specific expressive purposes, each selecting different starting points for running mental simulations about

the same past action. Sentences 23 and 24 function as alternative organizations of two entrenched component structures whose syntactic placement changes subtly the meaning of each, because the opposing participants enjoy different degrees of prominence. Translated into mental space grammar, each construed sentence builds two sets of two temporally distinct mental spaces embedded into a larger narrative framework. Figure 3.3 presents the contrasting configurations of mental spaces for each example (appendix).

*Pointing and clicking*

Anyone fluent in contemporary standard American English will recognize the sentence,

(25) Come visit us on the web at. . .

as a frequently occurring invitation in the world of "e-commerce". Anyone fluent in standard American English will likely interpret the utterance

(25a) Go visit us on the web at . . .

as a command rather than an invitation. It is typical for addressees of such messages to take an object-centered perspective of the web site itself. We can do so because of the presupposition that websites are the creation of conscious, intentional agents like ourselves, and the conception that these texts are the products of human subjectivity and intentionality in an over time. The invitation emanates from an intentional subject hailing the addressee-browser.

With the recent ascendancy of e-commerce as a dominant economic force, the verb and noun collocation *point and click* has acquired rich constructional variants in English.

As a frequent web browser, I routinely receive e-mail messages in which the subject line reads:

(26) Grocery savings are *just a click away*.

A recent television advertisement for *Ditech.com,* a site offering low-interest home-equity loans (often lent for the purpose of consolidating debt), begins with this slogan:

(27) So you've gotten yourself into too much debt. Now you can *click your way* out of high interest rate, non-tax-deductible credit-card-debt with Ditech.com.

I will submit both examples to extended analysis; however, I do so only after analyzing them as variants (i.e., clipped forms) of the productive verbal construction with special reference to the lexical contribution of *way*. First, the verb phrase *point and click*.

I begin with this generalization: *point and click* lexicalizes a sensorimotor routine that usually operates at the level of proprioception during fluid performance. In other words, those well versed in computers do not have to focus explicit attention on what the hand is doing with the mouse, nor do they have to hold in working memory the explicit proposition that "clicking the mouse button will change what appears on the computer screen". In the manner described above, the utterance of this verb phrase "disturbs" the conscious present (it is a stable attractor) by shifting momentarily the figure and ground, so as to foreground the background of the represented ongoing activity. Once in the foreground, content associated organizing frames and mental models other than computer use can be integrated with that domain of computer use, forming a single, coherent scene whereby the micromovements of the person sitting at a computer terminal using Netscape Navigator or Microsoft Explorer relate causally to remote actions and effects associated with the macroscopic of this other domain.

A blending mental space network analysis for *pointing and clicking* as an explicit conceptualization can be summarized in the paragraphs that follow Figure 3.4 (appendix).

Figure 3.4. presents a schematized network of three spaces, an input space corresponding to the ongoing activity of browsing the World Wide Web, a space that can reflect the addressee's actual ongoing activity (as in sentence 26) with respect to Base space or a mental simulation (as with sentence 27) with respect to Base space. The Web Browsing space contributes the participant role "Browser" (x), whose

value matches that of the "Addressee" (a) in the Base and the element "Website" (z).[18] The most salient material anchor for this space is the computer mouse and the clicking sound it makes. There is a structured coupling between the feel and sound of the mouse when manipulated and the organizing frame used to build this mental space, which means that the material properties of this space can be salient or active in attention or remain dormant but accessible.

Linguistically, the verb *point* represents (via metonymy) a computer user grasping a mouse with one hand, moving it around a pad resting on a table and visually tracking a cursor (arrow, white hand with extended index finger); *click* represents (again by metonymy) the same user pressing one of the mouse buttons that produces a familiar sound and visually witnessing the cursor change from a pointing hand to an hourglass (assuming a default setup) and, eventually witnessing a change from one site to another. While the verb *point* can encode a number of actions and possible affordances, its pairing with *click* narrows the interpretation down to the haptic, auditory, and visual sensations accompanying its use (i.e., its resistance or lack thereof when being moved, its smooth trajectory, the sound it makes as it moves over the pad or table, and the cursor's trajectory). The verb *click* encodes haptic and auditory sensations (i.e., the browser feels the short downward movement of the button, feels its resistance, and hears a short, sharp sound) as well. The conjunction, *and*, encodes sequence (not just collocation), so that the act of pointing the mouse and visually tracking the cursor movement on the screen precedes the act of clicking the mouse and seeing the hourglass and the site change. Sequentiality becomes evident if you try to alter the order of the two verbs (as one can do with *come* and *go*). I found no instances of *click and point* or *just a click and point away* in corpus searches on Lexis-Nexis or CoBuild databases.

My searches revealed additional formal constraints. For instance, one can vary the inflectional morphology of the two verbs, but whatever change is made to the first verb has to made to the second verb as well. This symmetry constraint means that changing first verb to *pointing* necessitates changing

[18] I use the designation "browser" to refer to the person sitting at the computer terminal and not its more common designation as the program enabling the user to "browse" the World Wide Web.

second verb to *clicking* to preserve meaning. One can modify the idiom to *pointing and clicking* but not to **pointing and clicked*, and (at least in my grammar)*?pointing and click* is an unlikely (though perhaps not impossible) combination.

Taken as a unit, *point and click* and its inflectional variants, prompt interpreters to construct a causal action chain between user, input device, terminal, and site (output information). Perspective is another salient property of this idiom. Since the causal action chain accessed through the sensations associated with the human-computer interface, an implicit but important element is the construction of a fictive user sitting in front of the computer. The mere presence of the construction prompts the reader to (at least momentarily) adopt the perspective of the user—which is an easy, automatic feat if one encounters this combination of verbs while surfing the web or reading email. As we shall see momentarily, speakers of English can choose which possessive pronoun comprises the direct object of the composite construction, either *your way*, *their way*, *our way*, *his way*, *her way*, or *its way*, respectively. Thus, the pronoun choice can suggest whether the addressee is being asked to cast herself in the role of "browser," or whether to cast a third party in that role. With respect to perspective, we take the evoked physical (ego-centered) position of the user as she interacts with a computer (not the object-centered perspective of the computer or website); with respect to social framing, we take on some of the possible roles associated with browsing websites, but which the verb component itself leaves unspecified. I am thinking principally of such social roles as "consumer," "researcher," "office worker," or "professional". These participant roles come online with respect to the second input space.

In the schematized analysis, input space 2, the Telos space, operates as the topic space (the darker circle represents the intentional priority of the Telos space in the network), containing conceptual structure discourse participants are supposed to care most about. The organizing frame for Grocery Shopping comprises the Telos space for example 24, while the organizing frame for Debt Consolidation comprises the Telos space for example 25. Schematically, the Telos space contributes the semantic roles

of agent (x'), action (y'), and location (z') with the value of the agent role corresponding to an intentional subject, the value of the action role corresponding to a particular happening initiated by said agent, and the location role corresponding to a particular place (or institution) where the agent acts. In this schematic version of the mental space network, none of these roles has a specified value.

The blend recruits structure from the Web Browsing and Telos space to generate a representational scene whereby the "browser" and "agent" fuse, becoming the "peruser-agent" (x") therein. I change the designation to "peruser" to underscore that the participant in the blend is not just browsing the World Wide Web anymore, but is in fact carefully parsing the contents of the site in order to satisfy a particular goal (irrespective of whether that goal was established prior to sitting at the terminal or whether the browser maintains this goal in working memory for the duration of the ongoing activity). In the blend, but not in input space 1, the addressee is using the World Wide Web as a means of satisfying a goal specified in the Telos space. Thus, the emergent structure of the ongoing activity can be characterized as "local actions at the computer terminal (mouse clicks) have a causal effect in the world at large". In the blend, a subject, "the peruser," consciously *intends* to point her cursor and click her mouse button in order to accomplish a goal. An exception would be some kind of context in which a person inadvertently "goes" to a website, as illustrated by

(28) I inadvertently clicked onto the Ku Klux Klan's homepage.

The coding time for point and click is certainly within the three-second window of attention, and it stands to reason that individual instances of pointing and clicking take place in these 3- second intervals. This fact contributes to the overall entrenched feeling that the human-computer interface occurs in and over short time spans. Users can spend hours, even days browsing, but the feel of each instance of pointing and clicking are comparatively short. This is an important affective property of the situation that is often exploited rhetorically.

Now let us consider the other components, the adverbs *way* as an optional adjunct to verb phrase.

With this introduction of *way* (plus possessive pronoun), the Point & Click blend acquires an additional element: the simulation of movement from one place to another. As Goldberg (1995: 199-217)argues, the lexeme *way* usually encodes a path that the agent creates by using a non-stative verb (i.e., verb implying motion). But the verbs *point* and *click*, even as isolated components do not represent protypical instances of human motion (other than the motion of extending an arm or the motion involved in producing a sound, which is usually minimal) and even less by way of a body moving along a path it has created for itself. The mental simulation that runs in the blend, however, includes body movement along a path and that movement is further elaborated by manner of motion. In the mental simulation, the verb phrase "point and click" contributes manner of motion component, focusing attention on the minimal bodily actions of the simulated self, while the direct object *X's way* focuses attention on the means of motion from starting point to destination. The introduction of *way* entails a TRAJECTORY image schematic transformation in which we are to trace mentally a path from our present starting point to some imagined destination.

It is in specific instantiations of these two constructions that mental simulation further compresses relations of identity, time, cause-effect, change, and part-whole that do not typically fit within the usual scope of computer use (such as buying a car or home, refinancing your mortgage, conducting research, going grocery shopping, and so forth) into this mental simulation (more about this later). The combination of these two constructions generates an initial blend that compresses to an immediate, perceptible human scale a complex chain of cause and effect that without the aid of the computer and Internet would remain diffuse, complicated, and time consuming (for better or worse).

The last point structural property to consider before analyzing examples 26 and 27, is the presence of the adverb *just* and its cousin, *only* in the nominal construction. These adverbs function as *scalar operators*. A scalar operator is a grammatical expression with a pragmatic interpretation that "derives from the fact a scalar model is taken empirically to contain a set of propositions as part of the shared background of speaker and hearer at the time of utterance" (Kay 1997:53). In these, the implicit

scalar model is a shared measurement of distance such that the topic of conversation in the real world is viewed as entailing travel over a longer distance as compared to the same topic in the virtual world. Tasks in the virtual world that influence results in the real world are shorter. There is an implicit origin, a stated destination, and a span in between. The adverbs *just* and *only* cue interpreters to construe a contrasting distance, one shorter than the other. As we shall explore more elaborately in the next chapter, this scalar model can be linked to another scalar model in which shorter distances are preferable to longer distances, producing a "utopic" rhetorical effect; or contrastingly, this scalar model can be linked with the opposite evaluative scale in which longer distances are preferable to shorter distances, producing a "dystopic" rhetorical effect (see next chapter).

A short Lexis-Nexis Universe database(8/11/00) sample set of major Newspaper, Magazines, and Journals published in the last two years reveals a statistical preference for *Just a click away* over *Only a click away,* with a hundred-and-fifty matches for the first and thirty-three matches for the second. Some instances found in the corpora add a temporal dimension to the scalar model with the adverb *now*, as illustrated in

(29a) . . .now just a click away,

and

(29b) . . .now only a click away.

Another interesting finding is the presence of both adverbs, as illustrated in

(29c) . . .only just a click away,

evidence that *just a click away* functions in some speaker's minds as one whole unit. However, the syntactic placement of these adverbs is not reversible. Sentence 24d,

(29d) *just only a click away,

is unacceptable in my grammar and highly questionable generally, and no corpora searches produced any match, suggesting that it cannot function as a stable attractor for this scenario in the grammar of attention.

Given the character of this distribution, what might be the difference between *only* and *just?*

According to the Oxford English Dictionary, the first definition of adverbial *just* is that of "exactly, precisely," and "in the appropriate manner," whereas the first definition of adverbial *only* is a "single or solitary thing or fact". These adverbs focus attention on complementary aspects of the process associated with getting what you want: the adverb *just* focuses on the manner or duration of execution; the adverb *only* focuses on the number of executions required. *Just* construes perspective as internal to the event of clicking, a sound of short duration (the perceptual duration of a single clicking sound), whereas *only* construes an external perspective of iterative instances of clicking and the fact that only one instance is necessary as opposed to many. *Just* profiles short duration while *only* profiles the number of sequenced steps required of the computer user. We can stipulate that the exceptional instance of *Only just a click away* is a blending of these two alternative expressions in which the adverb *only* is added so that the entire construction, focusing attention simultaneously on the number and duration of executions, whereby emphasizing the singularity also emphasizes short duration. Number reinforces scope. Because the use of *just* in the phrase initial position is statistically more frequent than *only*, one can conclude that it is entrenched and easier to access..

The nominal *a click* summarizes the entire set of events associated with web browsing, construing input space 1 as a static, completed action rather than an ongoing presence. Following Langacker (1987, 1998) the clipped form (*a click* instead of *a point-and-click*), creates an active zone—picking out and profiling a part of the object, event, or action and using it to metonymically evoke the entire scene; it is a reference point for constructing a more elaborate simulation. Thus, the coordinated actions of pointing and clicking while sitting in front of a computer can be accessed easily by merely coding a small portion of that scene. The later clipping suggests the once novel scene of interacting with a computer using GUI's is now thoroughly assimilated and easily accessible; thus, the only salient part of the activity is the clicking motion, an act that "moves" *you* from one site to another. Notice, however, that speakers of

English typically do not use the verb *point* to clip the verbal or nominal instantiations of *point-and-click*; thus,

(26a) ?Grocery savings are just a point away,

would not be found. (The Lexis-Nexis corpus search revealed no instances *just a point away* as it relates to interacting with a graphic user interface.) The act of clicking entails the act of pointing, so we can say that the focus of attention is on the results of browsing rather than on the preliminary conditions for accessing a site.

The final adverb for sentence 24, *away*, also deserves commentary, for it reveals that we conceptualize this activity in terms of spatial distances. More specifically, we often conceptualize web browsing as going to a specific location (even though it is just as valid to construe the information on the site is coming to us as we are going to it). In contrast to *way* as verbal adjunct, the introduction of *away* as a nominal adjunct can be analyzed as a PATH-END FOCUS image schematic transformation.

Let us now focus deliberately on sentence 26. I encourage you to consult figure 3.5 in the appendix as you read.

I routinely receive an e-mail message from the corporation that owns the store where I do my weekly shopping. I am asked to print out the attached text (a list of grocery items titled "ValuPage") and then take it with me to the store. These conditions constitute the specific features of the Base space in which I am the "addressee" and I am engaging in the ongoing activity that constitutes a key representational component of the message. By virtue of bodily orientation, attention to matters of computer use are primed for activation because current practice makes them accessible to selective attention.

The input space for Web Browsing is identical to the one presented in Figure 3.4. The distinction is that the Telos space becomes the Grocery Shopping space. In this space, the participant role "shopper" (x'), location role "store" (z') and theme "groceries" (y) comprise the elements of the organizing frame for

a future, habitual activity. In addition, the element "coupons" (w) is also recruited from long-term memory as a salient element of Grocery shopping (more about this later). The blended space for Web Shopping, includes the participant role "peruser-shopper" (x") a fusion of the participant roles from inputs 1 and 2, a locative element, "virtual-store" (z"), and a symbolic artifact, "electronic-coupons" (w'). In the blend, the shopper inhabits a virtual store that, unlike its real counterpart, does not have tangible goods, but rather has lists and pictures of tangible goods. Additionally, the Web Shopper virtually "clips" coupons for items sold at the real store. The blended space, in effect, constitutes a planning space prior to the real activity, wherein the shopper acquires the means (coupons) to save money on items he will purchase. The metonymic relation between coupons and prices is compressed to a direct causal relation such that clicking a mouse on an item in the virtual store will reduce the actual price the shopper will pay at the real store. Cause-and-effect and time compress to a single manipulable scene; the addressee witnesses the complex downstream effects of his actions at the computer terminal.

This cause and effect relationship develops through completion of input space 2 with respect to the metonymic relation between coupons and the prices for goods. Figure 3.5 adds two additional "satellite" spaces (i.e., spaces that profile a facet of an input space for the purpose of adding new structure) "orbiting" input 2. The two spaces I call Cost α and Cost β, with the former representing the retail price of groceries and the latter representing the discount price of the same items (the dotted circle indicates that Cost α space is a logical entailment of the Cost β space).[19] In short, the noun *savings* operates a cue for constructing two pragmatic scales that feed into the blend via a metonymy connector. Thus, clicking the mouse saves money at the real store.

The presence of *just* and *away* complete this blend by metaphorically construing it as a shorter distance between two points. As noted above, the web browser clicks her mouse and subsequently sees a new site appear on the screen. This phenomenon has all the attributes of apparent motion; that is to say,

[19] In this case, the alpha and beta distinction does not reflect the order of presentation, but the reconstructed entailment that the retail price exists before the discount. It is the initial point on a scale from which the subsequent point descends, metaphorically speaking.

we seem to organize our experience of web browsing around image schemas of intentional motion. The conceived real world outcome associated with input space 2 (*grocery savings*) is now being construed metaphorically as a spatiotemporal destination, a desirable one at that. In the blend, the virtual-shopper is closer to the destination than someone who is at the store but who has not activated his electronic coupons. Figure 3.5 diagrams this spatiotemporal metaphor with an analogy connector running from the virtual-shopper to the large dot in the Away β space. The adverb *just* cues the addressee to focus on the end of the path near the desired destination as licensed by a DESIRES ARE DESTINATIONS metaphor. As with the Cost spaces, the Away β comes with an implicit comparison, the Away α, an imaginary shopper farther removed from the destination. It is implicit NEAR-FAR image schema that governs the metaphoric aspects of this slogan.

The *point and click* construction used in the Ditech.com advertisement, on the other hand, windows the medial point in the referent scene. Figure 3.6 (appendix) represents a summary view of the mental space network for example 27.

In contrast to (26), the *Ditech.com* slogan appears as a television commercial. The base conditions of computer use (as represented by the Web Browsing spaces in each figure) have to be mentally simulated, since the default situation is of a person watching television and not operating a computer, a significant variable with respect to the alerting and orienting, because the referent scene is offline. The advertisers are banking on the notion that simulations of web-browsing are so entrenched and easily accessible to the consumer's psyche that simply mentioning the verb *click* will be enough to elicit the mental simulation of web surfing for those who are watching television and not using their computers.

Before getting deeper into the specifics of this example, let us consider in greater detail the *way*-construction as an entrenched, easily accessible lexical construction. As discussed in detail by Goldberg (1995: 199-217), the *way*-construction comprises a subject, a nonstative verb, a possessive pronoun that co-references the subject, and an oblique phrase coding direction. Instances of this construction imply a

subject moving along a path designated by the prepositional phrase. I offer two of Goldberg's (1995) examples as an opening illustration:

(30) Frank dug his way out of prison.

and

(31) He belched his way out of the restaurant. [199]

In both cases, the speaker construes the subject as an agent moving along a delineated path. What is of interest is the choice of verbs. In both cases, the speaker uses a nonstative, action verb; however, that relationships between the kind of action expressed by the verb and the kind of action encoded by the construction differ greatly. In (30), the relationship between *dug* and the rest of the construction is clearly one of means: the act of digging enables Frank's motion. His digging is not simply a coextensive action. In (31), the relationship between *belch* and the rest of the construction is ambiguous. A case can be made that the peristaltic convulsions of the subject's alimentary canal is merely coextensive with his movements through space, thus eliciting a manner of motion interpretation. Goldberg's findings suggest, however, that the means interpretation is primary and that even in cases where the verb would likely elicit a manner interpretation, interpreters (and context) often supply accounts suggesting means interpretation ( i.e. that belching caused him to leave the restaurant). Means and manner blend together, where the manner verb is the attentionally salient causal element in the whole referent scene.

Since these constructions often take the imperative mood, it is possible not to specify the grammatical subject, especially since the possessive pronoun is coreferential. In these cases, point and click operates as a lexical construction that fits into the larger way construction, such that manner of motion and means of motion are compressed together, and where the task at hand is often metaphorically construed as motion.

Another feature of *way* important for this analysis is that it calls attention to the intentional nature of the activity, since the subject of this construction (as co-referenced by the possessive pronoun) is

understood as the "creator" of the path. She is not merely wandering from site to site. The cliche, *You make your (own) way in this world*, for instance, suggests that the subject is finding or creating her own pathway as she moves along it. Because the default semantics of verbs like *dig* do not include directed motion along a rectilinear path, the strong reading of intended motion comes from the construction itself. The use of this construction focuses and sustains attention on the actions of the agentive subject.[20]

Given these preliminary remarks, let us analyze this example using mental spaces. Input space 1, Web Browsing, is substantially the same as in figures 3.4 and 3.5, but with one difference: the contents are encircled with a thicker line. Input space 2, the Debt Management space, contributes the focal participant role "debtor". In the present context, the "debtor" owes money to a lending institution, such as a credit card company, and is paying a high rate of interest.

At this point, interpreting aspects of Figure 3.6 will clarify. Like the Grocery Shopping space for sentence 26, the Debt Management space has two satellite spaces orbiting it: Loan α and Loan β. It is the Loan β space that represents the initial state of the Debtor at time *t*. Thus, the initial pitch, *So you've gotten yourself into too much debt* casts the addressee in the participant role "debtor". (Remember that in the Base, the value of "debtor" is watching television.) The adverbial phrase, *too much debt*, draws the addressee's attention to the component of being in debt, which is paying interest on money borrowed, which presupposes that money paid on interest cannot be used to buy other goods and services, even

[20] It is possible to use a non-agentive subject in this construction, as in *The car plowed its way through the crowded streets*; however, the reason we can do so is through a rather direct metonymic connection to an animate agent acting with volition. In short, the presentative condition of the sign suggests that the mentioning of car entails a driver who, at the moment preceding coding time, was controlling the vehicle. Notice that the sentence, *The asparagus plowed its way through the crowded streets* requires that we either attribute animate qualities to the vegetable (as in a drug induced hallucination) or to a novel situation like a person wearing a large asparagus costume and marching in a parade. The sentence, *The asparagus stalks grew their way through the crowded streets*, is easier to interpret because growing is an easily accessed process associated with vegetables. In this case, the act of growing is being construed as horizontal motion along a path in which the normal temporal dimensions associated with plant growth and which typically are not observable without the help of time-lapse photography is construed as an event with a temporal dimension which is easily observable. It is worth noting, however, that the subject of this sentence is plural, not singular. The sentence *The asparagus grew its way through the crowded streets* would either be interpreted collectively (despite the grammatical cues indicating singularity) or would represent an exceptional, outlandish situation, such as a giant asparagus stalk taking over New York City.

necessary goods and services like food and shelter, or capital-producing goods and services like education, property, or stocks and bonds. Examine the lower right hand corner of Figure 3.6, in it you will find a schematic drawing representing the IN-OUT image-schematic transformation, the full significance of which will be discussed shortly. The first dot represents the "debtor" at this initial time. In summary, the Debt Management space at time (t) comes with a negative evaluation, meaning that the debtor is not properly managing his debt load, creating the need in the addressee-debtor for the service about to be advertised.

Assuming the addressee in Base space (television viewer) fits the profiled debtor role described above, and assuming he mentally simulates computer use, we can now posit the existence of a blended space with the salient participant role, "virtual-debtor," a person web browsing who fits the financial profile and matches the person in the Debt Management space. In the blend, the virtual debtor has access to a service that promises to substantially change his present financial situation for the better, so that merely clicking a mouse button will produce the desired effect of making that person debt free. In contrast to the person in the Debt Management space who is at the mercy of the lenders and who must use more of his income to service debt, the virtual debtor is in control and can create a "path" to financial independence. This empowerment scenario depends on the presence of *way* followed by a particle (*out*) and prepositional phrase (*of debt* ). Like *grocery savings* in (26), *debt* is a metaphorical place; unlike *grocery savings*, however, *debt* is not a desired destination. Rather, *debt* is implicitly conceptualized as points on a trajectory, becoming a region from which the debtor is striving to emerge. By clicking his mouse, the debtor creates a path out of this metaphoric region, which, in addition exploits a *center-periphery* image-schematic transformation, where center corresponds to the high-debt load at time (t) and where the periphery corresponds to financial independence at time (t+2). The service offered by Ditech.com promises to help the debtor move from center to periphery (time t+1), and from there, the addressee can infer that he is on his way to becoming debt free.

Like the previous example, the "virtual-debtor" in the blend focuses attention on the causal relation between visiting a web site and becoming debt free. His attention is so focused because the blend builds a scenario that compresses relations of cause-effect, space, and time to a singularity, allowing the addressee to envision a better financial future.

What is the difference between these two constructions?

As we have seen, the words point and click can be used as a complete functional unit or in clipped form as both verbs and nouns. The verbal uses often are embedded within the more common *way*-construction and always require a directional preposition; the nominal use often occurs as a functional unit consisting of the adverbs *just* or *only* (or both) preceding it and the adverb *away* following it. Both constructions highlight computer use and the downstream effects associated with it; both constructions do so by metaphorically construing the complex causal relationship between computer use and its effected or affected activities as motion along a path from one place to another. But the similarities end there.

The difference is in the degree of prominence afforded to the act of pointing and clicking the mouse and its relation to the discourse topic. The nominal *Just a click away* focuses attention on the endpoint; the thing you want exists in a fixed location that is very close to you. The mental simulation is construed statically as plotting a course form one fixed location to another and assessing its relative distance. Websites uniformly "reduce" or compress the conceived distance. The verbal *click your way*, on the other hand, focuses attention on the process of moving to the desired location. The mental simulation is construed dynamically so that the person doing the pointing and clicking is creating her own path to a desired site rather than traversing an already plotted trajectory. In fact, the number of pointing and clicking executions remains in the background. Satisfying a goal may require the user to execute several pointing-and-clicking routines; in short, the path might be long and meandering. Nevertheless, the user is simultaneously sitting still at her computer and moving through space.

Now that we have an idea of what the difference is, the next question to tackle is why the

difference?

The rhetorical difference may be strategic. The nominal use of *click* occurred in a context where the reader was already pointing and clicking his mouse. The active verb was already primed in the motor affordances of the medium itself. The focus of attention can shift from an ego-centered perspective to an object-centered perspective because the ego was already doing what the advertisement wanted him to be doing. The verbal use of *click* occurred in a context where the television viewer was not already pointing and clicking, thus the notion was not already active in the motor affordances of the medium. The basic simulation needs external prompting, and verbal forms are critical for making salient that which is barely active. While all this seems quite plausible to me, I still cannot rule out the larger contributing factor which is the subject matter itself. Debt consolidation is a long, complex process and the emotional valences associated with it run along a continuum of fear, frustration, and helplessness. Therefore, creating a referent scene whereby the person experiencing this kind of debt can actively "make his way" to a new financial destiny is an important precondition for seeking this service. In this case, an ego-centered perspective vis-à-vis the Internet matches the ethical and emotional dimensions of the situation better than an object-centered perspective. In fact, it may be the case that the viewer accumulated his debt through Internet use itself. In this case, it would be best for the advertisers not to construe their product as coming to them, after all they have been seduced by easy purchases.

*Blowing and tooting your own horn*

Coulson (2001: 93-96) has analyzed the idiom *Digging your own grave* as more than a straightforward metaphoric mapping between the source domain of "grave digging" and the target domain of "trouble", because the default interpretations of this idiom entail the inference that the deeper one digs, the closer one gets to dying. The implicit causal relationship between grave digging and death is clearly counterfactual, for by all accounts grave digging and death are not causally related (unless when applied

to atypical contexts in which a sexton suffers a massive heart attack while digging a grave). The idiom is also fantastic in that the agent and patient roles fuse in the blend; that is, the frame slot for "grave digger" and "body" are filled by the same referent. This idiom can be applied to an indefinite set of representations of an individual's actions that are thought to have untoward results, and which the speaker thinks the individual does not recognize as harmful. Example target spaces can include anything from financial ruin, academic failure, or even injury or death, such as:

(32) You're digging your own financial grave by investing all your money in start-up Internet stocks,

or,

(33) You're digging your own grave if you don't study for next week's Chemistry final.

I will examine in greater detail a close relative to this idiom: *Blowing [tooting] your own horn*, an idiom with similar patterns of linguistic productivity in that it takes part of the larger construction type X your own Y, an with examples 31-33.

(34) Get your own drink.

(35) Take your own notes

(36) Play your own instrument.

Like *Digging your own grave*, this idiom is a metaphoric blend with fused counterpart roles. First I will provide a basic blending analysis of the idiom and then account for its use in two texts (both gathered from the Lexis/Nexis database).

My basic analysis poses three mental spaces, two inputs and a blend, in which the first input functions as a source space for the second.

The verb *blow* (or *toot*) and the noun *horn* provide the verbal cues for opening a mental space representing the act of playing a musical instrument. More particularly, it invites us to simulate playing a brass or wind instrument in particular, rather than a string or percussion instrument. In this mental space,

the Horn Blowing space, the focal element "horn" provides the reference point for accessing and filling other slots in the frame, such as "musician" and "musical score". There is a part-whole relation between the musician and her instrument encoded by the adverb *own*. In this case, "own" represents an intentional agent's possession and manipulation of an object (patient) such that to own it is to use it. The event represented in this space is the effects associated with producing a certain kind a sound; that is, the musician blows into the horn, which produces a distinctive (and loud) sound competing for the attention of other "auditors". The "auditors" notice the musician because she produces a sound that stands out against the background of other sounds. In terms of a grammar of attention, the sound of the horn is conceived as filling and dominating the immediate semiotic landscape.

Just as there is a close part-whole relation between a musician and her instrument, there is an even closer part-whole relation between a speaker and her voice. This common part-whole "topology" establishes a close relationship between the source representation of "horn blowing" and its target representation of "praising" in the Encomium space.[21] Since native speakers of English know that this idiom is about acts of praising, the mental space activates the focal element "speaker" and "voice". Once activated, the speaker role and her most relevant feature map onto the focal elements in the Horn Blowing space, namely the "musician" and "horn". But a salient element of the Encomium space is also covertly active, since without it the idiomatic interpretation of these words would not emerge. That element is the "subject". In this context, the subject is the accomplishments, character, and deeds of a living person. In the Encomium space, the speaker uses language to single-out for special attention aspects of her subject such that auditors focus their attention on the subject's life. We can presume that were it not for the efforts of the speaker, the subject's life would remain anonymous, never dominating the immediate semiotic landscape. Connecting these two spaces is the commonplace notion that distinctive sounds produced by a human being will claim the attention of other minds. The "subject" role in the Encomium space can

[21] *Encomium* refers to a formal genre of speech in which a speaker or writer praises the achievements and deeds of a living person.

activate an element like “music” or “musical score” in the Horn Blowing space, where the sounds conspire to produce music and song. Music/song then maps onto the topic in the Encomium space. The topic of the Horn Blowing space is music; the topic of the Encomium space is a living person's biography. Before discussing the blend and why we need to posit its existence, I want to call your attention to the mismatches between these two spaces. The number of focal participants in the source space is 1, the musician, whereas the number of focal participants in the Encomium space is 2, the speaker” and her “subject”. In the first input space, the musician is the “agent” and the “horn” is the “patient,” while in the second input space the speaker is the “agent” and the subject is the “patient”.

Since the source space represents an agent manipulating an object and since the target space represents a different agent voicing her opinion about another person in a public forum, we need to posit a blended mental space that accounts for the asymmetry in the number of participants. In fact, this asymmetry is the source of emergent structure in the blend: self-aggrandizement or bragging. The organizing frame for the blend is an agent working on an instrument, but the topic is issuing verbal acclaim. This incompatibility leads to the central inference that the speaker in the blend is bragging about herself. In the blend, but not in the target space, “the speaker is the subject of the encomium”; and in the blend, but not in the source space, “the musician is the speaker”. Therefore, “speaking is blowing a horn” and “blowing a horn” is the conceptual equivalent of “bragging about one's accomplishments”. I therefore will refer to the blend as the Bragging space. The blend connects to the cultural schema associated with formal speech acts of praise. One salient rule of etiquette associated with encomia is that one should not praise one's self, because there is a perceived conflict of interest between speaker and subject when that speaker's life IS the subject. He or she appears to have a hidden agenda, or the act implies something venal about the speaker and by extension the subject. (As we shall see, however, this idiom is often used to describe situations when it is deemed necessary to break this fundamental rule of etiquette.)

But my account has left a crucial element out. Why the instrument *horn*? Why not a string or

percussion instrument? They, too, produce distinctive sounds that stand out and potentially dominate the semiotic landscape. Two reasons come readily to mind. The first is, of course, that these instruments need human lungs to produce their sounds, just as the prototype instance of praising someone else requires a human voice powered by the same vocal articulatory tract. But that would not explain why the noun *trumpet* can substitute for *horn*, as a recent Lexis/Nexis search shows, but other instruments such as *trombone, tuba, clarinet* or *oboe*[22], cannot. This referential asymmetry points to a second reason. The western cultural schema for praising someone often comes with a rich set of assumptions about the formality of a situation: these acts take place in ritualized ceremonies. Musicians and their instruments are included as a salient part of these ceremonies, and they are usually playing brass instruments. In these ceremonies, horn playing is often a scripted part of the procession, hailing the attention of the crowed. When playing begins, the arrival of a very important person is announced such that the attention of the crowed is directed on that person in question. Given that ceremonial role, the horn becomes a reference point (an alerting and orienting device) for the entire public act of praising an individual, and it would be bad form indeed for the VIP to announce his own arrival. Thus, the choice of instrument has much to do with accessing the ceremonial nature of the speech act as it does with providing the metaphoric basis of the conceit itself (see appendix).

Now let us consider two specific examples. The first instance of this idiom appeared in a September 22, 1998 story in the metropolitan section of the *New York Times* on President Clinton's visit to New York University and quotes its president, L. Jay Olivia, as saying:

(37) "I firmly believe that if you're doing something interesting, you ought to tell people about it," Dr. Olivia said. "And if you're blowing your own horn, do it loudly. There's no sense giving it a little toot".

[22] It is rather easy to imagine contexts in which specific instrument names like oboe, clarinet, and bassoon can be used. One can even imagine situations which call for the names of percussion and string instruments as well. However, only horn and trumpet appear as the instruments of choice in a variety of contexts, such as "Lawyers like to blow their own trumpets" or "Nick Smurthwaite [special-effects artist]. . .doesn't believe in blowing his own trumpet Hollywood-style" [Lexis/Nexis search 7/31/00].

The second instance of this idiom appeared in an October 10, 1999 story in the *Chicago Sun-Times* on job interviewing , where the writer, James E. Challenger, gives this advice:

(38) "If you have a hard time 'tooting your own horn,' step outside yourself and view yourself as someone else might. Rather than thinking that you are talking about yourself, pretend you are recommending a good friend".

I touch briefly on some issues here, but reserve extensive comment for the next chapter. The first thing to notice about the first example is the self-conscious distinction the speaker makes between "blowing your own horn" and "tooting your own horn". After asserting the need for self-promotion (the topic of both stories) by invoking the blowing your own horn idiom, Olivia sustains attention on this topic by further focusing on the manner of self-promotion. Thus, the Bragging blend gets elaborated by recruiting additional representations of manner of articulation from the source *space*--blowing the horn is more likely to be heard than *tooting* it. What is more, the addition of a new referential space, New York University, fills out the "subject" role in the target input and blend. Attributes from the NYU space capable of being projected into the blend can includes such designations as "multicultural", "cosmopolitan", or "a world compressed into a village". In doing so, the completed blend allows for an institution to be praised rather than an individual; however, such extensions of the genre are typical given that institutions are conglomerates of persons. Critical to this simulation of self-promotion is the exploitation of our mental models of loudness and softness, suggesting that the latter, when applied to the bragging scenario, would be ineffective in the "intensely competitive world of higher education". The adjective *competitive* should be read in terms of rising background noise and the challenge of having the sound of your horn heard among a cacophony of "horns".

The second example also makes use of the tooting version of this idiom to justify self-promotion, again understood as a violation of etiquette. In this instance, the addressee, the "you" who presents a prospective employee with a resume and letters of reference, is represented as reluctant to engage in self-

promotion, thinking that these documents should be sufficient. The writer then offers advice for "tooting your own horn" by suggesting that you act as though you are someone else. While the idiom itself draws attention to the mismatch between the single participant schema of the source space and the dual participant schema of the source space. This instance creates a counterfactual pretence version of this blend representing the momentary consciousness of the addressee in mid-interview. The shy addressee, in effect, has to construct a pre-blend and a final blend. In the pre-blend, the person compresses his identity so as to conceive of the situation as calling for self-aggrandizement. Since this compressed situation produces discomfort, he is asked to construct a final blend that preserves the intentional structure of the pre-blend but decompresses the identity relations initially compressed .The pretense blend offers an illusion that the acceptable dual participant schema is being preserved during the interview, thereby preserving the etiquette of acclaim so internalized in the shy addressee's mind

As a preview of the next section, consider the adjective *own* as a component of a possessive phrase that functions reflexively. By reflexive I mean that the coded agent or experiencer arguments apply to the grammatical subject while the patient argument applies to the grammatical predicate (thereby selecting and sustaining attention on different facets of the same referent), as sentences 39 and 40 illustrate:

(39) Blow your horn.

(40) Blow your own horn.

The first opens a mental space for Horn Blowing in which an agent, coded through the possessive pronoun, does something to a patient. This construal is not reflexive because the agent (presumably the addressee) and patient arguments (the instrument) are clearly delineated. The second opens a similar mental space for Horn Blowing but profiles the relations differently. Two semantic properties come to mind, suggesting that this second utterance combines two of the prototype meanings of the reflexive outlined in van Hoek (1997: 172-178): the *contrastive* and *semi-subjective* meanings.

The contrastive meaning emphasizes a key difference between the designated entity and some other (often unspecified) entity or set of possible entities, whereas the semi-subjective meaning can represent canonical instances of reflexivity in which the participant referenced by the reflexive is viewed by the agent not objectively as a different person but semi-subjectively as the recipient of the action being performed, thereby fusing the roles of "viewer" and "viewed". The viewer in the enunciation space and the object of his enunciation take the same perspective.

The second clause exhibits both meanings in that the adjective emphasizes an implicit contrast, "your horn" and not "someone else's horn". Construed literally, this clause could represent a fellow musician refusing to let a fellow musician play his instrument, calling attention to topic of "ownership". Construed figuratively by mapping "horn" from the Musician space metonymically onto "speech" in the Encomium space, this clause exhibits the canonical semi-subjective meaning in which the agent and patient/theme roles fuse, calling attention to the "speaker-topic" relationship. The referent scene can be interpreted such that the person holding the horn is the person delivering the speech. This "onstage" perspective then does not fit the typical cultural schema of praising a living person.

*Reflexives*

Review of the literature on reflexive pronouns reveals three uses of reflexives: 1) as emphatic markers; 2) as point of view indicators; and 3) as coreference markers. In this section, I will consider each as they apply to these examples.

(41) I bought a Honda myself [spoken by a salesperson at a car dealership]

(42) The four surviving U.S. presidents toured Madame Tussaud's Wax Museum and Gerald Ford tripped and fell on himself. [inspired by van Hoek (1997: 184)]

(43) Travis Fryman is trying to play himself back into shape. [spoken by a television announcer for the Cleveland Indians in August 1999 during a late season game]

(44) Awakened by the sound of the building shaking itself to pieces, the family of five ran apartment to apartment, ringing doorbells and pounding on doors. [from the November 12, 1999 edition of the *Washington Post*]

In (41), the reflexive functions as an emphatic marker. More specifically, the speaker, presumably speaking to a buyer, profiles herself as a member of a class of buyers, thus she switches her status within the cultural schema of "commercial transaction". By calling attention to her previous role as "buyer," the speaker shrinks the social distance between her and the addressee by identifying with the same role, thus mitigating (if only momentarily) the potentially adversarial relationship between buyer and seller.

Example (42), analyzed by van Hoek, makes use of the reflexive pronoun to code point of view, such that the real person, Gerald Ford, falls on the wax statue bearing his likeness. It does not mean the reverse, i.e., that the wax statue falls on Gerald Ford. Reflexives focus attention on subjectivity and intentionality. In this case, *himself* does not corefer to an intentional, volitional subject who momentarily loses control of his actions (prompted by the verb *trip*), but to an object bearing his likeness. In (42), the point of view shifts to the "onstage" scene of Ford's tripping and falling, as opposed to the point of view remaining "offstage," as would be that case with

(42a) The four surviving U.S. presidents toured Madame Tussaud's Wax Museum and Gerald Ford fell on his statue.

Sentence 43 will receive extended commentary.

The context of this comment surrounds the fact that Travis Fryman, the third baseman and batter for the Cleveland Indians, was placed on the disabled list for several weeks due to a shoulder injury, and Fryman had just returned from "rehab" from a minor league team (what baseball officials call a "farm" team).

At the time the announcer makes his comment, Fryman is standing at home plate and practicing

his swing after his first strike. When the speaker uses the verb *trying*, he focuses attention on the subject's intentions. When he uses the non finite form of the verb *play*, he represents the kind of action. Notice that *play* takes on a narrower meaning in this context. It not only means "play" the game of baseball (i.e., batting, fielding, etc.) but it means playing major league as opposed to minor league baseball, and the prepositional phrase *back into shape* represents a pre-existing physical condition which the present Travis Fryman does not currently enjoy. The use of the reflexive just after the intransitive verb *play* sustains attention on the subject, but this time by splitting him up into an agent and patient counterparts, in which the agent corresponds to an intentional subject who remembers a previous level of competence, and the patient correspond to the recalcitrant present self (or body) that does not manifest that level of competence, but is striving to regain that level of competence.

A mental space analysis of this sentence goes as follows. Figure 3.8 provides a pictorial view of the frame compatible blending network for (43). The base space includes the focal participants "the game announcer" and the subject, "Travis Fryman". It is in this space that the meaning of *in shape* gains specificity as "the ability to play competitive professional baseball, such that the player can get hits, score runs, play good defense (3$^{rd}$ base) and help his team win. Temporally, this space corresponds to the moment when the subject emerges from the dugout and approaches the batter's box. On viewing his emergence from the dugout, the announcer makes the subject the topic of conversation, thereby selecting the viewer's attention on his game.

Since all the participants share the same background knowledge (i.e., they all know they are playing or watching a major league baseball game), the next mental space is the Game space. Temporally identical to the base, the Game space corresponds specifically to the subject's present activity, standing at home plate trying to hit the ball. In this space, Fryman's body and its present state is the focus of attention. He is presently unable to perform at the level of competitiveness he would like, because his body is not in pique condition. In this space, Fryman is unable to consistently put the ball in play and

score runs for his team. The progressive verb *trying* opens an Intention space that, in this network is the primary focus of attention. Grammatically, the introduction of the subject selects viewer's attention, while the predicate *trying to play* sustains attention on the selected entity, providing crucial information about his intentions, which is to put the ball in play and score runs for his team. If the Game space represents Travis Fryman's "body," this mental space represents his "mind". The Base, Game, and Intention spaces serve as inputs for the initial blend, the Getting in Shape space, the space housing the referent of *himself*.

This blend preserves the organizing frame in which Fryman is standing at the plate trying swinging at pitches (i.e., preserving topology permits sustained attention on the subject), but not hitting them. Temporally identical to Base, Game, and Intention spaces, the Getting in Shape space represents the present body of Travis Fryman as 'out of alignment' with the present intentions: body and mind are "out of synch". However, in this space, the act of playing pro baseball itself will eventually bring about the realignment of body and mind in the subject. The blend compresses cause and effect, such that "playing the game of baseball will enable Fryman to play the game of baseball".

The contrasting conceptions of the self (body) as unfit and fit for the game of baseball (i.e. shape) takes shape in the final blend, the Return space. The scenario evoked in the Return space is based on a common inference that acquiring something depends on your location relative to whatever you are trying to possess. In the blend, Fryman is attempting to go from out-of-shape to in-shape by playing the game of baseball. Fryman's participation in the game becomes forward motion along the path, even though the game itself does not have these attributes. That is, with each game, with each at bat, and with each inning, Fryman moves closer to his goal. In this space, the initial conception of using baseball as a means of getting into shape, is structured image schematically using the TRAJECTORY transformation, grammatically coded by the prepositions *back* and *into*. But a peculiar conceptual structure emerges as a result of exploiting the trajectory schema:, Fryman is traveling backwards in space even though time continues to progress. He is, in effect, trying to occupy a location previously occupied. This scenario

feeds into a new mental space that is temporally future. In the Future space, Fryman is once again playing in peak condition, getting hits and scoring runs for his team. The blended space creates a conceptual structure that allows us to render the Future (t+1) and Past (t-1) spaces a conceptually equivalent structure, by creating a situation in which the present Travis Fryman plays professional baseball, and in playing it moves along a trajectory that returns him to a previous location, the metaphoric source domain for conceptualizing physical and mental conditioning.

Grammatically, the reflexive pronoun places onstage three versions of Travis Fryman: his distinguished past, his undistinguished present, and his brilliant future.

The use of the reflexive also influences conceptual point of view. The announcer uses the reflexive that, in effect, zooms in and aligns his perspective with that of Travis Fryman. The viewpoint shifts from an Olympic one to one in which we get inside Fryman's head and body. The announcer could have said

(43b) Travis Fryman is trying to get back into shape by playing the game.

Similar to the actual utterance, the speaker asserts something about Fryman's intentions and actions, but the point of view is offstage rather than on stage. The speaker, and hence the viewer regards Fryman objectively rather than subjectively. Our alignment is with the speaker, whose position is of an olympian offstage narrator.[23] This does not prompt us to simulate as much the body movements of Travis Fryman, but it does preserve the causal and spatiotemporal blends, whereby the subject traverses a metaphoric path as he is swinging at pitches. In a grammar of attention, the difference is that the reflexive puts Fryman, the speaker, and the viewer on the same stage, thereby explicitly marking for attention Fryman's own body movements. (See figure 3.8 in appendix.)

Example 44 uses the neuter reflexive pronoun as a coreference to *building*. Unlike most reflexives representing intentional acts (sometimes with unintended consequences, as in sentence 3), this

[23] The terms olympian and olympus come from Line Brandt's thesis "Explosive Blends: From Cognitive Semantics to Literary Analysis". Unpublished MA Thesis: Roskilde Universitetscenter, Denmark, 2000.

use of the reflexive does not code intentionality. In other words, there is no need to personify the building when interpreting this sentence. In fact, absent additional contextual guidance, both producer and interpreter will not personify the building. To what representational end, then, is the reflexive being used in this instance?

I begin answering this question by offering contrasting versions of this opening adverbial clause. Recall that the subordinate clause in 44 appears thus:

(44) Awakened by the sound of the building shaking itself to pieces. . .

Now consider the same phrase minus the reflexive pronoun:

(44a) Awakened by the sound of the building shaking to pieces. . .

Or, consider a passive version:

(44b) Awakened by the sound of the building being shaken to pieces

Both alternatives eliminate instruction for placing the family of five *inside* the crumbing building. Without additional information, it is possible to infer that the family of five live in an adjacent building, were awakened, and ran into the crumbling building in an heroic attempt to save lives. The addition of the reflexive orients the readers attention by narrowing the onstage region being represented, while the two alternatives potentially expand the onstage region to include adjacent buildings and locations.

*Prefixes as reference point constructions*

One of the most productive units in the English language is the morpheme *pre-*. Not only is this form prefixed to any verb of Latin origin, it is now attachable to nearly any lexical item. According to the second edition of the OED, pre- has replaced *ante* as the opposite of *post*, thus speakers now prefer *pre-Darwinian* over *ante-Darwinian*, *pre-Columbian* over *ante-Columbian* and so on. The items I will consider in this section include the verbs (and their gerund counterparts) *pre-drill*, *pre-prime, pre-install, pre-write,* and *pre-heat*

In most contexts these terms refer to the actions that should logically follow them; thus, to pre-drill something is not to prepare the drill or the surface for drilling, but actually to drill a hole; to preinstall something is not to create the conditions for installation but to actually install it; to pre-write is not to sharpen a pencil or clear your text editor, but to actually write something down; to preheat an oven is not to do something to an oven before it is turned on (e.g., clean it), but to start heating the oven; and to pre-prime a surface (to prime a surface is to apply a base coat of paint) is not to sand the surface smooth or remove old paint, but to actually apply a base coat of paint.

What is happening here? Many of my colleagues find these term barbarous because they are "illogical". How can you "pre-write" when you are already moving pen across paper (Zeugma)? How can you pre-drill when you are already creating the hole? How can you preinstall a program if you are already loading it onto your hard drive? How can you pre-prime clapboards if you are already painting them? The standard view of morphology is that prefixes alter the meaning of the lexemes to which they affix It seems, however, that the prefix *pre* has not been fully grammaticalized and has a lexical status of its own. It behaves like a morpheme, a unit of form affixed to a lexical item but which itself represents temporal precedence rather than the lexical item. Hudson (2000) for instance has pointed out that English morphology, such as word-final *-s* acts as a marker of plurality when applied to nouns and as an agreement marker governed by third-person singular noun subjects when applied to verbs. This fact leads him to conclude that plurality is a property of the lexeme not the morpheme.

According to Langacker (1999:173), an indispensable and pervading feature of quotidian experiences is the ability to "invoke the conception of one entity for purposes of establishing *mental contact* with another" (173, italics in original). Thus, consistent with the semiotic notion of indexing, reference points form the backbone meaning making in language, perception, and thought. For the most part, this ability lies below the threshold of explicit attention, but we use reference points all the time, as suggested in my discussions of walking to the library to borrow a book. If reference points are

fundamental elements of the semiotic function of language, then attention is too another fundamental cognitive function of language structure and use, and since, reference points occur via a conceptualizer, target, and context, a formal theory of context is also necessary.

In this section, I argue that the notion of cognitive reference points and accessibility are critical for understanding the discourse semantics of *pre-*

This sample text appeared in the *Boston Globe* in a question and answer forum for do-it-yourself handymen. It is a reply to a question asking how to repair loose screws in a door hinge that will neither tighten nor back out of the wood.[24] The columnist's reply includes the verb *pre-drill*. He writes:

(45) . . .You'll have to back out the screws that are tight. Once those screws are removed, pry up the hinge leaf, which will pull out those screws that cannot be backed out. To make the holes tight so the screws will hold, your best bet is to drill a large hole (⅜ to ½ inch) at least two inches deep, fill the hole with waterproof glue (resorcinol is a good type) and insert a wood dowel to fit the hole snugly. Let the glue dry overnight. *Pre-drill* pilot holes in the dowels and redrive the screws holding the hinge leaf. . .[My italics]

Base space is the writer and reader; the writer is an experienced handyman giving advice to the reader who is a less experienced handyman. It is important to keep this mental space in mind, because it establishes the rhetorical conditions governing the exchange. The base is the viewpoint, but the space in focus covers the sequence of steps for repairing the door hinge. The most salient participant is the antecedent of "your" who is also the implicit subject of the imperative clauses, both of whom map onto the unnamed addressee (and general reader) in the base space via an identity connector. Once in focus the embedded Problem space opens a series of subordinate or (step) spaces that "explode" the summary

[24] The precise wording of the question reads as follows:

> Two of the hinges on my front door have loose screws, but they won't back out. I know the holes are two big for the screws, so they are not holding the hinge, but how do you get the screws out? Two of the four screws in each hinge are the culprits. The other two screws in each hinge are tight. How do you make the loose screws tight again?

nature of the represented task into selected parts. The purpose of the discourse is to highlight those components and arrange them sequentially. The step spaces include a mental space representing the act of removing the tight screws, a mental space representing the act of prying the hinge plates loose along with the loose screws, a mental space representing that act of drilling larger holes in the same spots as the holes for the loose screws, a mental space representing the act of filling those holes with glue, a mental space for inserting wood dowels in those holes, a mental space representing glue drying, a mental space for drilling new holes in the dowels, and finally a mental space representing attaching the hinge leaves to the door or door jam. It is these last two mental spaces that receive prominence in the discourse. Notice that the motor image associated with "drilling holes in wood" is the same for the verb stem as it is for the more complex prefixed form. In other words, pre-drilling holes does not bring into existence "pre-holes," but, rather, holes. The difference is the selection of attention on specific elements in these mental spaces. The manner in which the representation of drilling these new holes in the door is "prospective" such that the represented act gives mental access to the next step, distinguishing the entire step as preparatory for the next step, calling attention to the status of these holes as 'guides' for something else, hence the appearance of the nominal compound *pilot holes*, where *pilot* accesses a functional schema in which something is guiding something else, and *hole* accesses a basic path schema for realizing this function. The coordinating conjunction *and* opens the next mental space, shifting attention to the next step in the process, which is viewed as the desired result.

The use of the prefix, then does not appear to alter the semantics of the verb, nor does it alter the larger representation in that mental space; rather, it references or helps the reader shift attention to the next step to create an integrated sequential package. In short, the mental space configuration is such that the Step 7 space functions as the viewpoint space for the subsequent structure emerging in the Step 8 space. By calling attention to a salient component in one mental space as preparatory for the salient component in a yet to be established mental space, the speaker uses a linguistic strategy that is likely to

keep the interdependent nature of these two steps in the forefront of the reader's mind.

This is manifestly not the case with the initial use of the verb *drill* in the previous sentence *you're best bet is to drill a larger hole*, which discourse participants are likely to view as a discrete step whose access path comes directly from the Repair space, since initial non-finite clause essentially paraphrases the principal goal of the repair. The motor imagery of drilling a hole in wood is essentially the same (though the resulting hole is bigger), but the representation focuses attention on the step itself and not on the relation of that step to the others. (See figure 3.9 in the appendix.)

*The rhetoric of inaccuracy*

It is possible to use a relatively fixed repertoire of signs in alternative ways. As Langacker notes, "the symbolic resources of a single language provide an enormous range of options for construing any given situation, and a speaker shows great dexterity in shifting from one to the other" (1998:5)

As an example of this symbolic dexterity, consider in detail the language of Boswell's biography, the object of my trek that started this whole mess. Example 46 is Mrs. Hester Lynch Piozzi's (formerly Mrs. Thrale) account of an incident involving Hannah More and Samuel Johnson (Quoted by Boswell):

(46) *That natural roughness of his manner so often mentioned, would, not withstanding the regularity of his notions, burst through them all from time to time; and he once bade a very celebrated lady, who praised him with too much zeal perhaps, or perhaps too strong an emphasis, (which always offended him) consider what her flattery was worth, before she choaked* [sic] him *with it.* (1966 [1799]: 1327-1328, italics in original)[25]

[25] The full passage from Mrs. Piozzi's *anecdotes* (1974 [1786)] reads:

> That natural roughness of his manner, so often mentioned, would, notwithstanding the regularity of his notions, burst through them all from time to time; and he once bade a very celebrated lady, who praised him with too much zeal perhaps, or perhaps too strong an emphasis (which always offended him), 'consider what her flattery was worth before she choacked *him* with it.' A few more winters passed in the talking world shewed him the value of that friend's commendations however; and he was very sorry for the disgusting speech he made her. (122, author's italics)

Boswell contrasts Mrs. Piozzi's anecdote with Edmund Malone's account of the same event:

(47) Now let the genuine anecdote be contrasted with this. At Sir Joshua Reynold's one evening, she met Dr. Johnson. She very soon began to pay her court to him in the most fulsome strain. "Spare me, I beseech you, dear Madam," was his reply. She still *laid it on*. "Pray, Madam, let us have no more of this;" he rejoined. Not paying any attention to these warnings, she continued still her eulogy. At length, provoked by this indelicate and *vain* obstruction of compliment, he exclaimed, "Dearest lady, consider with yourself what your flattery is worth, before you bestow it so freely".

How different does this story appear, when accompanied with all these circumstances which really belong to it, but which Mrs. Thrale either did not know, or has suppressed. (1966: 1328, italics in original)

These two passages occur in a section of the biography in which Mrs. Thrale's inaccuracies are contrasted with supposedly more accurate testimony in order to "guard my readers against the mistaken notion of Dr. Johnson's character, which this lady's *Anecdotes* of him suggest. . ." (1327).

Consider these passages from a cognitive linguistic perspective. From a grammatical standpoint, spoken and written discourse are assemblages of symbolic structures, comprising a number of constructions in which component structures blend together, both semantically and phonologically, to form composite structures. In elaborate texts, as well as ongoing discourse, the composites, in turn, combine to form ever more complex assemblies. In this view, constituencies do not represent a separate and purely grammatical dimension of linguistic representation (as mainstream generative research

---

Note that Boswell does not include the final sentence that asserts Johnson thought his speech "disgusting," and does not include the comma between "manner" and "so" that brackets "so often mentioned" from the grammatical subject, nor does Malone include quotation marks around the clause beginning "consider what her flattery is worth. . ".. The above text is printed from the British Museum copy of the first edition and collocated with the fourth edition. The editor's list of principal textual variants indicates no changes to this passage from the first to the fourth edition.

suggests) they simply reflect an order in which simpler assemblies (resembling the initial conditions or ’units' of linguistic symbolization) form progressively elaborate assemblies (approaching the boundary conditions of ’usage events').

Mrs. Piozzi's text exhibits complex composite structures. A text linguist would likely note that the first complete sentence is a complex assembly, the first component, *[[that] [[natural] [roughness]]]*, a clause functioning as a grammatical subject (i.e. nominal periphrasis), combines with two other discernable components *[[of] [[his] [manners]]]*, and *[[[so] [often]] [mentioned]]*, each contributing to a composite portrait of Johnson's demeanor verified repeatedly by others besides the speaker herself. The whole usage-event of discerning his character is treated as a reified unit, as if one could point to it the way one could point to an object in space. In the grammar of attention, the focus of attention is on the nominal phrase *that natural roughness,* with the prepositional phrase, *of his manners*, lending background support . Why? My argument assumes that part of the presupposed inferences held in common among writer and reader is that everyone has a discernable “manner”. While Dr. Johnson is not the only one with manners (read “demeanor”), it is his specific, idiosyncratic manner that interests the reader. Hence, the nominal phrase *natural roughness* characterizes or elaborates on the head of the prepositional phrase, and together the two components focus attention on a purported unique attribute of Johnson's manner. The component *so often mentioned*, in which “so often” functions as a single grammatical unit--with “so” indicating intensity of the speaker's commitment to proposition and “often” indicating the frequency of testimony--combines with the lexical semantics of the verb “mention”. These structures converge to produce the impression of an incontrovertible fact regarding a dimension of Johnson's personality. The structure of Piozzi's sentence leaves out a filler for the agent role, implying that anyone who knew him would make mention of it.

The lexical connotations of “natural” further suggests that this roughness is beyond Johnson's own control, the overriding rhetorical thrust of the entire sentence.

Initiated by the modal verb *would*, the final clause, *burst through from time to time*, and its parenthetically embedded adverbial phrase, *not withstanding the regularity of his notions,* expresses a struggle between his natural and studied demeanor. The elaborating clause calls for a force dynamic reading. In order to understand the semantic and pragmatic structure of the final clauses, we need an approach that captures how language represents the force dynamic interactions in a referent scene.

One basic cognitive category governing linguistic structure is force dynamic opposition. First formulated by Leonard Talmy (1988), the force dynamics model accounts for such stereotype conceptualizations as the exertion of force, the resistance to the exertion of force, overcoming the resistance to force, blockage of a force, and a removal of said blockage. These force dynamic conceptualizations have motivated the creation of symbolic repertoires at all levels of language structure, particularly in verbs and modals. "Overall," writes Talmy,

> force dynamics thus emerges as a fundamental notional system that structures conceptual material pertaining to force interaction in a common way across a linguistic range: the physical, psychological, social, inferential, discourse, and mental model domains of reference and conception. (1988: 50)

The basic structure of the model is simple: it consists of two focal participants, *Agonist* and *Antagonist* (one of which is always explicitly coded, while explicit mention of the other is optional). The agonist is said to manifest a force tendency, either toward motion or stasis, while the antagonist can manifest a tendency to stop motion, hinder motion, cause motion, enable motion, or alter the direction of motion. It is important to note that the antagonist responds to the agonist. A focal participant is construed as being either stronger or weaker than the other, and this strong-weak ratio determines the force dynamic result, either toward stasis, motion, or change in the direction of motion. The complexity of this system requires that I pause to introduce a sample of the force dynamic category and its range of representational domains. I follow Talmy's procedure by first offering a force dynamically neutral expression for maximal

contrast.

In the physical domain, speakers of English can say *The skateboard was rolling along the driveway* to encode a force dynamically neutral event, or they can construe that same event force dynamically with *The skateboard keeps rolling along the driveway*. Similarly changing the preposition in the first expression to *up* subtly marks the contrasting force dynamic relationship between the skateboard and the driveway. In the physical/psychological domain, speakers of English can say *Peter doesn't go swimming,* or they can say *Peter can't go swimming* to suggest that some internal inhibition or external force prevents him from going swimming. In the purely psychological domain, speakers of English can say *Peter did not go swimming* to suggest that a single, integrated conscious agent subject made the decision, or they can say *Peter refrained from going swimming* to suggest that a divided, conflicted conscious agent made the decision. Interestingly, an elaborated clause, *Peter couldn't bring himself to go swimming* suggests the same divided, conflicted self, but this time the grammatical subject is coded as the weaker agonist role rather than as the stronger agonist in the previous example, because a stronger antagonist is coded in reflexive pronoun. In the social-psychological domain, speakers can say *Peter's got to go swimming* to indicate an obligation while at the same time leaving it ambiguous as to the source of that obligation, whereas they can express the same situation force dynamically with *Peter gets to go swimming*, implying an implicit, external antagonist granting Peter permission. In the first version, it is not clear that Peter wants to go swimming, but it is not clear either who or what is making him go swimming. In the second version, it is clear that Peter wants to go swimming, but cannot do so without permission.

Keeping force dynamic structures in mind, we can now resume analysis, first by noting some specific linguistic structures that guide meaning construction. In Mrs. Piozzi's account, linguists would peg *natural roughness* and *burst through from time to time* as force dynamic properties used to express the nature of Johnson's personality. The writer of the anecdotes construes Johnson's own comportment as

an intra-psychological struggle between two selves (see Lakoff 1996): rough and refined. This struggle is made evident by his inconsistent public behavior. The force dynamic reading of this passage would assign the natural roughness as the agonist whose tendency is toward motion (i.e., "burst through"), whereas the part of Johnson that tries to inhibit these outbursts, "the regularity of his notions," occupies the force dynamic role antagonist. As the phrase "not withstanding" suggests, the force dynamic outcome does not favor the antagonist. The verb "withstand" is itself force dynamically marked as an instance of sustained stasis against an antagonist, as in

(48) The shed withstood the gale-force winds.

The negation of the verb's semantic meaning indicates that the subject of withstand (construed as a antagonist) failed to inhibit the outburst. The figurative result is motion, but the literal result is insult. The reader's attention is fixated on Johnson's dual personality, an abstraction, by metaphorically fixating attention on the commonplace notion of a physical struggle among two unequal forces, yielding a specific physical result that is metaphorically applied to social interaction.

Construing the opposition of forces as entirely internal to Johnson facilitates another strategy, which is to suggest that the recipient of his outburst, Hannah More, did nothing significant to provoke Johnson's sharp reply. Thus, it is also noteworthy to focus on Mrs. Piozzi's strategic use of the adverb "perhaps," twice in this passage, first in the final syntactic position followed immediately after a coordinating conjunction in the fronted (or prejunct) position. What kind of work is this word doing?

This adverb functions as a hedge word that at once concedes the possibility that the celebrated lady was praising him and that he did not like it, but represents her offenses as too weak to count as a significant force, and, hence, a significant breach of social mores.

We should also note the presence of the verb "choked" ending this passage. The use of a transitive verb in which the young woman's words are the means of asphyxiating its recipient. In Piozzi's anecdote, the verbal presence produces an impression that Johnson has overreacted by issuing a visceral

image. In other words, Piozzi tries to focus attention on the inappropriateness of Johnson's response rather than on the act precipitating his response.

I now offer a detailed mental space analysis of both passages. Piozzi's initial sentence makes a general claim about Johnson's character that is subsequently supported in the following sentence. Interpreting each sentence entails composing two different mental space networks.

The opening assertion may be analyzed as an interconnected network of four spaces and a generic schema.

The first space in the network is the Dialogue space. In this mental space, the focal participant is "Johnson's self" (or body) and his verbal and gestural interactions with unspecified. From the context established in both Piozzi's and Boswell's books the focal topic is Johnson's demeanor in mixed company and how he deals with flattery from women. The mental space represents his bodily actions as they are reported to occur in such situations.

The second mental space is the Subject α space. In this mental space, the focal participant in this mental space is, again, "Dr. Johnson," but this time the issue is his personality or manner, what Lakoff (1996) distinguishes as the "subject". The relationship between this mental space and the Dialogue space is indexical. Observable speech and demeanor emanating from Johnson's body indicate or reveal his internal, unobservable characteristics or traits. In this case, they reveal a natural roughness of his personality, which, reflects in his somewhat grotesque appearance. The surface of the self matches the core of his personality.

The third mental space, the Subject β space, counteracts the second mental space. The focal participant in this mental space is, once again, "Dr. Johnson", but this time the representation is his artificial manner. This regularity, a tutored trait (like ape language) is a subjective personality trait created to "mask" his natural roughness. It is a conscious, deliberate manner. The surface of this self does not match the natural core of his personality, but it exists nonetheless. Thus, these three mental spaces operate

in a network that represents his self-subject conflicts as they play out in the social arena.

A schema for force dynamics governs these three mental spaces. The focal participants in the mental space are the “agonist” and “antagonist” roles and their various “force tendencies”. It is commonplace to induce this schema to represent social interaction generally; thus, the force dynamic model functions as an organizing frame for the network.

The fourth mental space, which I call the Occasional Outburst space, represents a force dynamic metaphor for intrapersonal conflict that has a manifest, public result. “Dr. Johnson’s subject” is the “Agonist-Rough Manner”. The purpose of the Subject β space is to represent the psychological act of keeping this rough manner from revealing itself. It is represented as the “antagonist” to the subject represented in the counterpart space. In moments of outburst, the “agonist” wins control over Johnson’s body, controlling his public manner. However, this control is intermittent and irregular. Roughness, then, is represented in two ways, as a core personality trait and as an imperfect ability to regulate or keep that roughness private. His self reveals this tension. (See figure 3.10 in the appendix.)

Interpreting the next sentence as an illustration of Johnson's personality can be captured in a four space analysis, where the first input space is the same Dialogue space described above, but now organized around the inference that the “natural rough” subject wins out over the “regular” subject. Johnson's body is currently being controlled by his subjectivity represented in the Subject α space.

The phrase “and he once bade a celebrated lady” builds a second mental space representing the context of Johnson's response, which I call the Praise spaces, with Johnson (Y) and Hannah More (X)as focal participants. The organizing frame of this space is that of participant X praising participant Y and Y responding to X; it also contributes the background knowledge that X is female and Y male, that the setting is a party, and that it is their first meeting. Johnson's response, particularly the verb “choke” sets up a third mental space, the Choke space. In this space, two focal roles are “victim” (Y') and Perpetrator (X'), where the perpetrator is responsible for causing the victim to gag. This is an odd version of the

choke space, since choking is typically associated with the lodging of an inanimate thing, such as piece of food or small object, in his throat. There is not usually an external agent responsible for the choking. The same is not the case with representations of "strangling".

The use of "choke" in conjunction with accusative "him" in the context of receiving words of praise sets up a metaphoric blend in which "being praised vulgarly is being choked" and all the negative, visceral connotations associated with choking. What Mrs. Piozzi objects to is Johnson's characterization of More's speech; such vulgar images associated with choking, so publically issued, evidence his rough and sometimes disgusting manner. The emergent structure in the blend (i.e., that "certain kinds of speech act like physical objects that can obstruct an individual's airway, producing graphic visceral responses by the body") is then projected back to the initial Dialogue space representing Johnson's public demeanor. It was not More's speech that was vulgar, but Johnson's reply. Note in passing that Malone characterizes More's words as an "indelicate and vain *obstruction* of compliment" for the seeming purpose of focusing attention on the Verbal Choking blend, this time to impute vulgarity on the speaker and not on the respondent.

Let us now examine Malone's contrastive account of the same incident with respect to force dynamics. It should be noted that Malone's account is a direct response to reading Mrs. Piozzi's book. While Piozzi construes the situation as an intra-psychological battle between the self and two competing subjects. Malone construes the situation as a social and inter-psychological conflict between two individuals. Johnson's responses to Hannah More's verbal overtures are the focus of attention. In this case, Hannah More is represented as manifesting a certain tenancy that can be translated into force dynamic nomenclature as a agonistic force imposing herself on another entity. Johnson, by contrast, plays the antagonist role of resisting her impositions. The reason this metaphoric vocabulary makes sense is due to the fact that Malone characterizes her actions. The use of the adverb *still* has the force dynamic property of persistence against an opposing force in "She still *laid it on*, and with the verb particle adding

a directional component to her speech, which is metaphorically represented as “laying”. Malone further characterizes her persistence by claiming “she continued still her eulogy”. Again the use of *still* in conjunction with *continue* are force dynamically marked. Let me add that the lexical item *eulogy* is a particularly pointed way of representing More’s inappropriate behavior by suggesting that she continues to forcibly impose her will on Johnson even when that will violates basic social decorum. One simply does not eulogize the living. It is More’s manner that is rough, not Johnson.

What is of interest next is the way Malone represents Johnson as the antagonist. Johnson is by all means socially more powerful; we presume that he has the power to stop More. What Malone represents is Johnson's measured attempts to restrain from exercising too great a force against More. But More’s inability to heed Johnson’s first attempts necessitates that he ratchet up the relative force of his opposing strength. What is being represented over several clauses is the dynamic nature of the magnitude of force.

The first mental space is the First Meeting space, with “Dr. Johnson” (X) and “Hannah More” (Y) as focal participants. In this space, participants X and Y greet each other, exchanging polite speech and establish familiar discourse patterns. The next space is the Flattery space, a space whose temporal logic follows the Initial Meeting space. In this space More praises Johnson. When completed, the representations suggest that she is doing most of the speaking and, further, is not really listening. She is not taking turns or being a good discourse participant. Since both spaces represent facets of social interaction, they are subject to force dynamic organization. As stipulated above, Johnson acts as the antagonist, More as the agonist. The Blend composed is what I call the Warning space. In this space, the force dynamic roles fuse with the social interaction roles. Once composed the blend develops over time such that the alpha and beta versions of the blend represent More’s persistent but inappropriate domination of the conversation. They correspond linguistically to *She still laid it on* and *She continued still her eulogy*. It is the gamma version of the blended space that represents the power shift among agonist and antagonist. Johnson finally applies the force necessary to stop More. Unfortunately, the

necessary force was extreme. (See figure 3.11 in the appendix.)

**Conclusion**

As this chapter has shown, the place of attention in linguistic theory is anything but settled. Generative linguists like Chomsky have argued that its role in setting linguistic parameters is tangential and beside the point, while functional and cognitive linguists of many stripes continue to regard attention as a crucial ingredient in the structure, acquisition, and use of language. In seeing a linguistics of attention as part of a greater grammar of attention, I am naturally drawn to the functional-cognitive linguistic approach. The linguistics of attention outlined in this chapter follows and extends the cognitive tradition by examining real usage data from an enunciation perspective that even cognitive linguists sometimes take for granted. The detailed descriptions of data in their precise contexts shows how a linguistic theory based on alerting, orienting, selecting, sustaining, controlling, and sharing attention can yield rich nuances to online meaning construction unappreciated, certainly, within the formal linguistic tradition, and only dimly appreciated within the functional-cognitive linguistic tradition.