THE NEW PSYCHOLOGY OF LANGUAGE

Cognitive and Functional Approaches to Language Structure

✳

Volume 2

Edited by

Michael Tomasello
Max Planck Institute for Evolutionary Anthropology
Leipzig, Germany

Discourse and Grammar

John W. Du Bois
University of California, Santa Barbara

Differences that separate grammar from discourse are not hard to find. Grammar describes sentences; discourse goes beyond the sentence. Grammar limits options by rule; discourse is what speakers do with the freedom that is left. Grammar is general; discourse varies at the will of its speakers and whim of their topics. Grammar is meaningless,[1] proudly so; meaning and pragmatic force lie at the heart of discourse. Grammar is pointless in a sense, possibly a good sense; discourse realizes the ends, whether communicative, cognitive, interactional, ideological, aesthetic or otherwise, that its producers seek to attain. It is no surprise that the study of grammar and the study of discourse are so often seen as worlds apart, pursued with different goals and different methodologies by different people. And yet language itself, in its actual occurrences, would seem to display at once the characteristics that attract both the grammarian and the discourse specialist. If language responds to either approach taken, can the distance between grammar and discourse be so great as the dichotomies imply?

Perhaps this perceived gulf makes it needlessly difficult to reach a full understanding of language. If we hope to learn how language works we will need to pursue multiple vantage points, while encompassing grammar and discourse within a single field of view. We need to exploit a stereoptical vision to integrate the two into one unified domain of phenomenal inquiry,

---

[1]This and the following point about grammar are associated with the approach of Zellig Harris (1951) and his intellectual offspring in the formalist tradition of generative syntax.

even as our theoretical constructs retain the character distinctive of their separate origins. Indeed, a potent trend has become evident in recent attempts to illuminate the foundations of linguistic structure and language function (Ariel, 1998; Chafe, 1987, 1994, 1998; Du Bois, 1985, 1987; Fox & Thompson, 1990; Givon, 1979, 1992, 1998; Hopper, 1998; Hopper & Thompson, 1980; Nichols & Timberlake, 1991; Thompson, 1997, this volume). We propose to study grammar and discourse together in order to understand how language comes to be what it is.

In this chapter I illustrate this trend with reference to a particular pattern lying at the intersection of discourse and grammar, what I have termed *Preferred Argument Structure* (Du Bois, 1985, 1987; Du Bois, Kumpf, & Ashby, in press). Preferred Argument Structure represents neither a discourse structure nor a syntactic structure per se, but a preference in discourse for a particular syntactic configuration of linguistic elements, both grammatical and pragmatic. Roughly, the claim is that in spontaneous discourse, the distribution of nominal referential forms (such as full lexical noun phrases or pronouns) across the various syntactic positions (subject, object, oblique) is systematically skewed. Speakers freely realize full lexical noun phrases in intransitive subject position or transitive object position, but strongly avoid placing them in transitive subject position. In a pragmatic parallel to this, new information (typically expressed by full lexical noun phrases) freely appears in intransitive subject or transitive object roles, but not in transitive subject role. This strong tendency—not a categorical rule—is evidenced widely in the spontaneous discourse of virtually all languages investigated to date (Du Bois, 1987; Du Bois et al., in press). In the following, I illustrate this pattern via its linguistic consequences, presenting short excerpts from English conversations, supplemented by a brief summary of the quantitative evidence from several languages. Along the way I try to point out some of the general implications of this work for the mutual connection and influence—even co-evolution—of discourse and grammar.

We have been used to thinking of grammar as the preserve of whatever generality and systematicity can be extracted from the phenomena of language. Such a conception leaves us with an inconvenient residue of randomness, to be disposed of in the lap of some field of language study or other. Lexicon used to be the dumping ground of irregularity. But the once-despised lexicon has now been cleaned up, its reputation refurbished as it becomes a bright new field of generalization. We recognize that the fine-grained patterning that permeates the field of lexical organization can become a foundation upon which generalizations of grammar are built. The social sphere has served its time on scapegoat duty, too. Linguistic variation between one social group or one society and the next could be dismissed as a mere surface perturbation—instigated by haphazard influences from local cultures—of the otherwise tidy picture of rule-bound normative

grammar. Psycholinguistic processing could be tapped to supply exculpating performance factors that would allow competence to remain pure and unaccountable. Pragmatics could be counted on to tie up any loose ends that grammar couldn't sort out, an all-too-convenient passing of the buck that should be considered suspect unless it's accompanied by delivery of the pragmatic goods in the form of specific working analyses.

But in the end it is discourse that has remained as the final stronghold of random fluctuation, if only because it stands as the domain of the speaker's ultimate personal prerogative. Here the freedom to express one's unique, unconstrained intention—what the self unfathomably wishes to say—would appear to preclude any generalization. At best the discoursist is limited to narrow claims of form–function symbolization: If what you want to say is a (certain) thing then say that (certain) thing using this form. A repressive tolerance for the untrammeled individualism of speaker intention leads us to assume that once grammar has skimmed its generalizations off the top there will remain no linguistic pattern of any consequence to be found among the whimsy products of talk.

The surprising thing, then, is that powerful trends of systematicity do remain to be discovered within discourse, even granting first-born grammar its due. If we know how to look—with a theoretical framework that helps us to see the patterning for what it is—we can recognize the systematicity of discourse, appreciating both its distinctive character and its critical impact on grammar. The real story of discourse and grammar research is that there is a place for pattern, and generalization, in both domains. Discourse and grammar each claim a distinctive type of patterning, neither of which is reducible to the other. And yet—here's the challenge—grammar and discourse interact with and influence each other in profound ways at all levels, so that in real life neither can even be accessed, not to mention explained, without reference to the other.

I make three theoretical assumptions about discourse and grammar. First, speakers exploit available grammatical structure to realize their goals in speaking. Second, the aggregate sum of what speakers do in discourse exhibits recurrent patterning beyond what is predicted by rules of grammar. Third, grammatical structure tends to evolve along lines laid down by discourse pattern: Grammars code best what speakers do most. From this perspective, forged through investigations of grammar and discourse-pragmatic function in typologically diverse languages, we are led to seek out cross-linguistically recurrent patterns of grammar on the one hand, and of discourse on the other. If we learn that a certain distinctive pattern of grammatical relations turns up in a variety of unrelated languages around the world, we take this wide occurrence of similar grammatical structure, arrived at through independent historical development, to be evidence of a fundamental pattern: something that needs to be explained. What we have

found is that wherever there is fundamental patterning in grammatical structure, we are likely to discover parallel, yet not identical, fundamental patterning in discourse function. One such discourse-and-grammar parallel is to be found in Preferred Argument Structure. But first we cast a brief glance at another approach, which assumes a rather different role for discourse in the description and explanation of grammar.

WHAT IS DISCOURSE?

Even where the value of discourse as a locus of grammatical research is granted, there may remain differences regarding what goals are to be pursued and what methods established. How one conceives the nature of discourse has much to do with how one conceives the nature of language itself. It will be instructive to briefly consider some alternative approaches, without any pretense to surveying the rich diversity of viewpoint that characterizes the present intellectual ferment in the arena of discourse.

One longstanding role for discourse is represented in the Americanist tradition of grammatical description (Boas, 1911), in which linguists and ethnographers of unfamiliar languages are urged to document the trilogy of grammar, lexicon, and text. The approach flourishes to this day and remains useful and valid for the goals it sets. Here the role of discourse ("text") relative to grammar becomes, in a sense, to substantiate the claims made in the grammar. Once a text in the language (usually a narrative) has been elicited from a native speaker, transcribed, and glossed, sentence tokens are culled from it and displayed to exemplify the grammatical types posited for the grammar of the language (cf. Aissen, 1987). The practice of culling individual examples from texts indeed goes back much further, to the traditions of historical linguistics and comparative philology, where it has long proved indispensable for documenting the facts of languages known only from older written sources. The hallmark of the approach is its use of discourse as a data source for documenting grammatical types.

Saussure, who as a historical linguist was well versed in such uses of texts as evidence about grammar, held that parole—in effect the aggregate of tokens of speaking (or writing)—could not be studied directly (Saussure, 1916). Only language, conceived as a system of types, was an accessible and viable object of study. On this approach the discourse data, however invaluable, remain transparent in a sense, almost invisible. They are taken as representative instantiations of types, the types constituting the actual focus of interest. Aside from its useful supporting role as a source of evidence for linguistic types, discourse as such is not seen. One does not look there for pattern, generality, or explanation; these rather remain the purview of the language as system. Yet even Saussure recognized a most intimate connec-

tion between discourse and grammar, when he acknowledged that language is both instrument and product of speaking.

In one modern variant of this approach, pragmatics and even discourse may be usefully described in terms of types. In place of the type categories of sentence grammar, it is now discourse-pragmatic types whose function is to be described via evidence drawn from discourse. To characterize a discourse particle like *anyway* or *well* whose impact extends beyond the sentence, a discourse-pragmatic function is invoked—*anyway* signaling, for example, "end of off-topic excursion," whereas *well* might be said to mark "dispreferred speaker response." In effect, the function of a word (*anyway, well*) is stated as the pairing of a form type with a meaning type. Of course the discourse data in this case cannot be invisible, because they must be analyzed to apprehend the pragmatic meaning of the particle or structure in question. But after the discourse tokens have yielded their service to the types, they may remain elusive in themselves. This approach is on target for analyzing specific forms bearing well-defined meanings in the realm of discourse-pragmatics. The classic case is discourse particles (Schiffrin, 1987), whose meanings often commingle semantic and pragmatic components, the latter being difficult or impossible to discover without an astute analysis of extended stretches of natural discourse. The success of the method lies in its ability to recognize an extended discourse pattern, associate it with a specific linguistic form, and reduce the whole complex of use to a compact description of the form's meaning. This is the familiar linguistic sign of Saussure (1916)—what could be called the "sign function," to be contrasted with what I describe later as the "structure function" of grammar.

The temptation, though, is to extend this potent method beyond its effective range. Because so much of language is describable in terms of form–meaning pairings, one can be led to see all of language in this light. (For an approach that seeks to take the symbolizing role of grammar as far as it will go, see Langacker, 1987, 1998.) But the form–meaning method as applied to discourse may reach its limit once one undertakes to analyze the more abstract and highly grammaticized roles such as subject and object (as we see later).

Where some methodologies treat discourse tokens as a mere means of arriving at, or confirming evidence of, a type description, the present approach to discourse has no wish to efface the tokens, neither in the process of analysis, nor in the summation of generalizations, nor even in the framing of explanation for why the type system of grammar is as it is. To achieve the explanatory goals we have set ourselves, the body of utterance tokens necessarily becomes an object of study in its own right. If the goal is to understand language, even in its aspect as a system of grammatical types, we find it essential to simultaneously seek knowledge of the fundamental pat-

terning of the mass of instances of language use, what I have called the *token aggregate* (Du Bois, 1987). The accumulated linguistic experience of a lifetime amounts to a body of utterances encompassing form, meaning, and contextualization. This aggregate of realized tokens of language use exhibits patterning that in its broad outlines, allowing for a certain degree of analytical and statistical abstraction, remains remarkably constant from one speaker to the next and even from one language to the next. The challenge is to probe the patterning of discourse in a way that speaks to the language as a whole: to discover how talk shapes grammar as much as grammar shapes talk.

Why pursue specifically speaking, if writing equally constitutes discourse? Speaking has several key features to recommend it, despite the admitted convenience of working with written material. What, after all, is discourse? What do we hope to find in it? Natural discourse encompasses any authentic instance of language use in all its manifestations, which is to say, any organically motivated act of producing or apprehending language, via any medium, spoken, written, or otherwise, subsuming the full scope of the situated utterance in all aspects of form, meaning, and contextualization. Leaving aside artificial discourses, such as might be constructed for purposes of experimental control, we face an almost unlimited variety of genres or types of language use: newspaper editorials, novels, horoscopes, comics, boilerplate legalese, grocery lists, love letters, lectures, credos, religious rituals, roll call, sporting news, enthusiastic recitations of bad poetry, longwinded jokes, chit-chat, put-downs, the clipped commands of workplace talk. For natural discourse there can be no a priori requirement of truth, sincerity, beauty, respectability, coherence, or even grammaticality in the normative sense. Each of these uses of language is natural discourse, and each merits study on its own terms. Yet amidst the diversity there is one kind of use that is often singled out as worthy of special attention, that of face-to-face conversation. Not so much a genre as a matrix within which particular genres may be invoked, conversation, while exhibiting far-reaching variability within itself, can be characterized as pervasive, spontaneous, interactional, and contextualized. It is pervasive, constituting the commonest use of language for virtually all language users, the first learned by children, and one of the few forms of language use found in all languages, at all stages of human history (Chafe, 1994, 1998).[2] No written genre can stake such a broad claim, even for today. Conversation is typically spontaneous, giving us the opportunity to witness on-line planning and other processes that attend the unfolding of discourse as it is produced. Conversation occurs

---

[2]Ironically, one of the other forms of discourse most pervasive in human life is one that is the opposite of conversation in many respects: ritual language, which appears in all human cultures (Du Bois, 1986). Arguably it has been with us from the earliest stages of humanity.

curs in an environment of dialogic interaction between speakers, where give and take among interlocutors affords us additional clues as to how the discourse is being interpreted and evaluated by the participants themselves. And finally, conversation is highly contextualized, filled with subtle cues at all levels marking the relation of utterances to contexts of prior discourse, to situational and cultural contexts, to contexts of social relations between speech event participants, and even to the mutual cognitive context within which the dialogic interaction is embedded. Spontaneous discourse reflects the speaker's cognitive frames and processes, as when a pause reflects a stage of planning for the next utterance (Chafe, 1987, 1994; McNeill, 1992); and it reflects the speaker's cognitive model of the hearer's cognitive model, as when a referential form is selected that takes into account what the listener does and doesn't know (Ariel, 1990, 2001; Chafe, 1994, 1998; Clark, 1996). To be sure, cognitive processing takes place whenever we write for others to read, too. But if we do our writing alone in a room, all evidence of cognitive processing that would be carried by pauses, ruminations, restarts, rephrasings, corrections, rhythm, and more is carefully and deliberately effaced through successive waves of editing, as we strive to present our audience with a seamless piece of prose. To the extent that we would hope to access the cognitive processes that underlie discourse and ultimately grammar, we need to be present to witness the innumerable cues that attend its actual moment-by-moment coming into being. This wealth of information is most fully available in spontaneous spoken discourse. In all, spoken language, and specifically conversation, remain closest to the living embodiment of language. If we seek to understand the system of grammar, and if grammars code best what speakers do most (Du Bois, 1985), then it is to spoken language that we must devote our most scrupulous attentions.

There are additional reasons for preferring authentic discourse data over inspecting one's intuitions about invented discourse. There is reason to believe that native speakers simply lack reliable intuitions about linguistic patterning on a scale as large as that of extended discourse. Whatever ability individuals may have to judge grammaticality or acceptability at the sentence level (and of course the reliability of such claims is controversial even at the sentence level) seems to desert them when the higher level patterning of discourse is broached. For example, decades of work on the syntax of isolated sentences failed to elicit any articulate awareness of a key patterning of grammar, Preferred Argument Structure, which is apparently followed quite regularly by all speakers. On the other hand, certain sentence structures that appear perfectly unproblematic in isolation are nevertheless strongly avoided in spontaneous discourse. Evidently it takes a systematic, theoretically conscious assessment of a mass of authentic spontaneous discourse data to discover the fundamental organization of discourse.

In summary, spoken discourse most transparently reveals grammar in use. Crucially, this use is use in context. Discourse and grammar research undertakes to analyze grammatical elements and structures in their ecological relationship to each other and to overall functional patterns of use in the total context. By context we understand no less than the sum of the salient situation, prior discourse, interlocutors' actions, models of mutual knowledge, cultural frames, and even the publicly visible processes of marshaling cognitive resources in acts of communication. The ecological perspective on grammar (Du Bois, 1980, 1985) becomes especially important once we undertake to examine the strategic use of grammar across extended sequences in discourse (see "Constraint and Strategy" later).

DATA

The data in this chapter are drawn from naturally occurring conversations and other speech events in the Santa Barbara Corpus of Spoken American English (Du Bois, 2000). For reasons of space, most examples cited are very brief, usually just a line or two, without the larger discourse context. In some cases the transcription has been slightly simplified for the sake of conciseness and accessibility.[3] At the start of each example, the citation includes the short title (e.g., ATOMS, denoting the discourse whose full title is "Atoms Hanging Out"), followed by the line number in the transcript. The published corpus (Du Bois, 2000) is available for consulting the larger discourse context and full transcription detail for many of these examples, if desired.

A key aspect of the selection and presentation of discourse examples in this chapter concerns the intonation unit: Most excerpts constitute exactly one intonation unit, no more and no less. The intonation unit can be defined as a stretch of speech occurring under a single coherent intonation contour, as reflected in several specific prosodic cues including initial pause and final lengthening (Chafe, 1987, 1998; Du Bois, Schuetze-Coburn, Paolino, & Cumming, 1992). In the transcription system used here (Du Bois et al., 1992; Du Bois, Schuetze-Coburn, Cumming, & Paolino, 1993), each line represents one intonation unit. Intonation units play an important role in much of current research on spoken discourse and grammar, because of their status as perhaps *the* fundamental unit of cognitive processing, social interaction, and other domains (Chafe, 1987, 1994, 1998;

---

[3]The main simplification pertains to cited utterances in which the speaker's words overlap with another utterance, not included in the citation, by a second speaker. Had both utterances been cited, each would contain square brackets [ ] marking the point where overlap begins and ends. But wherever only one of a pair of overlapping utterances is cited in the excerpt, all overlap notations are left out, to avoid confusion.

Du Bois et al., 1992; Ford & Thompson, 1996; McNeill, 1992). Intonation units tend to show a fairly close correspondence with simple clause structure; consequently, in many of the examples cited, the intonation unit is roughly coextensive with a clause (e.g., a verb plus its nominal arguments).

Because the type of discourse and grammar research represented here makes claims about recurrent tendencies in discourse, its verification ultimately requires a numerical accounting of the commonest ("preferred") configurations in a body of discourse data. Such counts have long been a hallmark of discourse and grammar studies (Ariel, 1990, 1998; Clancy, 1980; Du Bois, 1980; Fox & Thompson, 1990; Givon, 1979, 1983, 1992; Hopper & Thompson, 1980; Thompson & Mulac, 1991), particularly in those establishing a common Preferred Argument Structure across a wide range of languages (Du Bois, 1987; Du Bois, Kumpf, & Ashby, in press; and references therein). Though an explication of the quantitative methodology is outside the scope of this chapter, the results of some of the relevant quantitative studies in several languages will be briefly summarized in Tables 2.1 to 2.5.

The observable pattern that concerns us most in this chapter involves the differential distribution across the clause of full lexical noun phrases versus pronouns. To make this contrast more visible to the reader, I adopt two supplementary formatting conventions for the examples cited: (a) full lexical noun phrases are boldfaced, and (b) pronouns (and ellipted "zeros") are underlined.

PREFERRED ARGUMENT STRUCTURE: WHERE GRAMMAR MEETS DISCOURSE

The verb is often accorded a dominant role in the grammatical structuring of its clause, which it realizes principally via its argument structure. In one popular view of clause structure, a given verb's argument structure grammatically specifies how many nouns will accompany it, and which roles they will play, within the clause (Chafe, 1970; Fillmore, 1977, this volume; McNeill, 1992, p. 253ff). From a cognitive point of view, an argument structure is nothing more than a structure of expectations triggered by a verb. Specifically, each use of a particular verb token raises the reliable expectation that a certain predictable configuration of nominal roles will occur in meaningful relation to it. The verb *enjoy*, for example, can be predicted to co-occur with a subject argument and an object argument, corresponding to the experiencer of enjoyment and the thing enjoyed, respectively. The particular nominal roles and their configurations will vary from verb class to verb class. Though subject to certain qualifications (Goldberg, 1998), this account will suffice for present expository purposes.

Among the arguments of a verb, some are core (subject, direct object, indirect object) and some are not (obliques, such as locative or temporal adverbials). The core arguments tend to be those most central to the clause structure. They are highly grammaticized, which is to say, obligatory rather than optional and relatively bleached of specific semantic meaning. For Preferred Argument Structure it is the clause core that is the primary unit of analysis, that is, the verb with its core arguments. To this must be added a discourse perspective: We are interested in the clause core as bounded by the intonation unit. In general all clause core elements tend to fit within the unified intonation contour that demarcates a single intonation unit.[4]

Referential Form

The noun phrase, considered as a structural position within a clause, can be realized by a wide variety of forms, such as a full lexical noun phrase with or without various modifiers, a personal pronoun, and so on. Grammarians would normally leave the selection among these realizations to factors outside the domain of syntax. Recognizing that discourse-pragmatic considerations such as referential continuity, information status, accessibility, and so on influence the choice of referential form (Ariel, 1990; Clancy, 1980), nevertheless from a narrowly structural perspective the variation counts as optional. As far as the grammar is concerned, apparently, the speaker of a clause like the following is free to choose whether to realize the subject of the verb *speeding up* with either a full lexical noun phrase or a pronoun:

(1) {ATOMS 682}
    PHIL: . . . **the molecules** are speeding up,

In (grammatical) principle, the full lexical noun phrase *the molecules* could be substituted by the pronoun *they* with no ill consequences (at least for grammaticality). Conversely, any pronoun should in theory be substitutable by a full noun phrase.[5]

Of course, once we take into account discourse-pragmatic factors (such as those involved in information management) as opposed to merely structural factors, the noun–pronoun alternation is demonstrably not free:

---

[4]Clausal elements may also exceed the boundaries of one intonation unit, spilling over into the next. Although this phenomenon is of considerable interest in its own right as evidence of the limited information-carrying capacity of the intonation unit, and of important cognitive processes of verbalization, consideration of this topic would carry us well beyond the scope of this chapter.

[5]We leave aside phenomena known as binding, where some current discussions treat so-called pronominals (as opposed to anaphors) as subject to grammatical limitations of a very different order from those treated here.

(2) {DEADLY 1354}
    JOANNE: **His tail** is curved like that.
               Does <u>it</u> always get curved like that?

This first mention in discourse of the tail of a turtle (visible in context) is realized by the full lexical noun phrase *his tail,* whereas its second mention, in syntactically similar environment, is realized by the pronoun *it.* Taken in isolation from its discourse context, either sentence could in theory be realized with either noun or pronoun. But within the authentic discourse context, substitution of the grammatically alternative form cannot be treated as inconsequential: It could change the conveyed meaning or even induce incoherence. Speakers in discourse are careful about keeping track of which referents have been previously introduced and which are only now being introduced for the first time, mindful as they are of their interlocutors' current state of shared knowledge (or lack of it). Strategies for information management partly govern linguistic choices like that between lexical noun and pronoun, a phenomenon subsumed under the rubric of discourse pragmatics.

In the next example there appears a grammatical alternation between a proper noun (*Harold*) and a first person pronoun (*me*) in parallel syntactic contexts:

(3) {LAMBADA 1269}
    PETE: Without even telling **Harold**?
    HAROLD: . . Without telling <u>me</u>?

From the isolated-sentence syntactic perspective, the selection between noun and pronoun here would again count as free, as evidenced by (nearly) equivalent syntactic environments. (The appearance of two distinct words in nearly the same syntactic context precludes formulation of a syntactic rule predicting which must occur.) But from the discourse perspective the choice is evidently not free, being governed rather by pragmatic factors (including interactional criteria for self- and other-reference that need not concern us here).

So when we ask whether the selection of referential form (noun, pronoun, etc.) is *optional,* we need to add: With respect to what governing theoretical domain? If we take care to distinguish between syntactic and discourse-pragmatic determinants of referential form, we note that in each of these cases, syntactic structure seems to have no determining influence on the selection. Grammar is neutral about the choice between noun and pronoun. As long as there is an argument role available, as defined by the argument structure of a given verb, either noun or pronoun may equally occur, without preference. The discourse-pragmatic perspective, on the other

hand, is not neutral, treating the selection as governed by factors in the total context of use such as referential continuity, degree of accessibility, speaker self-reference, and other information management factors. In these cases we conclude that there exist discourse-pragmatic, but not syntactic, determinants of referential form.

To ask where *in a syntactic clause structure* a full noun phrase can occur would seem therefore to be mixing up theoretical domains, or at best inviting a trivial answer: anywhere. That is, anywhere an argument role (say, subject) has been licensed by the argument structure of a verb (say, a transitive verb), a full noun phrase could supposedly occur.

But there is another perspective, in which this kind of question can hope to receive a more meaningful answer. Preferred Argument Structure carves the matter up a little differently, not necessarily respecting the traditional division of labor between discourse and grammar. Although recognizing the influence of discourse-pragmatic factors on the choice of referential form, it proposes an additional influence mediated by the syntactic structure itself. This comes not in the form of a syntactic rule categorically determining grammaticality, but rather in the form of an enabling influence: The syntactic role within the argument structure of a verb constitutes a factor influencing the probability of a lexical noun (as opposed to a pronoun) occurring there. For this reason we are especially interested, in what follows, in the question of where in the clause a given referential form occurs. In the end this will lead us to a conception of syntactic configurations as not merely classified into grammatical and ungrammatical, but as fulfilling specific structural roles in the cognitive management of information processing.

To prefigure my direction of argument, I propose that the grammatical structure of the clause contributes a kind of architectural framework for the management of cognitive processing tasks. Relatively demanding tasks (like processing the introduction of a new referent into the discourse) are largely confined to specific grammatical environments within the clausal architecture. Non-demanding tasks (like tracking an accessible referent already introduced), on the other hand, may occur anywhere in the clause. But we are getting ahead of ourselves. In order to build toward a picture of the grammatical ecology of information management, we must first lay out a survey of the available linguistic resources: the architectural repertoire of basic grammatical structures employed by speakers in discourse.

One-Place Predicates

We begin with the most basic configuration of arguments, that defined by a one-place predicate. An intransitive verb like *run* or *come* articulates an argument structure specifying that it will be accompanied by a single core argument noun phrase, typically functioning as its subject. By convention we

label this the S role (a label that applies specifically to *intransitive* subjects, not to subjects in general). For clarity's sake, the first example of each set will be labeled for the argument roles S, A, and O, and subsequent examples repeating the same pattern will be left unlabeled.

(4) {CONCEPT 416}
   MARILYN: .. **The fish** are running,
                 S

(5) {RUNWAY 637}
   RANDY: **The Caravan**'s coming across,

(6) {DOLLARS 629}
   DAN: .. **policies** change.

The intransitive subject or S role may be filled by a noun phrase of almost any shape, size, and meaning. These examples represent relatively modest full noun phrases, with just an article plus noun (*the fish, the Caravan*), or even a bare noun alone (*policies*). But the S role can easily accommodate more substantial noun phrases, containing adjectives and all kinds of other modifiers including relative clauses. (See later examples for a wider range of lexical noun phrase types.) In all, we can say that the S role is *free*, allowing noun phrases of any form and size without constraint.

The array of verbs and other predicates that invoke a single core argument is semantically diverse, as is the semantic relation that obtains between subject and predicate. This relation can range across agent, actor, source, recipient, theme, patient, and others. What remains constant is a syntactic argument structure that reserves a place for a single core argument.

Two-Place Predicates

A transitive verb, like *enjoy* or *eat*, invokes two core arguments: its transitive subject (termed the A role) plus its direct object (the O role). Hence there is the possibility of two full lexical noun phrases. But is this grammatical potential fully exploited in discourse? The evidence of discourse patterning suggests not:

(7) {HOWARDS 1102}
   JANICE: (H) But I enjoyed .. **the movie**.
                 A   O

(8) {ZERO 489}
   NATHAN: ... She's eating **that bu=g**.

(9) {TREE 519}
   MARY: So I stopped **the car,**

(10) {JURY 871}
   REBECCA: you called . . **the police.**

The first thing to note about this highly characteristic transitive pattern is that in discourse there regularly appears just one full lexical noun phrase. The other argument, in contrast, is represented by just a pronoun (or other such reduced form). The prevalence of this pattern in discourse is what originally led me to formulate a constraint on grammatical quantity in core argument positions, termed the *One Lexical Argument Constraint* (Du Bois, 1987):

(11) Avoid more than one lexical core argument

That is, even if a transitive verb provides two core argument positions, only one of them is to be exploited to introduce a lexical noun phrase. The other argument position will sustain only a pronoun, or ellipted "zero." Note that this is a soft constraint, not a categorical grammatical rule. The pattern is followed generally but not without exception.

The second aspect of this pattern pertains to grammatical role. Lexical nouns do not appear with equal likelihood in all roles. In the direct object (O) role, lexical nouns occur freely, but within the transitive subject (A) role, we find that pronouns are used far more frequently. Put another way, the single lexical argument allotted by the first constraint tends to favor the O role but avoid the A role. Based on counts confirming this pattern as pervasive in spoken discourse, I have formulated a constraint on the syntactic role of lexical arguments:

(12) Avoid lexical A

This preference, dubbed the *Non-Lexical A Constraint,* represents a very strong tendency in discourse across numerous languages (Du Bois, 1987; Du Bois et al., in press; see following). But again, this is a soft constraint rather than a categorical grammatical rule, and exceptions do occur.

Three-Place Predicates

Ditransitive verbs like *give, tell,* and *show* raise the number of core argument positions to three: subject (A), direct object (O), and indirect object (I). But the question remains as to whether they thereby raise the actual capac-

ity of the spontaneously uttered clause to carry full lexical noun phrases. The evidence from discourse suggests not (Schuetze-Coburn, 1987). We see a pattern reminiscent of that for transitive verbs, in which some argument positions are left underexploited (in the sense of not carrying full noun phrases). The question is, how many?

(13) {LAMBADA 1126}
   MILES: I told you **that story,**
           A   I   O

(14) {DEADLY 1561}
   JOANNE: You have to buy em **feeder fish.**

(15) {HOUSEHOLD 1245}
   RON: Yeah she showed me **all that stuff.**

Despite the three possibilities (grammatically speaking) for lexical noun phrases in each ditransitive clause, in the prevailing discourse pattern only one is realized. The other two arguments are regularly realized as pronouns. Ditransitives thus dramatically confirm the One Lexical Argument Constraint. Even with three opportunities, only one is taken.

Moreover, ditransitives in discourse are consistent as to where the single lexical argument falls. As with transitives, it falls predominantly in the direct object role, as seen in the previous examples. The transitive subject role is avoided, conforming to the Non-lexical A Constraint. In a related constraint, lexical noun phrases also tend to be avoided in the indirect object role (though not, apparently, as strongly as for A role).

Quantitative Overview of Lexical Argument Distribution

Though a full quantitative justification of the claims of Preferred Argument Structure goes well beyond the scope of this chapter, we may offer a brief glimpse into some of the numerical evidence that has been adduced in studies over the last several years, based on analyses of spontaneous spoken discourse in a small sampling of geographically, historically, and structurally diverse languages of the world. All the studies summarized here specifically undertook to address the claims of Preferred Argument Structure, testing its viability within the language in question by amassing quantitative evidence on the distribution of lexical (and new) mentions relative to clauses and their argument structure roles. All the studies employ some variant of the methodology initiated by Du Bois (1987), usually using narra-

tive discourse data.[6] Even so, differences arise in the specific analytical choices made, so caution should be exercised in comparing across studies. As each language represents a somewhat distinct grammatical ecology, the implementation in context of even so seemingly straightforward a category as "lexical noun phrase" can vary subtly, or substantially, from one language to the next. This requires each analyst to make coding decisions on the ground, as it were. Moreover, the particular discourse genres and even subgenres employed in the various studies often differ in potentially significant ways, with consequences that are not yet fully understood.

With all these caveats, the similarities across languages still manage to shine through. Apparently the pattern of Preferred Argument Structure is robust enough that even differences of method, language, and grammatical type cannot obscure it. Evidence from five languages on lexical quantity within the clause is presented in Table 2.1. In all the languages, clauses containing two lexical core arguments are much rarer than clauses with one or zero lexical arguments, thus providing evidence for the One Lexical Argument Constraint.

TABLE 2.1
Lexical Argument Quantity: Frequency
of Clauses by Quantity of Lexical Arguments[7]

Quantity:	0		1		2		Total	
Language	n	%	n	%	n	%	n	%
Hebrew	261	(50)	252	(48)	9	(2)	522	(100)
Sakapultek	211	(46)	240	(53)	5	(1)	456	(100)
Papago	430	(57)	307	(40)	22	(3)	759	(100)
English	252	(47)	241	(45)	39	(7)	535	(100)
Gooniyandi	2318	(62)	1305	(35)	114	(3)	3737	(100)

[6]In making comparisons across languages, it is important to compare similar genres. Most studies of Preferred Argument Structure published to date report on narrative, so that is mostly what I present in the tables that follow. The data for Sakapultek, Brazilian Portuguese, English, Hebrew, Papago, and Gooniyandi are narrative, of which the first three (and part of the fourth) are narratives elicited from viewers of a short film known as the Pear Film (Chafe, 1980). The Spanish and French data are relatively monologic sociolinguistic interviews. The Japanese data are conversation. (Conversational discourse is documented more extensively in the studies to appear in Du Bois, Kumpf, & Ashby, in press.)

[7]The sources for each language are: Hebrew: Smith (1996); Sakapultek: Du Bois (1987); Papago: Payne (1987); English: Kumagai (2001); Gooniyandi: McGregor (1999). Pagago counts include all overt (i.e., not just lexical) mentions, but due to Papago's grammar this is not likely to greatly affect the results. Note that English counts reported by Kumagai include two instances of 3-argument intransitives and one instance of 3-argument transitives (not included in Table 2.1), suggesting that the counts include some non-core arguments.

TABLE 2.2
Lexical Argument Role: Syntactic Role of Lexical Core Arguments[8]

Role:	A		S		O		Total	
Language	n	%	n	%	n	%	n	%
Hebrew	18	(8)	103	(44)	111	(48)	232	(100)
Sakapultek	11	(5)	126	(58)	81	(37)	218	(100)
Papago	37	(10)	169	(47)	152	(42)	358	(99)
English	21	(8)	90	(35)	146	(57)	257	(100)
Spanish	35	(6)	215	(36)	341	(58)	591	(100)
French	32	(5)	290	(45)	324	(50)	646	(100)
BrPortuguese		(8)		(39)		(53)		(100)
Japanese	48	(7)	320	(48)	293	(44)	661	(100)

If the typical clause has at most one lexical core argument, the question arises as to where in the clause it will go. Table 2.2 shows evidence from eight languages suggesting that lexical arguments occur freely in the S role (35%–58%) or the O role (37%–58%), although tending to avoid the A role (5%–10%). The Non-Lexical A Constraint is thus supported across an array of very diverse languages.

Table 2.3 makes the simple point that the scarcity of lexical A's is not explained by a scarcity of A's in general, as there are plenty of transitive clauses in discourse (although intransitives almost always dominate). In these studies, from a third to a half of all clauses are transitive.[9]

Although this brief numerical interlude can only be suggestive, it does provide some idea of how pervasive the cited argument structure configurations are. For full quantitative analysis the reader is referred to the original sources cited, as well as to Du Bois (1987) and the studies in Du Bois et al. (in press).

Summary: Free and Constrained Roles in Argument Structure

How can we characterize the prevailing discourse patterns? Consider the transitive clause. With respect to allowing lexical noun phrases in transitive

[8]The sources for each language are as listed in the note for Table 2.1, plus: Spanish: Ashby and Bentivoglio (1993); French: Ashby and Bentivoglio (1993); Brazilian Portuguese: Dutra (1987); Japanese: Matsumoto (1997). Non-core lexical noun phrases (e.g., obliques) are not included in this table (but see Table 2.7). Note that the numbers for "intransitive" S combine all S role arguments of one-place arguments, including S of copular/equational clauses. For Japanese, Matsumoto's figures combine all overt forms including overt pronouns, not just lexical nouns. (This is presumably justified by the fact that most Japanese "accessible" reference is actually accomplished by zeros.)

[9]Note also that by definition we expect to see as many instances of A roles as of O roles, as each transitive verb token normally has one argument position of each type.

TABLE 2.3
Clause Type: Frequency of Intransitive Versus Transitive Verbs[10]

	Intransitive		Transitive		Total	
Language	n	%	n	%	n	%
Hebrew	256	(47)	286	(53)	542	(100)
Sakapultek	277	(61)	179	(39)	456	(100)
Papago	523	(69)	236	(31)	759	(100)
English	309	(58)	226	(42)	535	(100)
Spanish	979	(63)	571	(37)	1550	(100)
French	1025	(68)	481	(32)	1506	(100)
Japanese	764	(68)	357	(32)	1121	(100)

clauses, the O role can be said to be *free*, whereas the A role is *constrained*. "Free" means that the lexical argument *may* appear, not that it must. We can represent the argument configuration schematically using a capital letter to represent a free role, and a small letter for a constrained role. Thus the typical array of core arguments in a transitive clause in spoken discourse is:

(16) a O

As for one-place predicates, we saw that the single core argument position (S role) is free, in that it allows full lexical noun phrases. Applying the same notation to intransitives, the argument pattern for a one-place predicate is represented simply as:

(17) S

For ditransitive clauses, which introduce an additional core argument (the indirect object, or I role), the typical argument pattern is schematically:

(18) a i O

Here only the direct object is shown as a free role; the other two are constrained. Putting it all together, we can summarize the patterning of free and constrained roles for the three clause types. The arguments of one-, two-, and three-place predicates, respectively, are represented schematically as follows:

---

[10]The sources for Table 2.3 are the same as those for Table 2.2.

(19) S
   a O
   a i O

Note that, even as the number of core argument positions is increased, the number of allowed lexical arguments remains constant at one.

REFERENTIAL PRAGMATICS AND INFORMATION MANAGEMENT

There is more to the distribution of lexical and pronominal arguments than just a patterning in grammar, of course. Referential pragmatics, with its intimate connection to processes of information management, plays a role that cannot be overlooked. An aspect of information flow in discourse, referential pragmatics refers specifically to the pragmatic factors influencing selection of the form, structure, role, and interpretation of referential forms such as noun phrases.[11] For example, when there is a new entity to be introduced into a discourse, this will characteristically motivate the use of a full lexical noun phrase of some kind. (The precise size and shape of the nominal reference can vary tremendously, in response to several additional referential-pragmatic factors including the scale of cognitive accessibility [Ariel, 1990, 2001].)[12] Once the entity has been introduced, in subsequent discourse it will generally be tracked by a reduced form such as a pronoun. These linguistic choices reflect the speaker's assumptions regarding the listener's cognitive processing cost for accessing the intended referent. A full noun phrase, characteristically deployed when new information is being introduced, is thus taken to be an index of relatively high effort in accessing the referent (Ariel, 1990, 2001). A pronoun, invoked when the referent is more or less obvious from context, is interpreted as an index of relatively low cognitive effort to access.

It might seem that we are putting things backward. Shouldn't a full noun phrase be easier to process, since it carries within itself more complete information spelling out explicitly who or what the intended referent is? Isn't

---

[11]Noun phrases are labeled referential forms because they are of a type typically used to refer, not because they always refer. Ironically, the first challenge faced by the student of referential pragmatics is to separate out those uses of noun phrases that are *not* referential, in order to mount a coherent analysis of the remaining, truly referential cases.

[12]The concept of "accessibility" articulated by Ariel (1990, 2001) differs from Chafe's (1994) "accessible" in several significant respects. Ariel's term denotes the entire range of a continuous scale of accessibility, whereas Chafe's refers to the middle term of a three-way contrast among discrete cognitive activation statuses: given, accessible, new. In this chapter, the term "accessible" is used for the range from mid accessibility through high accessibility on Ariel's scale.

a pronoun harder to process, since it carries so little information that it requires the listener to infer the referent from other cues? This logic might be applicable if speakers selected their referential forms at random or uncooperatively; for example, thrusting a pronoun upon an unprepared addressee. But speakers are not so cavalier. Once we acknowledge the implicit contractual commitments that bind the collaborating participants in a conversation (Clark, 1996), we can understand the cooperative exchange of information in a different light. In general speakers use pronouns precisely when warranted, that is, when the contextual ground has been sufficiently prepared in advance that no more than a hint is needed to point the listener toward the right referent. That hint is the pronoun. More generally, speakers consistently take care to assess just how much help their addressee will need to access the intended referent, so as to select a referential form that will provide precisely the right amount of information: neither too little nor too much. Am I introducing a new referent that my addressee has never encountered before? Or is it obvious who I'm talking about because it's the same person I just mentioned in the previous clause? The first case will call for a full noun phrase of some kind, more or less elaborated, whereas the second will generally call for a pronoun (or even an ellipted "zero" if the cognitive context is right).

Yes, a speaker in the heat of conversation could make a mistake, misjudging the listener's difficulty of access for a particular referent and consequently selecting an inappropriate form. But such miscalculations are few and far between, as long as we are dealing with natural spontaneous discourse, in which speakers speak freely from their own motivations and can draw on their own situated awareness of the interactional and informational context they are embedded in. In general a speaker's choice of referential form can be counted on as a sensitive and reliable indicator of the perceived cognitive cost of accessing a referent. Supporting the reliability of this methodological assumption, we note that it is relatively rare for a listener to halt the flow of interaction in order to request clarification for a pronominal reference.

This brief account of the cognitive processes and costs attending the exchange of information between discourse participants, though necessarily simplistic, must suffice for present purposes. We might mention, however, that the emphasis on the hearer's cost of access—as judged by the speaker—is not meant to preclude a role for the speaker's own immediate cost in developing the conceptualization and its verbal formulation in the first place (Chafe, 1994, pp. 71ff, 1998; McNeill, 1992, pp. 218ff). The growth process of the speaker's emerging utterance carries significant cognitive costs, too, which I suggest will again correlate with the selection between nominal reference forms of greater or lesser elaboration, and their distribution across argument structures. All that is required at this point for

our analysis of discourse patterning is that lexical noun phrases reflect greater cognitive cost than pronouns, whether the cost accrues to the speaker, the hearer, or both; and that these costs can be consistently localized in relation to grammatical structures.

In the next three sections I revisit the available grammatical repertoire of intransitive, transitive, and ditransitive argument structures in order to consider what role referential pragmatics plays in their patterning and use, as speakers and hearers strive to systematically manage the cognitive costs of information exchange in discourse.

Intransitive Pragmatics

In speaking of *intransitive pragmatics* I purposely juxtapose a structural term (*intransitive*) with a functional one (*pragmatics*). But how could these two belong together? We have already noted a certain patterning of grammar in discourse, whereby one-place predicates freely admit lexical mentions in the S role. We can now ask whether there is a functional basis for this distribution. What kind of pragmatic function is responsible for the use of all those noun phrases? Consider the following clause:

(20) {CUZ 995}
ALINA: .. **this new wa=ve of people** comes in.
        S

A new referent is here introduced into the discourse for the first time, via *this new wave of people*, as subject of the predicate *come in*. Presumably this new introduction is accompanied by a certain amount of initial cognitive processing, as a new cognitive file (Du Bois, 1980) is opened for its referent. In the next example, similarly, the first mention of a particular child appears in the S role as an indefinite noun phrase:

(21) {VETMORN 351}
HEIDI: ... **a little kid** answered,

Information introduced in intransitive clauses is not limited to humans or other topically salient referents. In the next example (from a narrative about a car accident), *glass* is mentioned for the first time, in the S role:

(22) {FEAR 587}
WALT: (H) **g=lass** is flying everywhere,

One-place predicates, with their unconstrained S role, can be counted on to allow the speaker unhampered fulfillment of the more demanding cognitive tasks, such as introduction of a new referent. They provide the

simplest available clausal structure for information management. (In light of this role among adults, it is instructive to consider Tomasello's [1992] finding for children that one-place predicates play a critical role in the early development of verbal argument structure. In the earliest phases of development, even verbs that adults treat as two-place are subsumed into the congenial frame of one-place predicates.)

Transitive Pragmatics

With transitive verbs like *hire* and *devise* the presence of a second core argument role gives rise to a functional contrast between one free role, the O, and one constrained role, the A. The free role is where new entities are (optionally) introduced. In the following example, the O role lexical noun phrase *another voice* represents the first mention of its referent in the discourse, whereas the A role pronoun *they* refers back anaphorically to a previously mentioned referent.

(23) {NOTIONS 87}
JIM: they hired **another voice.**
          A   O

The same pattern appears in the following two examples:

(24) {VISION 359}
DANNY: (H) .. and so she devised **a plan.**

(25) {DEATH 818}
PAMELA: ... I certainly miss **my do=g.**

The lexical noun phrases *a plan* and *my dog* each represent first mentions in their respective discourses. In contrast, the pronouns *she* and *I* refer to referents that are cognitively accessible based on prior discourse mention (plus, in the case of first person *I*, presence in the interactional context).

We can discern in this patterning of new information a set of pragmatic constraints parallel to those identified earlier for the distribution of grammatical elements across argument structures. The first I formulate as the *One New Argument Constraint*:

(26) Avoid more than one new core argument

Beyond this pragmatic constraint on *quantity*, there is a further pragmatic constraint on *role*:

TABLE 2.4
New Argument Quantity: Frequency of Clause Types
by Quantity of New Core Arguments[13]

Quantity:	0		1		2		Total	
Language	n	%	n	%	n	%	n	%
Sakapultek	336	(73)	122	(27)	0	(0)	458	(100)
English	463	(87)	72	(13)	0	(0)	535	(100)

(27) Avoid new A

This may be called the *Accessible A Constraint* (a slight revision of what I previously formulated as the *Given A Constraint* [Du Bois, 1987]).

As with intransitives, the transitive argument structure can be seen as providing an architectural framework for information management. But with two available roles, transitives introduce a sharp differentiation of pragmatic function between them. The free O role admits the demands of new information, whereas the constrained A role avoids them.

Ditransitive Pragmatics

The distribution of new versus accessible information in the ditransitive clause follows a by now predictable pattern. The direct object (O) readily accommodates new information whereas the other two core arguments, the transitive subject (A) and indirect object (I), do not.

(28) {RETIREMENT 163}
SAM: .. I gave him **a red pepper.**
           A   I   O

Ditransitive verbs provide a particularly compelling confirmation of the One New Argument Constraint. With as many as three slots available, still only one is exploited for the heavier information management demands.

Cross-Linguistic Evidence for Pragmatic Constraints

The studies summarized in Tables 2.4 and 2.5 show that the previously cited patterns of new information distribution within the clause represent the prevailing ones in other languages as well. Table 2.4 confirms that clauses

---

[13]Sources are as for Table 2.1. Fewer languages are cited because fewer studies report the relevant data on new argument quantity in a comparable format.

TABLE 2.5
New Argument Role: Syntactic Role of New Core Arguments[14]

Role:	A		S		0		Total	
Language	n	%	n	%	n	%	n	%
Hebrew	6	(6)	40	(43)	47	(51)	93	(100)
Sakapultek	6	(6)	58	(55)	42	(40)	106	(101)
English	0	(0)	15	(21)	57	(79)	72	(100)
Spanish	2	(1)	56	(28)	142	(71)	200	(100)
French	0	(0)	75	(34)	143	(66)	218	(100)

with two new core arguments are very rare (or, within these two studies, nonexistent). Table 2.5 shows that the one allowable item of new information appears relatively rarely in A role, whereas no comparable avoidance characterizes the S or O roles.

Summary: A Discourse-and-Grammar Parallel

We have seen that Preferred Argument Structure has two parallel dimensions, a grammatical and a pragmatic. The grammatical dimension is expressed as a soft constraint on the quantity and the grammatical role of full lexical noun phrases relative to argument structures. The pragmatic dimension represents a soft constraint on introducing more than one item of new information (as opposed to accessible information, which suffers no such limitation) within the clause core, plus a constraint on the specific role within which this information may appear. (See Chafe, 1994, pp. 108–119 for a related constraint against more than one new idea per intonation unit.) Full noun phrases and new information both tend to be avoided in the transitive subject position, yet occur freely within intransitive subject and direct object positions. The quantity and role constraints on the distribution of grammatical and pragmatic elements across argument structures together constitute Preferred Argument Structure. These are summarized in Table 2.6.

How do the Quantity and Role constraints relate to each other? We can surmise that general cognitive limitations on information processing amount to an overall constraint on the quantity of new information that can be handled within a single processing unit. What the Role constraints add to this picture is a predictable locus for the heaviest cognitive demands. By segregating high-demand tasks from low-demand tasks and confining the former to a distinct, specified locus within the argument structure, us-

---

[14]Sources, and comments, are as in the note for Table 2.2. Fewer languages are cited because fewer studies report the relevant data on new argument role in a comparable format.

TABLE 2.6
Preferred Argument Structure Constraints: Quantity and Role

	Grammar	Pragmatics
Quantity	Avoid more than one lexical core argument	Avoid more than one new core argument
Role	Avoid lexical A	Avoid new A

ers of grammar are given a predictive advantage as to where they should direct their limited attentional resources.

Low Information Density

I have been emphasizing the challenge of managing the introduction of new information into discourse, and noting the link between such relatively heavy processing tasks and certain grammatical roles within the clause. But discourse is not always so demanding. The fact is, in some kinds of discourse there is relatively little new information to introduce, a situation we may describe as low information density.[15] When this situation obtains there may be no need to use full noun phrases at all to introduce new entities, if each of the entities relevant to what the speaker wants to say has already been introduced into the discourse.

Consider the one-place predicate. Despite its information-carrying privileges, there is no requirement that the S role's full processing capacity be exploited at every instance, whether needed or not. The intransitive S role cannot be reliably predicted to bear a lexical noun phrase, nor the pragmatic statuses that go with it. Often enough, the S role will be filled by a personal pronoun. In the following examples, each pronoun corresponds to an accessible referent previously introduced:

(29) {LUTHER 471}
FOSTER: He wavers.
             ‾S‾

(30) {CUTIEPIE 919}
JILL: And we started laughing,
              ‾‾

(31) {RAGING 365}
SHARON: she gets real embarrassed,
          ‾‾‾

---

[15]In Du Bois (1987) I spoke of low "information pressure." I now prefer the more neutral term "information density," modeled on Durie's congenial phrases "referential density" and "lexical density," which denote similar discourse measures (Durie, in press).

If the event that is being described calls unequivocally for a semantically one-place predicate (*waver, laugh, get embarrassed*), the S role will routinely be called into play regardless of whether there is a need for it to carry any new information. Verbs and their argument structures, like so many elements in grammar, are multifunctional: they are capable of serving both semantic and pragmatic functions. Despite the focus in this study on the most demanding aspects of information management, we should not be surprised to see that sometimes a particular argument structure (e.g., that of a one-place predicate like *waver* or *laugh*) is selected for its semantic function alone, even when its full information-carrying potential is not needed.

The same holds for transitive verbs. Whereas the O role is free to accommodate new information, it is perfectly compatible with accessible information as well. There is no *lower* limit on the number of new entities to be introduced within a transitive clause. In the following examples, all argument positions are filled by pronouns for the simple reason that all the referents are accessible (due to prior discourse mention plus the situational context):

(32) {CUTIEPIE 969}
     JEFF: .. (H) she didn't see me.
              ‾A‾   ‾O‾

(33) {RAGING 218}
     CAROLYN: Oh I believe it.
                   ‾   ‾

At such times it may happen that the only really new information carried in a clause is that expressed in the verb (*see, believe*) or in another element like a negative (*didn't*). Sometimes even the verb is already known and the only thing remaining to be specified is who did what to whom. In the next example, the concept of *kiss* as well as the kissing event's two participants have all been previously introduced. What remains is the selection of which referent goes into the subject role and which into the object, and the past-tense timing of the event itself:

(34) {HOWARDS 1124}
     LORI: He kissed her,
             ‾A‾   ‾O‾

Even ditransitive verbs entail no lower limit on the introduction of new nominal information. When pragmatically appropriate, all three argument roles may contain accessible information expressed by pronouns, as in the following example:

(35) {RETIREMENT 880}
     DORIS: You've told us that.
              ‾A‾   I O

Clearly, speakers have good reasons to utter clauses that contain very little new information, at least of the sort encoded in lexical noun phrases.

Under conditions of low information density it will matter little whether the verb has one, two, or three core arguments available. We are likely to find that none of the argument roles is filled with a lexical noun phrase. It is not uncommon, and indeed entirely natural, for both the A and the O roles of a transitive clause, or the A, I, and O roles of a ditransitive clause, to be realized with pronouns.

It is important to point out that such all-pronoun clauses do *not* constitute violations of any of the Preferred Argument Structure constraints. These constraints posit only *upper* limits on the introduction of new and lexical information within the clause. True, if low density discourse was the norm, it would be hard to justify the constraints empirically, since it is only in high density discourse that the constraints become fully evident. As the density of new introductions increases, available S and O roles are more and more filled up with new lexical noun phrases. Meanwhile, it becomes more noticeable that the A role is not contributing much to managing the influx of new information. In contrast, low density discourse neither supports nor refutes the constraints.

But there is one important lesson to be learned from low density discourse. It reminds us that even a relatively frequent association of a particular discourse function with a particular grammatical role need not rise to the level of predictiveness we expect of a linguistic sign. Despite their critical contribution to the management of new information, neither the S nor the O role can be said to *stand for* new information. No matter how much linguists love signs, there is simply no justification for a Saussurean sign function here, by which a form (say, the O role) symbolizes a discourse-pragmatic function (say, "new information," or even "lexical noun phrase"). The all-pronoun clauses, which are by no means aberrant or exceptional, preclude this. Rather, what the S and O roles do is to enable or facilitate. They reserve a place in a structural configuration that allows, but does not require, the performing of demanding processing functions. In my terms, the S and O roles lack a sign function, but do fulfill a structure function: They provide a predictable locus for unpredictable work.

THE IMPORTANCE OF BEING STRUCTURED

Although the discourse pattern identified here is not reducible to a rule of grammar, neither is it without structure. We are not dealing with some crude constraint on overall quantity of nouns or items of new information,

but rather a well-articulated preference that is systematically sensitive to the specific syntactic structuring of the surface clause. In particular, all the constraints identified apply precisely to *core argument* roles of the clause (subject, object, indirect object), and not to just any nouns in a clause—not even to nouns embedded within the core argument nouns themselves.

For example, prepositional phrases contain noun phrases, but insofar as these represent oblique rather than core roles within the clause, they are not governed by the accounting of quantity applicable to core arguments. Hence the oblique full noun phrases may occur freely, without heed of the limit that core arguments are held to:

(36) {DEADLY 1903}
KEN: (TSK) (H) <u>He</u> drops **the goldfish** into **the tank,**
                A   O   Obl

(37) {BANK 342}
FRED: and then <u>he</u> borrowed **some money** from **his uncle,**

Here the prepositional objects *the tank* and *his uncle* happily coexist in the same clause with the direct objects *the goldfish* and *some money*, respectively. Only the direct objects are core arguments subject to constraint. In fact, it is not uncommon to see multiple oblique roles, each non-core and each bearing a full lexical noun, without any violation of the Preferred Argument Structure constraints.

(38) {DEADLY 894}
LENORE: <u>0</u> take uh **one of those a day** on **an empty stomach,**
           A   O   Obl   Obl

Here we see three full noun phrases within one clause—and within one intonation unit—but two of these nominals (*a day* and *an empty stomach*) are in oblique roles, and hence not subject to constraint. Speakers have no trouble verbalizing all three noun phrases within a single intonation unit, even if more than one contains new information.

Bringing obliques into the overall picture of information management introduces some new complexities, which are mostly beyond the scope of this article. (Compare Table 2.7, which includes all lexical mentions, with Table 2.2, which includes only those in core argument roles.) I merely mention here that obliques, by falling outside the clause core and hence outside the scope of the Preferred Argument Structure constraints, represent another prime opportunity, beyond the S and O roles, for the introduction of new information. But the information introduced obliquely tends to be of a substantially different character, less topical and more ephemeral (Du Bois, 1980; Thompson, 1997), as the examples in this section suggest.

TABLE 2.7
Lexical Argument Role: Syntactic Role of All Lexical Mentions

Role:	A		S		O		Oblique		Other		Total	
Language	n	%	n	%	n	%	n	%	n	%	n	%
Sakapultek	11	(3)	126	(33)	81	(21)	118	(31)	48	(12)	384	(100)
English	21	(5)	90	(23)	146	(37)	120	(31)	16	(4)	393	(100)

Another piece of evidence that the Preferred Argument Structure constraints are sensitive to specific syntactic structure concerns internally structured noun phrases:

(39) {RISK 98}
JENNIFER: Yeah <u>you</u> get **a percentage** of **the amount** of **countries**
                  A   O
you own,

Although this clause could be said to contain at least three lexical *noun phrases* (e.g., *a percentage, the amount,* and *countries you own*), a more perspicuous analysis would discern just one lexical *argument*, namely the complex noun phrase *a percentage of the amount of countries you own,* which as a unit fills the role of direct object of the verb *get.* The noun phrases *the amount* and *countries you own* do not fill verbal argument roles of their own, but are subsumed under the direct object role within the complex noun phrase that has *percentage* as its head.

We conclude that the internal syntactic structuring of the clause has precise consequences for the cognitive processing of lexical information. It will not do to formulate a syntactically naive or agnostic constraint like "Avoid more than one lexical noun phrase in an intonation unit" or even "Avoid more than one lexical noun phrase in a clause." Such claims are readily refuted by the pervasiveness in discourse of configurations like those cited in this section. The constraints on distribution of referential form are sensitive to a finer specification of syntactic structure, specifically to the argument structure of the surface clause, rather than to a gross unit-as-container organization, whether the unit proposed is the clause or the intonation unit. Similarly for information management, which is equally sensitive to surface syntactic configuration of the clause.

CONSTRAINT AND STRATEGY

Putting narrow limits on how much new information may be introduced within a clause, and where it may go, poses a certain puzzle. How do speakers manage to say what they need to say? Particularly confining, it seems,

would be the constraint against introducing new information in the transitive subject role. Humans assuredly like to talk about humans. Often enough what attracts our interest is a human engaged in agentive action. When it comes time to verbalize such a content, we find that this would normally entail the human referent taking the role of agent of an active verb, hence implicating the transitive subject role. Must we forgo expressing such meanings due to obstacles thrown up by Preferred Argument Structure? Not at all. The way to fulfillment is via a simple principle of discourse: Speakers need not say everything in one clause. Facing cognitive constraints that could frustrate their expressive goals, speakers can simply mobilize their planning capacity to organize a series of successive clauses. In the following example, the narrator relates how a psychotherapist told him something interesting about herself. This state of affairs is most naturally conceived in terms of the psychotherapist playing the role of semantic agent of the telling event, which in verbal form would tend to place the *tell-er* in the role of syntactic subject of the verb *tell*. And there's the rub. To introduce the psychotherapist as new information in (di-)transitive subject role would be problematic, if we take seriously the constraints thus far identified. Fortunately there is another way:

(40) {LAMBADA 1284}
   MILES: . . I meet **this psychotherapist.**
             A   O
          . . . who tells me she's addicted to **this dance.**
           A   O S   Obl

The verb *meet* nicely fills the bill of taking the pressure off, by breaking down the task at hand into a sequence of manageable subtasks. First, *meet* takes the newly introduced psychotherapist as its direct object—an unconstrained role—and so deftly evades the strictures of the Accessible A Constraint. Then, once the new referent has been properly introduced, it immediately becomes eligible to fill the role of subject of *tell*, because by now it already constitutes accessible information, expressible by a pronoun (*who, she*).

Speakers show a certain ingenuity in developing strategies for maintaining a full expressive range of propositional semantics while accommodating constraints on discourse pragmatics. Many of these strategies appear to be routinized, memorized for reuse—even grammaticized as specialized constructions in the language. Like *meet*, the transitive verb *see* is handy in that it readily takes a human referent in O, thus supplying a free role that allows unconstrained introduction of new referents. Observe how *this girl* is introduced in the next example:

(41) {CONCEPT 1315}
   MARILYN: . . we're pulling up,
             . . and I see **this gir=l.**
              A   O
            Who I'd never seen before,
            0 sort of d=art out of **our driveway.**
            ((4 LINES OMITTED))
            and 0 watch us pull in,

I suggest that one of the things speakers know about the verb *see*, as about the verb *meet*, is that it fulfills a strategic function of introducing a new human referent without running afoul of the Accessible A Constraint. There may be semantic motivations for using the verb *see* here, but most of our reported experiences involve seeing, and yet we don't feel compelled to continually verbalize this information explicitly in every instance. I would argue that the choice to verbalize a verb like *see* (or *meet*) that takes humans in O role is often motivated more by information management factors than by semantic ones.

This motivation surely holds for certain intransitive predicates, whose bleached semantic content leaves little to motivate their use *except* for information management. There are numerous strategies that exploit the potential of the S role of intransitive and other one-place predicates. The key is not intransitive verb status per se, but simply avoiding transitive (and ditransitive) subjects. "Subjects" are perfectly compatible, it turns out, with the high cost of introducing new information—but only if the subject is that of a one-place, rather than a two- or three-place, predicate.

One class of intransitive verbs that has strategic significance beyond its propositional semantics is that of verbs expressing "existence or appearance on the scene" (Du Bois, 1987, pp. 830ff; Firbas, 1966; Kuno, 1972, p. 319), which are often called on to introduce new information. One much-analyzed construction is the so-called "existential" *there*-construction. This certainly qualifies as being bleached of semantic meaning, and yet it is very much in demand for its discourse-pragmatic properties. Note that introducing a new referent via *there*, as in the next example, successfully evades the constrained transitive subject position.

(42) {CUZ 400}
   ALINA: . . . and there's **a car** in front of me,

Once this new car is introduced into the discourse, it allows definite mention of its driver, first as *the guy* and then *he*:

(43) {CUZ 399}
    ALINA: (H) So I'm driving up to **the house,**
           . . . and there's **a car** in front of <u>me,</u>
           and **the guy** is just like sitting there,
           <VOX> in **the middle of the roa=d,**
           and <u>he</u>'s not moving,
           and,
           . . <u>you</u> know <u>I</u> wanna park **the car** </VOX>.

Similarly, an intransitive verb like *come,* semantically denoting motion toward the speaker, frequently fulfills the additional pragmatic function of providing a framework for introducing a new referent "appearing on the scene." The discourse-pragmatic function may even come to dominate over the semantic one. Consider the example presented earlier, repeated here, in which the first-mention heavy noun phrase *this new wave of people* serves to introduce a new referent into the discourse (after which it becomes the topic of subsequent discourse):

(44) {CUZ 995}
    ALINA: . . **this new wa=ve of people** comes in.

The choice of an intransitive verb for the introduction is no accident. Beyond the specific semantics of *come in,* the verb's intransitivity has the advantage of defining a single core argument position, which, lacking competitors, is unconstrained. Verbs like *come* are useful in part because they are free to welcome full lexical noun phrases of all shapes and sizes, along with the cognitive cost of introducing new information.

PATTERN WITHOUT RULE

We noted earlier that presenting accessible information where new is allowed does not constitute a violation of any Preferred Argument Structure constraint. But there are real exceptions, as Tables 2.1–2.5 document and the examples in this section will portray. Exceptions are not always a bad thing. In this case, they make it clear that despite its systematicity, Preferred Argument Structure cannot be reduced to a grammatical rule. It must remain within the domain of discourse, as a patterning *of* grammar with consequences *for* grammar.

It is noteworthy that when a departure from Preferred Argument Structure does occur in natural discourse, the resulting utterance bears not a hint of ungrammaticality. For example, speakers sometimes do put a full

lexical noun phrase in transitive subject position, with no ill consequences for grammaticality:

(45) {CUZ 1282}
    ALINA: **turtlenecks** don't hide <u>everything.</u>
             A   O

The A role sometimes even accommodates new information not previously introduced, as in this first mention of (generic) *turtlenecks.* In the next example, the A role *cats* is a first mention in the discourse, also generic, while the O role pronoun *those* represents an accessible referent (some vitamin pills that are visible in the situational context):

(46) {CUZ 688}
    LENORE: **Cats** love <u>those.</u>
              A   O

Sometimes a transitive clause will contain even two full noun phrases, in both A and O roles, thus violating both of the Preferred Argument Structure constraints on lexical mentions.

(47) {LAMBADA 78}
    HAROLD: . . **little kids** usually don't break **their legs** anyway.
              A   O

Each of these two noun phrases expresses a referent that is accessible in discourse context (there was prior talk of a specific child breaking his leg), so this particular example may not present a significant violation of the *pragmatic* side of Preferred Argument Structure. But on occasion even new information full lexical nouns appear in both A and O roles. In the following example, each of the two core argument noun phrases represents a first mention in the discourse.

(48) {FEAR 580}
    WALT: (H) **trailer truck** hits **his brakes,**
            A   O

To be sure, one might seek extenuating circumstances for all of these cases,[16] and it must be said that in various ways the offending noun phrases

---

[16]For example, none of the "exceptional" A role lexical mentions cited in this section contains an article or other determiner, raising the question of just how "full" these full noun phrases really are. This points to the important issue of a continuous scale of accessibility (Ariel, 1990, 2001), which is beyond the scope of this chapter. See also Chafe (1994, pp. 108–119, 1998, p. 109) for exegesis of some applicable mitigating factors.

do not elicit the heaviest cognitive demands. But for the present it seems prudent to conclude that the discourse preferences so far identified constitute *soft* constraints: They may be violated without precipitating either ungrammaticality or processing failure.

It should be no surprise that the constraints are soft, to the extent that the discourse preferences are cognitively based. It would be risky to operate always at the outer limits of cognitive capacity. Better to set a routine limit lower than the maximum; under special circumstances one may then momentarily exceed this flexible limitation.

Of course, if exceptions like these were frequent, there would be no reason to posit a preferred argument structure in the first place. But in all they are relatively rare, as is attested by the evidence from a number of languages (cf. Tables 2.1–2.5). Even infrequent occurrences, however, are sufficient to preclude a categorical rule of grammar. Whereas models that depend entirely on categorical rules might take such an outcome as a failure to be regretted, from a discourse perspective the interpretation is quite different. We see it as noteworthy that a pattern of behavior so consistent across speakers and across languages can emerge in discourse and be sustained in the absence of any mechanical rule. And we go on to seek for deeper explanations.

The exceptions to the Accessible A Constraint, however rare, have another important theoretical consequence. They argue against interpreting the A role as *signifying* "accessible information." We cannot justify a sign function here, because new A's are encountered regularly, if not very frequently. (With a true sign function like the word *white*, we don't find that 5% of the time speakers mistakenly use it to mean *black*.) Even less plausible is treating S or O as a sign. As we saw earlier, half or more of the S and O positions may contain accessible information, making it impossible to reliably interpret S or O as signifying "new information." I say this despite the fact that S and O together clearly constitute *the* place to put new information within the clause core: If you have new information to introduce, that's where you nearly always put it. But the implication in the opposite direction does not hold, because S and O do not even come close to consistently predicting the presence of new information. The Saussurean construct of the sign function is of no help in our attempt to model the discourse phenomena underlying Preferred Argument Structure.

THE GRAMMATICAL ARCHITECTURE OF COGNITIVE COST

If Preferred Argument Structure is not a rule of grammar and not a sign function, what can it be?

We began by identifying a curiously skewed patterning of full versus reduced noun phrases within the grammatical frame articulated by argument structures in discourse. We went on to observe a parallel patterning of new versus accessible information. But if we probe still further, we soon see that these discourse patterns are not ultimately about noun phrases, nor even new information, but about cognitive cost. Specifically, they point to a systematic exploitation of syntactic structure as a frame for organizing and managing cognitive costs in speech production and understanding. Some aspects of speech processing invoke high cognitive cost, whereas others are relatively undemanding (Ariel, 1990, 2001; Chafe, 1994, pp. 71–81). Other things being equal, new referents are costly, accessible referents are cheap. But there are other costs as well, like those associated with processing relative clauses—which have been shown to follow a Preferred Argument Structure pattern, preferring S or O role over the constrained A role (Fox, 1987; cf. Fox & Thompson, 1990). The distribution of cognitive costs across the grammatical architecture of the clause is neither random nor constant, but systematically skewed. Speakers know where in a clause to produce, and hearers where to expect, the heavy processing demands such as those associated with the introduction of new information. Rather than leave this to chance, the role constraints of Preferred Argument Structure effectively enlist the syntactic structure of the clause to provide a consistent shape within which the more demanding tasks can be carried out: a predictable locus for unpredictable work. In this sense the surface grammatical structure of a clause, in particular its argument structure, can be seen as defining an architecture of cognitive cost, or more precisely, an architecture *for* cognitive processing, in which certain locales are predictably specialized for high- or low-cost work.

The verbs of a language can be thought of as a diverse population of semantic-pragmatic-grammatical elements, each offering an argument structure capable of managing some configuration of cognitive cost linked to a meaning frame. One-place predicates like intransitive verbs can handle their cognitive costs without special restrictions because in their simplicity they do not attempt to juggle more than one core argument slot for (noun-based) cognitive processing. Transitive verbs push the limits by introducing two slots, but at the cost of introducing constraints that limit the carrying capacity of one of them. Elaboration of still more complex structures, such as three-place ditransitive clauses, causatives, and certain complex clause structures, is attainable only through the trade-offs of historical evolution, that is, via the process of grammaticization. Grammaticization as an adaptive process represents the crystallization of compromise between competing motivations (Du Bois, 1985; see also Bates & MacWhinney, 1982). Due to limitations within a compact domain like the clause core or the intonation unit, new opportunities for semantic expression may be added only through the imposition of limits on old ones. Grammar is responsive to re-

current patterning in the aggregate of language use tokens, as it constitutes an adaptive architectural framework for cognitive function. Once crystallized as grammatical structure, the architecture becomes a cognitive resource available to all members of the speech community.

CONCLUSION

Roman Jakobson once quipped that "Grammar without meaning is meaningless" (1990, p. 332). In the context of his time this could be heard as a pointed critique of an approach that radically severed grammar from meaning. Since then a number of functional theorists have sought to restore meaning to a central place in language, offering a picture of grammar tightly integrated with it (Chafe, 1970; Fillmore, 1977, this volume; Langacker, 1987, 1998). If we wish to carry this understanding to its fullest development, then discourse pragmatics, too, will have to be accorded its distinctive place in the emerging grammatical synthesis.

It should be obvious by now that we do not accept the dichotomies between discourse and grammar that were recited at the outset of this chapter. More precisely, we do not accept the conclusion sometimes drawn from them, that the gulf between grammar and discourse is unbridgeable. Not only can apposite theory create productive links between discourse and grammar, but what at first glance seemed to be the freer, wilder, less constrained half of the equation—what we call spontaneous discourse—proves to embody some of the most pervasive, regular, profound, and cross-linguistically stable trends. And these well-grounded patterns in discourse have the power to shape the very foundations of grammar (Du Bois, 1987).

And yet the idea that syntax remains partly aloof from meaning retains a certain appeal. Could grammar, by showing some resistance to being directly semanticized, contribute something more than just another means of symbolizing what is meant? Could the lack of specific meaning free syntax to serve another role, more abstractly framing and organizing a crucial, if neglected, dimension of language use: its actual process of coming into being?

In the idealized cognitive world assumed in some versions of functional linguistics, there is no friction. A concept conceived is a concept verbalized, instantly and effortlessly rendered into words. But where is the process? In the world of natural spontaneous discourse, there is audible friction. Thought does not go gently into words. Wheels grind as we strive to speak, working to manage all the tasks that confront us at once. And some parts of the work are more demanding than others. Initializing a new cognitive file for a just-introduced referent takes more resources than updating an existing cognitive file for a highly accessible referent. A mix of easy and demand-

ing tasks of speech production (or comprehension) are juggled simultaneously, in pursuit of the goal of verbalizing and conveying the idea striven for. All this takes place in real time as we strategically deploy limited cognitive resources. How do such real-time discourse processes impact grammar? The argument structure of a clause, although undoubtedly contributing to the expression of semantic roles like Agent, Patient, or Experiencer, is also called on to serve the demands of information management. Managing information and expressing a full range of propositional meanings are two functions that must be strategically integrated into the production of a single structured utterance. Thus discourse pragmatics takes its place alongside semantics as a driving force in language use and grammaticization, ultimately shaping the most fundamental structures of a language's grammar (Du Bois, 1987).

We recognize that there is no discourse without grammar. There is no raw speaking, nakedly expressing pure speaker intention or discourse function, without the imposition of grammatical category and structure. No pristine primordial world can be found in which discourse function operates on its own in splendid isolation, unhindered by grammatical form, its unique contribution transparently revealed. Perforce we take discourse as it comes, in its grammatical clothing. Conversely, no speaker ever encounters grammar except as it is manifested in discourse. For better or worse, discourse and grammar are inextricably linked. So it is that within discourse we analyze tangible grammar (noun phrases and their roles of subject, object, oblique) and, tipped off by the cross-linguistic patterns identified by grammatical typology (Comrie, this volume), we take the trouble to distinguish between subjects of transitive verbs and subjects of intransitive verbs—because we know that many languages treat this discrimination as fundamental even if English and most European languages don't. Simultaneously we note the discourse-pragmatic functions and cognitive processes realized via each nominal reference token (new vs. accessible information, high vs. low processing cost). The interwoven strands of data are tallied in tandem as we look for correlations between grammatical structure and discourse-pragmatic function. This program of discourse research probes into syntactic structures within the sentence, often at the clause level or smaller. Against the stereotype that the discourse analyst's proper domain lies beyond the sentence, I call this "discourse inside the clause." Of course what's happening inside the clause isn't really a world unto itself. Its discourse patterns have larger origins. The point is that there is massive interpenetration extending from sentence-internal structures at the level of noun phrase and clause up to the scale of the larger discourse units and the longest threads of referential continuity. The analysis crosses the lines between small and large, structure and function, token and type. The ultimate pay-

off is in understanding and explaining grammar. But the payoff doesn't come without a commitment to pursue grammatical and discourse-pragmatic patterns within an integrated frame of cross-disciplinary inquiry.

Surprises emerge from this approach to discourse and grammar. In principle, what the rules of grammars do not prohibit they should allow. Yet for all the freedom that their grammars afford them, speakers travel a straight and narrow path. Consider the free-willed speaker, who on each occasion of utterance is graced with the liberty—under the rules of grammar—to introduce a full noun phrase (perhaps one expressing an item of new information) into each and every argument position in a clause. Overwhelmingly, the speaker neglects this grammatical potential. Moreover the unused options are not random but consistent in their grammatical patterning. One specific syntactic position is avoided for new information while others are favored, even though none of this is demanded by grammatical rule. We are led to conclude that something is at work that goes beyond grammar. When we discover the same discourse regularity without grammatical obligation recurring in the spontaneous speech of different speakers, talking in diverse contexts about varied topics, and when we find this same pattern recurring across languages of distant regions, independent histories, and radically divergent grammatical types, then we have grounds to invoke a universal of discourse more broadly grounded, more stable, and more empirically confirmable than many that have been claimed for grammar. Such is the discourse universal of Preferred Argument Structure. It is a recurrent pattern of language use that cannot be reduced to grammatical rule. It stands in its own right as a generalization about discourse, one that involves grammar, in that it is defined over grammatical categories, and yet is not part of grammar.

If grammar is assumed to be a functioning part of the total system of language, it is necessary to ask what role it plays. The theory of discourse and grammar I have been advancing points to the conclusion that structure is functional. Meaningless grammar need not mean functionless grammar. Sometimes grammar serves function by refusing to mean. By decoupling from signification and escaping the narrow role of a specific sign function, core grammatical roles like A, S, and O are freed instead to support a broader structure function. What grammar contributes instead of meaning is an architectural framework within which cognitive processing is realized. In place of the sign function with its form–meaning pairing that well serves so much of language, grammar sometimes offers nothing more nor less than a structure function. If the emerging picture of language reveals a more complex relationship between structure and function than many might have wished for, it is nevertheless one that is more in keeping with that complexity which is now more and more recognized as characteristic of all evolved forms of life and culture.

REFERENCES

Aissen, J. (1987). *Tzotzil clause structure.* Dordrecht, Netherlands: Reidel.

Ariel, M. (1990). *Accessing noun-phrase antecedents.* London: Routledge.

Ariel, M. (1998). Cognitive universals and linguistic conventions: The case of resumptive pronouns. *Studies in Language, 23,* 217–269.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 29–87). Amsterdam: Benjamins.

Ashby, W. J., & Bentivoglio, P. (1993). Preferred Argument Structure in spoken French and Spanish. *Language Variation and Change, 5,* 61–76.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). Cambridge: Cambridge University Press.

Boas, F. (1911). Introduction. *Handbook of American Indian languages, Part 1* (pp. 1–83). Bureau of American Ethnology, Bulletin 40. Washington, DC: Government Printing Office.

Chafe, W. L. (1970). *Meaning and the structure of language.* Chicago: University of Chicago Press.

Chafe, W. L. (Ed.). (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production.* Norwood, NJ: Ablex.

Chafe, W. L. (1987). Cognitive constraints on information flow. In R. S. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 21–51). Amsterdam: Benjamins.

Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing.* Chicago: University of Chicago Press.

Chafe, W. L. (1998). Language and the flow of thought. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 93–111). Mahwah, NJ: Lawrence Erlbaum Associates.

Clancy, P. M. (1980). Referential choice in English and Japanese narrative discourse. In W. L. Chafe (Ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 127–202). Norwood, NJ: Ablex.

Clancy, P. M. (in press). The lexicon in interaction: Developmental origins of Preferred Argument Structure in Korean. In J. W. Du Bois, L. E. Kumpf, & W. J. Ashby (Eds.), *Preferred Argument Structure: Grammar as architecture for function.* Amsterdam: Benjamins.

Clark, H. H. (1996). *Using language.* Cambridge: Cambridge University Press.

Du Bois, J. W. (1980). Beyond definiteness: The trace of identity in discourse. In W. L. Chafe (Ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 203–274). Norwood, NJ: Ablex.

Du Bois, J. W. (1985). Competing motivations. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 343–365). Amsterdam: Benjamins.

Du Bois, J. W. (1986). Self-evidence and ritual speech. In W. L. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 313–336). Norwood, NJ: Ablex.

Du Bois, J. W. (1987). The discourse basis of ergativity. *Language, 63,* 805–855.

Du Bois, J. W. (2000). *Santa Barbara corpus of spoken American English, Part 1* [3 CD-ROMs]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Du Bois, J. W., Kumpf, L. E., & Ashby, W. J. (Eds.). (in press). *Preferred Argument Structure: Grammar as architecture for function.* Amsterdam: Benjamins.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45–89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Du Bois, J. W., Schuetze-Coburn, S., Paolino, D., & Cumming, S. (1992). *Discourse transcription. Santa Barbara papers in linguistics* (Vol. 4). Santa Barbara: Department of Linguistics, University of California, Santa Barbara.

Durie, M. (in press). New light on information pressure: Information conduits, escape valves, and role alignment stretching. In J. W. Du Bois, L. E. Kumpf, & W. J. Ashby (Eds.), *Preferred Argument Structure: Grammar as architecture for function.* Amsterdam: Benjamins.

Dutra, R. (1987). The hybrid S category in Brazilian Portuguese: Some implications for word order. *Studies in Language, 11,* 163–180.

Fillmore, C. J. (1977). Topics in lexical semantics. In R. W. Cole (Ed.), *Current issues in linguistic theory* (pp. 76–138). Bloomington: Indiana University Press.

Firbas, J. (1966). Non-thematic sentences in contemporary English. *Travaux linguistiques de Prague, 2,* 239–256.

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 134–184). Cambridge: Cambridge University Press.

Fox, B. (1987). The noun phrase accessibility hierarchy reinterpreted: Subject primacy or the absolutive hypothesis? *Language, 63,* 856–870.

Fox, B., & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language, 66,* 297–316.

Givon, T. (1979). *On understanding grammar.* New York: Academic Press.

Givon, T. (Ed.). (1983). *Topic continuity in discourse: A quantitative cross-language study.* Amsterdam: Benjamins.

Givon, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics, 30,* 5–55.

Givon, T. (1998). The functional approach to grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 41–66). Mahwah, NJ: Lawrence Erlbaum Associates.

Goldberg, A. E. (1998). Patterns of experience in patterns of language. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 203–219). Mahwah, NJ: Lawrence Erlbaum Associates.

Harris, Z. S. (1951). *Methods in structural linguistics.* Chicago: University of Chicago Press.

Hopper, P. J. (1998). Emergent grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 155–175). Mahwah, NJ: Lawrence Erlbaum Associates.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language, 56,* 251–299.

Jakobson, R. (1990). *On language.* Cambridge, MA: Harvard University Press.

Kumagai, Y. (2001). Preferred Argument Structure and discourse ergativity in English. *Mulberry, 50,* 77–90.

Kuno, S. (1972). Functional Sentence Perspective: A case study from Japanese and English. *Linguistic Inquiry, 3,* 269–320.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites.* Stanford, CA: Stanford University Press.

Langacker, R. W. (1998). Conceptualization, symbolization, and grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 1–39). Mahwah, NJ: Lawrence Erlbaum Associates.

Matsumoto, K. (1997). NPs in Japanese conversation. *Pragmatics, 7,* 163–181.

McGregor, W. B. (1999). "Optional" ergative marking in Gooniyandi revisited: Implications to the theory of marking. *Leuvanse Bijdragen, 87,* 491–534.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

Nichols, J., & Timberlake, A. (1991). Grammaticalization as retextualization. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization: Vol. 1. Focus on theoretical and methodological issues* (pp. 129–146). Amsterdam: Benjamins.

Payne, D. L. (1987). Information structuring in Papago narrative discourse. *Language, 63,* 783–804.

Saussure, F. (1916). *Cours de linguistique generale* [*Course in general linguistics*]. Paris: Payot.

Schiffrin, D. (1987). *Discourse markers.* Cambridge: Cambridge University Press.

Schuetze-Coburn, S. (1987). *Topic management and the lexicon: A discourse profile of three-argument verbs in German.* Unpublished master's thesis, University of California, Los Angeles.

Smith, W. (1996). Spoken narrative and preferred clause structure: Evidence from modern Hebrew discourse. *Studies in Language, 20,* 163–189.

Thompson, S. A. (1997). Discourse motivations for the core–oblique distinction as a language universal. In A. Kamio (Ed.), *Directions in functional linguistics* (pp. 59–82). Amsterdam: Benjamins.

Thompson, S. A., & Mulac, A. (1991). A quantitative perspective on the grammaticization of epistemic parentheticals in English. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization: Vol. 2. Focus on type of grammatical markers* (pp. 313–329). Amsterdam: Benjamins.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development.* Cambridge: Cambridge University Press.