

## Thematic Proto-Roles and Argument Selection

David Dowty

*Ohio State University*

As a novel attack on the perennially vexing questions of the theoretical status of thematic roles and the inventory of possible roles, this paper defends a strategy of basing accounts of roles on more unified domains of linguistic data than have been used in the past to motivate roles, addressing in particular the problem of *argument selection* (principles determining which roles are associated with which grammatical relations). It is concluded that the best theory for describing this domain is not a traditional system of discrete roles (Agent, Patient, Source, etc.) but a theory in which the only roles are two cluster-concepts called *Proto-Agent* and *Proto-Patient*, each characterized by a set of verbal entailments: an argument of a verb may bear either of the two proto-roles (or both) to varying degrees, according to the number of entailments of each kind the verb gives it. Both fine-grained and course-grained classes of verbal arguments (corresponding to traditional thematic roles and other classes as well) follow automatically, as do desired "role hierarchies". By examining occurrences of the "same" verb with different argument configurations (e.g. two forms of psych predicates and object-oblique alternations as in the familiar *spray/load* class), it can also be argued that proto-roles act as defaults in the learning of lexical meanings. Are Proto-Role categories manifested elsewhere in language or as cognitive categories? If so, they might be a means of making grammar acquisition easier for the child, might explain certain other typological and acquisitional observations, and may lead to an account of contrasts between unaccusative and unergative intransitive verbs that does not rely on deriving unaccusatives from underlying direct objects.

### 1. Introduction

There is perhaps no concept in modern syntactic and semantic theory which is so often involved in so wide a range of contexts, but on which there is so little agreement as to the nature and definition of the concept, as *thematic role* (or *thematic relation*) and its derivative, *theta-role* in Government-Binding Theory (GB). In addition to the "argument-indexing function in GB (see below), thematic roles have been invoked in the statement of multifarious syntactic generalizations in that and in other syntactic theories, and the existence of thematic roles is so taken for granted that psycholinguists now attempt to study their role in mental processing experimentally (Carlson and Tanenhaus 1988, Stowe 1989), and an introductory text in formal semantics (Chierchia and McConnell-Ginet 1990) offers a technique for formalizing roles while presupposing their necessity in a linguistic theory.

Yet apart from some syntactic correlates of thematic roles, there is in fact a notable absence of consensus as to what thematic roles are (cf. below). At best, they are obviously creatures of the syntax-semantics interface, and thus require a sound semantic theoretical basis as well as a syntactic one (and mutually consistent

ones) in order to be considered respectable parts of a linguistic theory. But at worst, appeal to them can be a confusion of notions from the syntactic, semantic and pragmatic domains, or a "thinly disguised wild card to meet the exigencies of syntax" (Jackendoff 87:371). Despite the mention of Thematic roles in the Chierchia and McConnell-Ginet text, they have hardly been studied at all in formal semantics<sup>1</sup>, which seems to have gotten by up to this point without any significant purpose for them to serve. Ray Jackendoff, the only semanticist who has studied the concept extensively (non-model-theoretically) and who is ritually cited by syntacticians at their first mention of the notion, has evolved a detailed understanding of thematic roles within his own theoretical framework (Jackendoff 1972, 1976, 1983, 1987) that is clearly quite different from and inconsistent with that of GB or of many current syntacticians (Jackendoff 1987).

Though the term *thematic relation* (later *role*) was introduced by Gruber (1965) and made widely known by Jackendoff (1972), as semantic categories they obviously corresponded to a great extent to the (semantic) *Deep Cases* of Fillmore's contemporaneous Case Grammar (1966, 1968) --- and this in turn harks back to ideas of structuralists such as Frank Blake (1930), and ultimately to Panini's *karakas* --- but Deep Cases played a quite different part within his theory from that of thematic roles for Gruber-Jackendoff or  $\bar{U}$ -Roles do in GB. Chomsky (1981: 35), in introducing  $\bar{U}$ -roles into GB and citing precedents for the idea, claimed thematic roles such as Agent had been primitives of Donald Davidson's event-logic (Davidson 1967a), but he was in error: Davidson did not analyze events in terms of Agent and Patient but in fact rejected Hector Casteñeda's suggestion (Casteñeda 1967) that the Davidsonian event analysis be modified to do so (in Davidson 1967b:125)<sup>2</sup>. Though many linguists seem to assume that linguistic theory should include a finite (and short) language-universal canon of thematic roles (that includes the familiar members such as Agent, Patient, Goal, Source, Theme, Experiencer, Instrumental, etc.), no one that I know of has ever attempted to propose a complete list<sup>3</sup>. There is disagreement even on the most familiar roles (e.g. whether Theme, usually "something that moves or changes state", can be

<sup>1</sup> The exceptions I know of being Chierchia (1984), Carlson (1984), and Dowty (1989), the last discussed below.

<sup>2</sup> That is, what Davidson did propose was that adjuncts (temporal, locative and adverbial modifiers) were predicates of an existentially-qualified event variable in logical form, but subject and object were not: they are traditional "arguments", related to the event variable by the n-place predicate denoted by the verb.

<sup>3</sup> The most comprehensive list I have seen is also the earliest: Blake (1930), who argued that semantically-defined "case relationships" (clearly a similar notion to today's thematic role-types) are "numerous but not infinite; they are not indefinite and subjective, depending on the lucubrations of the individual mind, but objective, definite, and determined once for all by general grammatical principles and the laws of thought", offered as a "pioneer study" an organized system of 87 temporal and locative roles and 26 other roles, including such subsequently ignored roles as *additional* ("he gave him a sum of money *besides the cattle*"), *substitutive* ("he gave me promises *instead of money*") and *similative* ("he barked *like a dog*").

'assigned by' a stative predicate; and whether Theme is the same role as Patient or whether they are distinct), and new candidates for Thematic Roles are being proposed all the time (e.g. *Figure* and *Ground* in Talmy (1985a), *Neutral* in Rozwadowska 1988), *Landmark* in Jackendoff (1982), even *Subject* in Baker 1985). A paper such as this cannot begin to do justice to all the diverse literature on the subject, and a warning to this effect to the reader, plus apologies to the authors who are omitted or overlooked, are hereby issued.<sup>4</sup>

Among the various understandings of thematic roles, we can distinguish two kinds. What I will call the *argument-indexing* view of thematic roles is demanded by the *Ú-Criterion* of GB: each NP argument of a predicate is assigned exactly one *Ú-Role*, and the same *Ú-Role* is not assigned to two NP arguments of the same predicate (Chomsky 1981:36, 139). By clear implication, the *Ú-Roles* that Chomsky originally had in mind to fulfill this criterion were the familiar Agent, Patient, etc. from Gruber, Jackendoff, and others. By virtue of the *Ú-Criterion*, *Ú-roles* served (originally at least) two main purposes in the GB theory, (i) distinguishing "real", semantically contentful arguments of a predicate from dummy arguments such as *it* and *there*, and (ii) helping to keep track of identity and distinctness of NPs as particular semantic arguments of a predicate during the course of a derivation. From the structure of the early Case Grammar theory (Fillmore 1968), it is obvious that Deep Cases also served an argument-indexing function there, since in Deep Structure each NP argument bears exactly one case label (Agentive, Instrumental, Dative, Locative, etc.), and subsequent syntactic transformations are stated in terms of these labels, not arbitrary or tree-structurally positioned NPs, and this seems to presuppose that there is not more than one Agentive (etc.) NP per clause. (This was modified later: cf. below.)

In order for such systems to work, in an account in which the roles Agent, Theme, Goal, etc. are given explicit semantic content, the meanings of all natural language predicates must turn out to be of a very particular sort: for every verb in the language, what the verb semantically entails about each of its arguments must permit us to assign the argument, clearly and definitely, to some official thematic role or other--it cannot be permitted to hover over two roles, or seem to "fall in the cracks" between roles---and what the meaning entails about every argument must always be distinct enough that two arguments clearly not fall under the same role definition. This is a very strong empirical claim about possible meanings of natural language predicates, and, as soon as we begin to try to be precise about exactly what Agent, Patient, etc. "mean", all too subject to

<sup>4</sup> In this paper I have tried to follow the practice of citing papers in which, in my view, the essence of a proposal or insight was first made, but not necessarily later discussions of the insight (under the same or different terminology) unless I believe they contributed something new that is relevant here. Hence relatively more references are made to early literature by Fillmore, Jackendoff and their contemporaries, and relatively fewer references to recent literature on roles, than is sometimes found elsewhere.

difficulties and apparent counterexamples.

Doubts as to whether the familiar short lists of Roles/Deep Cases (or refinements thereof) would ever really work this way already arose in the Case Grammar days (e.g. Cruse 1973, Huddleston 1970, Mellema 1974, Fillmore 1971a, and many others). Later GB writers saw the danger too, and proposed to circumvent the problem by refraining from committing themselves to the traditional roles, what I will henceforth call *thematic role types*, and instead invoked *individual thematic roles*---these terms from (Dowty 1989)<sup>5</sup>. That is, we simply call the thematic role of the subject of the verb *hit* the "hitter role", that of the subject of *kill* the "killer role", of *build* the "builder role", and so on, with no assumption made that there is one thematic role type common to these arguments (Marantz 1984, van Riemsdijk and Williams 1986)---though the possibility that role types also exist need not be ruled out, either. Trivially then, there will be enough distinct  $\bar{U}$ -roles around (i.e., the individual roles) to permit the  $\bar{U}$ -Criterion to be satisfied, i.e. to preserve the argument-indexing view of thematic roles.

Nevertheless, many syntacticians working within the GB framework have continued to appeal to the traditional thematic role types to state syntactic generalizations (Rappaport and Levin 1988, Nishigauchi 1984, Belletti and Rizzi 1986, etc.). And appeal to a particular *hierarchy* of thematic roles, such as Nishigauchi (1984) does in stating control principles (i.e. Source >... ), requires ALL arguments of predicates (at least those that ever occur in control relationships), to have roles mentioned in the hierarchy, i.e. a role-type, not an individual role. Thus for such hypotheses, it IS a crucial question whether there is a small set of distinguishable role-types that effectively index all arguments.

In contrast to the argument-indexing view, Jackendoff's research on thematic roles is of a fundamentally different kind. For him, thematic relations (the term he prefers to *roles*) are most importantly notions of conceptual structure, as elucidated in Jackendoff (1983, 1987), rather than basically syntactic or interface notions; they are not theoretical primitives but are defined by particular configurations of primitive operators such as GO, STAY and CAUSE in conceptual structure; one discovers their nature and distribution empirically by looking at certain lexical and syntactic patterns in natural language in relation to their meanings, e.g. the distribution of prepositions in particular (though not, perhaps surprisingly, by psychological experiment). And the thematic roles one finds by this method do not by any means turn out to obey the theta-criterion: some verbs turn out to assign more than one role to the same argument, others assign the same role to two different arguments, and some verbs "have" thematic roles that they do not assign to any NP, e.g. *to butter* has both a Theme and a Goal role, but the Theme is

<sup>5</sup> For convenience, I will continue to use *Thematic Role* for role types, when no confusion between role types and individual roles can arise.

"completely expressed by the verb" (1987:387). Whether *all* arguments of all verbs receive one of the thematic relations he has mentioned may not be stated, but his view of roles would not seem to require that they all do. In short, Jackendoff's interest in thematic roles arises purely from his desire to describe semantic patterns in lexical subcategorization and in syntax (that to him reveal conceptual structure), not to index arguments, and that thematic biuniqueness does not seem to result is of no concern. The individual-thematic-role escape hatch is of no interest to the Jackendoffs and Fillmores (or to this writer), for it ignores precisely the semantic generalization of role-type across verbs that gives the notion its interest. It should be added that Fillmore's later work on Case Grammar also permitted more than one case per argument (Fillmore 1977) (and of course he had never advocated a one-to-one relation between Deep (semantic structure) Cases and *surface* constituents).

Alas, this paper is not going to solve all these problems and does not purport to offer a theory of thematic roles that serves everyone's needs perfectly---nor does it attempt to demolish the notion once and for all. Its goal are more modest: (1) to lay out some methodological groundwork for studying thematic roles with the tools of model-theoretic semantics, and to propose some new strategies for attacking the area one step at a time, (2) to propose one new account of thematic roles (not unrelated to some other recent proposals) that seems to have merit at the "first step" in the strategy, and (3) perhaps most important of all, to make syntacticians and all linguists recognize the dangers of continuing to take this notion for granted and of assuming thematic roles are as well-motivated as phonemes or phrase-markers, and to encourage others, by this one example, to invent and explore other novel theories of thematic roles. And though this is not a psycholinguistics article and I am not a psycholinguist, I believe the linguist making a theoretical proposal about an area such as this has the responsibility to point out what psycholinguistic implications his proposal could have (the extent is it correct) and what questions it raises; thus the paper will include some speculations of this kind.

As is customary in model-theoretic semantics, I begin with the question of what *logical type* thematic roles should have, summarizing briefly the results of Dowty (1989) in §2. As the traditional empirical difficulties with arriving at a well-motivated set of role-types (most of all an argument-indexing set) may not be well-known today, I survey these in §3, including some pitfalls of misidentifying roles. It is argued in §4. that a fundamental methodological problem is that we have no agreement on what *kind* of linguistic evidence is appropriate for identifying a role-type correctly, and as a remedy, a strategy is defended of examining first the domain of *argument-selection* alone. As a further constraint on legitimate kinds of roles, I argue in §5 that *event-dependent* but not *perspective-dependent* roles be admitted. The inventory of role-types must, in view of the definitions in §4, be widened to involve a new kind of

role, *Incremental Theme* (§6.). With this preparation, I introduce a new theory of roles, in which roles are "prototypes", here called *thematic proto-roles*, rather than discrete categories (§7); the argument-selection principles for this theory and their workings are discussed in §8. Most interesting for this account of roles are three cases of subtly-contrasting argument selection: partially symmetric interactive predicates (§9.1), psychological predicates (§9.2), and the *spray-load* alternations (§1.3). Comparisons of the present view of roles with related proposals in the literature are made in §10. Some psycholinguistic implications that this account suggests for the place of thematic roles in the acquisition of grammar and of lexical meanings are considered in §11, and finally, what this account might imply about the so-called "unaccusative" phenomenon is considered in §12. The paper concludes with a brief summary of its proposals in §13.

## 2. The Logical Type of Thematic Roles

Because this paper uses model-theoretic semantics as its main investigative tool, we should begin our semantic investigation by asking what *logical type* thematic roles must be given in a formal semantic theory, in order for our theory to model the properties linguists have traditionally attributed to them. As Dowty (1989) is devoted to that question, I summarize here only very briefly the results of that paper and refer the reader to it for further details.

From the semantic point of view, the most general notion of thematic role (type) is *a set of entailments of a group of predicates with respect to one of the arguments of each*. (Thus a thematic role type is a kind of second-order property, a property of multi-placed predicates indexed by their argument-positions.)

For example, consider the subject argument of the two-place predicates *x murders y*, *x nominates y*, *x interrogates y*: entailments they all share include that *x* does a volitional act, that *x* moreover intends this to be the kind of act named by the verb, that *x* causes some event to take place involving *y* (*y* dies, *y* acquires a nomination, *y* answers questions--or at least hears them), that *x* moves or changes externally (i.e. not just mentally). The first entailment is not shared by *kills* (traffic accidents also kill), the second not by *convinces* (once can convince, or kill, inadvertently but cannot murder inadvertently), the third not by *looks at*, the last not by *understands*. By *entailment*, I mean the standard logical sense: one formula entails another if in every possible situation (in every model) in which the first is true, the second is true also. Since we are discussing entailments of "non-logical" predicates, I take this to be the same as an *analytic* implications (for which I also use the term *lexical entailment*: the implication follows from the meaning of the predicate in question alone). That is, a role-type like "Agent" is defined semantically as whatever entailments of verbs about NP referents are shared by the verbal argument-positions we label with the term "Agent" (and excludes whatever is entailed for those arguments that differs from one verb to the next); this sidesteps the question of whether "Agent" has a more "atomic" meaning underlying it, but it is precisely the point here to have an exact way of semantically characterizing roles that avoids such a presupposition, that can describe a possibly "arbitrary" as well as a "natural" role type, so that we can investigate and compare theories which do and don't involve the traditional notions.

Some of the lexical entailments that will be under discussion are perhaps also correctly described as presuppositions (in which case they correspond to the *selectional restrictions* of Chomsky (1965), but I assume it is now uncontroversial that these are correctly analyzed as semantic properties, not syntactic properties, of words). But the difference between presupposition and lexical entailments will not be important for our purposes.

This definition has the advantage that it is compatible with a theory, like Jackendoff's (1972, 1987) or Foley and van Valin's (1984), in which thematic role-types are defined by certain configurations of the (explicitly or implicitly interpreted) logical structures into which natural language predicates are translated<sup>1</sup>; with a theory, like Zaenen's (1988) or Rozwadowska's (1988), in which thematic roles are sets of semantic features (as long as we can fix a definite set of entailments, within some formal semantic framework, to correspond to each such feature of those accounts), as well as a theory in which there is no internal "structure" to lexical meanings and in which entailments of lexical meanings are all listed independently (e.g. by meaning postulates) and do not completely "cross-classify" by semantic primitives in any neat way as structuralist semantic theories assume lexical meaning does. It is also neutral as to whether thematic roles are argument-indexing or not.<sup>2</sup> Yet the definition allows us to be as precise as possible in describing the substantive semantic *content* of thematic roles: as precise or more so, I believe, than any kind of current semantic theory does. When "entailments" are mentioned below, the reader should keep in mind that this notion is neutral among these various theoretical views.

<sup>1</sup> See Dowty (1979) for a demonstration of how English predicates can be interpreted (compositionally within a sentence) by translating them into a "logical form" or "semantic representation" where they are decomposed into elements such as CAUSE and BECOME, these translations then being part of a formal model-theoretic interpretation of English.

<sup>2</sup> Dowty (1989) also points out that if there is a set of effectively argument-indexing thematic role types for all predicates of a language, then an expressively equivalent language is one in which n-place predicates are represented in the "neo-Davidsonian" way with such predicates and their arguments replaced by event-predicates with thematic roles as relations between events and participants, e.g. in which "Mary kissed John yesterday" is represented by (ii) rather than (i):

- (i) yesterday[kiss(Mary, John)]
- (ii) Ee[kissing(e) & Agent-of(John,e) & Patient-of(Mary,e) & yesterday(e)]

(but of course this conversion is not possible if thematic roles are not effectively indexing). However, it is not clear what kind of conceptual or computational advantage, if any, (ii) achieves, once lexical entailments are also taken account of (Dowty 1989). The view of thematic roles as second-order properties of relations indexed by argument is equally adequate whether thematic roles are indexing or not.

### 3. Traditional Problems in Identifying Thematic Roles and Using Them to Distinguish Arguments

#### 3.1 Role Fragmentations and Unclear Boundaries

Of various examples that might be cited of the question how "finely" to divide thematic roles, perhaps Agent is most striking: this is one of the most oft-cited and in some sense a very intuitive role, but it is one of the hardest to pin down. Jackendoff (1983) divides it into Agent vs. Actor, but D. A. Cruse (1973) split it four ways,

- (1)
  - a. *Volitive* "an act of the will is stated or implied" (p. 18).
  - b. *Effective* "exerts a force...because of its position, movement, etc." (p. 19)
  - c. *Initiative* "initiation of an action by giving a command" (p. 20)
  - d. *Agentive* "performed by an object [living things, certain types of machine, and natural agents] regarded as using its own energy in carrying out the action." (p. 21)

and cites syntactic tests to isolate each new role type. Possibly Lakoff (1977: 244) offered the largest fragmentation of Agency ever proposed, in which there were fourteen supposedly distinct characteristics (though properly speaking, some of Lakoff's characteristics involved the RELATIONSHIP between agent and patient, not agency by itself). The dilemma is, if we adopt the finer categorization of roles to achieve certain distinctions, do we thereby not miss generalizations by not being able to refer to the grosser Agent category as well?

Linguists have often found it hard to agree on and motivate the location of the boundary between role types. The sentences in (2) show one example of the difficulties that can be involved:

- (2)
  - a. I walked a mile  
I swam 30 meters,  
I slept twelve hours.
  - b. This weighs five pounds  
The piano measures 6'5".  
It took me an hour to grade the papers.  
The book cost me \$5.
  - c. I paid \$5 (this amount) (?this \$5-bill) for the book.  
The book cost me \$5 (?this amount) (#this \$5-bill).  
I bought the book for \$5 (this amount) (#this \$5-bill)

- d. I paid for the book with ?\$5 (#this amount) (this \$5-bill).  
I bought the book with ?\$5 (#this amount) (this \$5-bill).
- e. I'll trade this record for the book.

These sentences may involve a little-studied thematic role that has been called *Extent* (Andrews 1985). Note first that in (2a) the phrases *a mile*, *30 meters*, *12 hours* are adjuncts rather than subcategorized elements (they may be freely omitted without loss of acceptability or, apparently, change in the meaning of the rest of the sentence) and have an "adverbial function". Can adjuncts, or adverbs themselves, be assigned a thematic role? Fillmore (1988) said yes, but there would seem to be room for doubt. If we can assign measures of distance or weight a thematic role, how about measures of rate, as in *He drove the car 50 m.p.h.*? But then where do we stop? E.g. does *too fast* have a thematic role in *He drove the car too fast*, or does *quickly* in *She walks quickly*?

On the other hand, similar NPs are clearly subcategorized argument NPs in (2b), so surely they ARE assigned thematic roles here, and their meaning seems quite parallel to (2a): If we say these NPs have thematic roles in (2b) but not in (2a) (contrary to Fillmore, I presume), then it seems we ignore the semantic parallelism and say that it is a matter of syntactic form, not the meaning of a sentence alone, that determines what thematic roles are involved. (Perhaps indeed this is the correct conclusion, but the concept of thematic role becomes quite a different one if the conclusion is accepted rather than rejected, and if we cannot use meaning alone to decide thematic assignment, then we need to justify *which* syntactic differences we allow to indicate role differences and which we do not.)

What do we make of the differences between (2c) and (2d)? *five dollars* and *this amount*, like *a mile*, etc., seem to refer to an measurement of monetary quantity (in the abstract), while *this five-dollar bill* refers to a concrete object, a piece of paper that has such a value. So perhaps the correct thing to say is that the verb forms in (c) make reference to the measurement (and have the Extent role) while those in (d) refer to a physical quantity of currency (and have some other role, say Theme, parallel to (e)). The break, however, is not quite that clean. One can also say *That bad investment cost me my house in the country* (where *my house...* is not merely an Extent NP). And in the temporal domain, we have *John spent Tuesday writing the paper* as well as *John spent an hour washing the car* (suggesting *Tuesday* can express Extent, though cf. *#It took John Tuesday to wash the car*).

But the confusing part is about (c) and (d) is that in the common commercial transaction there exist BOTH a concrete pile of currency that changes hands AND a particular measurement of value that this currency has (except that in purchase by check or credit card, the former may be absent in concrete form, yet presumably there is still a "virtual Theme" of currency). So, we wonder, by analogy to Jackendoff's

analysis that *butter* has a Theme "expressed by the verb", should we say the Theme is verbally expressed in (2c) and the Extent expressed by the NP, while the reverse is true in (2d)? Or are Theme and Extent mutually exclusive in these sentences? How do we decide?

Perhaps these questions do have consistent and justifiable answers obtainable by diligent research. But the point is, thematic role-type assignment is, at best, not always transparent. Surely Jackendoff would agree, and he has constructed some very intricate arguments for some rather non-obvious assignments. For example, Jackendoff, following Gruber, says *money* is NOT the Theme in *Nelson ran out of money* and *Fred came into a lot of money*, but rather the Goal (Jackendoff 1976: 134), so Theme is not always simplistically "that which moves or changes". And for Jackendoff and Gruber, *The circle surrounds the dot* has Theme as subject and Location as object, but in *The circle contains the dot*, the subject is Location (1976:97-98), even though elsewhere, the subjects of locative sentences seem consistently to be Themes (e.g. in both *x is to the right of y* and *y is to the left of x* (1976:98)). This is not to deny either that Jackendoff presents arguments for these assignments which have their appeal or that his resulting analysis is self-consistent, but it is only fair to observe that other linguists, having somewhat different methods and emphasizing different data, can come up with reasonable but different assignments.

The methodological dilemma in the view that thematic role type identification cannot be made from meaning alone but can be affected by syntax as well is that it all but excludes it from the possibility of empirical falsification. That is, when it is pointed out to a syntactician that there is a semantic inconsistency in her appeal to a certain thematic role in her analysis of a new syntactic construction, she can reply that this is simply one of those cases where syntax or the existence of certain lexical items, as well as purely semantic criteria, determine role-type distribution. Of course, in principle there MIGHT be an independent way to validate or falsify such a claim with further data, but in practice this independent justification can be hard to find, so that appeal to roles in this not-strictly-semantic way seems perilously close to the "wild card to meet the exigencies of syntax" that Jackendoff himself cautions us about.

### **3.2 Cases Where There May Be No Motivatable Role That Can Distinguish Two Arguments**

Another familiar problem with thematic roles, which is the complement of the previous one (and would be fatal for argument indexing by role type), is the case where two arguments of the same verb do not seem to be distinguished from each other by any entailments that the verb produces whatsoever, and if so, there can of course be no motivation for assigning distinct roles to them on semantic grounds:

among the clearest examples are probably symmetric stative predicates as in (3):

- (3) This is similar to that  
equal to  
near  
resembles  
weighs as much as

That is, if this is similar to that then that is similar to this and vice-versa, with no apparent asymmetry in what is predicated of the two arguments on which to pin a distinction in role type. The same difficulty arises with conversely entailing predicates, e.g. *x is to the left of y* and *y is to the right of x*. (That there might in fact be a subtle difference in subject vs. object that signifies a role difference both here and in (4) below, say a difference in "perspective", is a position I will address in §5. below.)

Another familiar conundrum of this kind is verbs which refer to commercial transactions such as *buy*, *sell*, and similar verbs, e.g. *rent*:

- (4) a. John sold the piano to Mary for \$1,000.  
b. Mary bought the piano from John for \$1,000.

As Jackendoff (1987:381) and many others have noticed, both buyer and seller must act agentively (voluntarily) whenever such a transaction takes place, and one or the other (or both) must act to effect transfer (signing names, or moving or taking the object or the money, the meaning of the verb being indifferent to how the change of possession is caused, as long as both participants desire both reciprocal transfers of possession to occur), and there is no obvious reason why either is entailed to act "more agentively" than the other. (Likewise, both currency and the purchased item necessarily change hands, so there is a danger that there are two Themes for such verbs, as well as a Goal and a Source for each transferred entity, namely the buyer and seller in each case.) Of course, such verbs in fact distinguish the two agents semantically according to which acquires the quantity of cash (or equivalent medium of exchange) versus which acquires the desired object of some other kind, but labeling such a difference a "thematic role" seems ill-motivated; it would violate what I think is an implicit principle that we should not postulate a thematic role-type that is limited to only one or two verbs (or a small set of near-synonyms) but expect each role-type to be applicable to a reasonable range of verb meanings.

### **3.3 Pitfalls of Misidentifying the Motivation for a Role:**

Though not an inherent problem in the concept of thematic role nor an

insurmountable barrier to identifying roles empirically, it is worth pointing out that a significant practical problem with correctly identifying evidence for particular role types has been the ease of confusing generalizations that should be stated in terms of thematic roles with generalizations of other kinds --- either purely syntactic generalizations, semantic generalizations (other than ones involving thematic roles per se), or pragmatic generalizations. These pitfalls should be kept in mind by anyone who reads the literature critically for evidence pertaining to the phenomenon.

### 3.3.1. The generalization in question is actually a purely syntactic one.

As an example of what at first appears to be a thematic role generalization turning out to be a purely syntactic one, note that Anderson (1977), Wasow (1980), Williams (1980), and Bresnan (1982) have all put forth the hypothesis that English lexical passives, such as the "un-passives" in (5) are only grammatical when formed on Theme objects:

- (5) a. A new car was sold to the customer.  
The customer was sold a new car.
- b. an unsold car  
\*an unsold customer

But the correct generalization is that lexical passives can be formed from all and only the lexical monotransitive verbs in English (Levin and Rappaport 1986), that is, from verbs that can appear with one object NP and no other complements in their active form, regardless of the thematic role type of this NP. The data in (6) illustrates this. (This particular generalization, incidentally, is predicted to hold by the categorial theory of lexical rules and relation-changing rules in Dowty (1978)<sup>1</sup>

<sup>1</sup>The theory of lexical rules in Dowty (1978) entails that lexical rules are defined over the same system of categories and expressions as syntactic rules are, with one key difference being that only the *basic* (i.e. lexical) members of a given category can be inputs to a lexical rule applying to that category, while both basic and syntactically derived (i.e. complex) expressions of that category can be inputs to a syntactic rule applying to that same category (this being, like other versions of Montague Grammar, a theory in which any category can have both lexical and syntactically complex members). A claim about English made in that paper is that it has both a lexical and a syntactic passive rule, each applying to the category of transitive verbs (possible phrasal ones, for the syntactic case). A ditransitive verb combines via syntactic rule with a NP to form a (phrasal) transitive verb; for example, *sell to the customer* and *sell a car* are phrases of this category, which, if combined with direct objects, give rise to examples such as *sell a car to the customer* and *sell the customer a car* (via "wrapping" operations); by using the phrasal TV's instead as input to the (syntactic) passive rule, the intransitive VPs *be sold to the customer* and *be sold a car* are produced. The lexical passive rule cannot apply to the ditransitive *sell* directly, since the rule is defined only on TV, not on the ditransitive category TV/T, and by the aforementioned principle,

- (6)      an unsold book (*cf.* John sold the book)  
           \*an unsold customer (*cf.* \*John sold the customer)
- \*an unfed hamburger (*cf.* \*John fed the hamburger)  
           the unfed children (*cf.* John fed the children)
- the unserved soup (*cf.* the waiter served the soup)  
           the unserved customer (*cf.* the waiter served the customer)

**3.3.2 The correct generalization is in terms of some semantic distinction other than one characterized by a thematic role type.**

As an example of this sort, Rappaport (1983:131) proposed that "no derived nominal inherits the argument structure (AGENT, EXPERIENCER) from its verb". This is supposed to explain the ungrammaticality of the *by*-phrases in derived nominals of the psychological verbs in (7):

- (7)      Amy's fright (\*by the scarecrow),  
           The class's boredom (\*by the lecturer),  
           Deborah's amusement (\*by Randy),  
           Sam's annoyance (\*by Dave)

However, Rappaport also noted that such derived nominals are always understood as referring to states rather than events, and she considered the possibility of stating this generalization in terms of stativity rather than in terms of thematic roles, the idea being that the *by*-phrases would be incompatible with a stative interpretation. But notice that the restriction against non-stative interpretations is needed independently to explain why adverbials implying an event-interpretation are ungrammatical with such nominals, as in (8), even though no Agent is present syntactically and even, as in the second example, there is an adjective like *unintentional* that excludes the understanding that an Agent was involved, syntactically present or not. Hence the stativity restriction is preferable to one in terms of roles.

a lexical rule cannot apply to the syntactically complex phrase *sell a car* (even though the category would be right); a lexical passive would therefore be possible for such a verb only if that verb independently had its valence reduced from ditransitive to transitive by another lexical rule. Now there are two ways to convert a ditransitive to a transitive--by suppressing the "Goal" argument or suppressing the "Theme" object--- and English has both kinds of "monotransitivizations" (*cf.* right-hand column in example (6)). But as can be checked from that data, the lexical passive is, as predicted, possible only if the corresponding "monotransitive" form exists with the appropriate argument omitted from the ditransitive, no matter whether the remaining argument is Theme (as with

- (8) the boredom of the class (#that happened ten minutes after the lecture started)  
 the unintentional fright of the children (#that occurred when they saw the scarecrow)

### 3.3.3. The generalization is actually a pragmatic one.

Third, a generalization that appears to be describable in terms of roles can turn out to be pragmatic in nature. Jackendoff (1972, 1987), Grimshaw (1975), Williams (1980), and Nishigauchi (1984) have proposed that the control of null subjects (N.B. *not* the object gaps) of infinitival relatives and transitive purpose clauses, e.g. in (9),

- (9) John bought a book to read to the children.  
 John bought Mary a book to read to the children.

is determined by thematic role or by a thematic role hierarchy (*Goal > Source/Location > Theme*). But Ladusaw and Dowty (1988) presented counterexamples to this hypothesis in the form of structurally and semantically parallel sentences which allow different NPs to control the infinitive. Following Bach (1982), we argue that extra-linguistic practical reasoning determines the control in these cases--i.e. reasoning about who would have what object at his/her disposal at what point in the action<sup>2</sup>. One vivid illustration of this is the example in (10) (a kind suggested by Bach), in which the subject controller can be understood as the addressee and speaker together.

- (10) Here is a bottle of wine. I brought it to drink with our dinner.

A revealing example from Ladusaw and Dowty (1988:68) is (11) (i.e. the italicized purpose clause):

- (11) John has been spending the night at Mary's house a lot lately and using her toothbrush, which irritated her a great deal. So to appease her, John bought Mary a second toothbrush *to brush his teeth with when he stayed at her house.*

Normally the Goal, or person who ends up as possessor of the object at the end of the action (here Mary), is the subject controller of the purpose clause, since that person will be in a natural position to use it for some future purpose. But our

*sell*), Recipient (*feed*) or whether both possibilities exist (as with *serve*).

ability to understand the unusual situation in which the owner of an object is different from its intended user is what permits us to naturally take the NP *John* as the controller in (11), in violation of generalizations in terms of role hierarchies. (One can also obtain the other control reading of this last sentence by putting it in a context where it is assumed that John customarily has Mary brush his teeth for him.) See Ladusaw and Dowty (1988) for further examples and discussion.<sup>3</sup>

### **3.3.4. The phenomenon in question is a consequence of general constraints between syntax and discourse structure.**

A possible instance of this category of misidentification, a second kind of pragmatic case, would be the status of roles such as Figure and Ground discussed in §5 below.

<sup>2</sup> But N.B. Ladusaw and Dowty make this claim is only about infinitival relative and purpose clauses, not control of the complements of *try*, *promise* *persuade*, etc., which is acknowledged to be syntactically governed, although ultimately a connection to "practical reasoning" is surely involved even here.

<sup>3</sup> This paper is criticized by Jones (1988), who shows that our claims about the verb *rob* are either wrong or, at best, in need of further explanation. However, Jones makes no comment about (11) at all. His solution to the observed variation in position of the subject controller of transitive purpose clauses is to posit a new thematic role *Location*, also characterized as "eventual possessor", which is assigned to the direct object in *John bought it* but not in *John bought it for Mary*. But surely our understanding of the difference in "eventual possessor" in these two examples is due to implicature, not the lexical meaning of *buy*, and such capriciousness in syntax-meaning correspondence as this role would need to display is not motivated elsewhere in the literature on roles. More important, the notion of "eventual possessor" is exactly what is relevant to our understanding of the control in examples like (10) as well, though no true "thematic role hierarchy" generalization can cover (10) and the other examples, since the controller in (10) is not an NP in the sentence. Thus I believe that while "eventual possessor" is a good intuitive description of how we understand these controllers, Jones' proposal only really makes sense as the pragmatic solution of Ladusaw and Dowty under a new name, not as a true "thematic role" analysis.

(Incidentally, Jones' discussion of obligatory control (with *try*, *promise*, etc.) does not seem to recognize that Ladusaw and Dowty take the position that obligatory control is grammatically fixed (just as Jones' own position holds) and only motivated by parallel semantic-pragmatic patterns, which to be sure in very rare cases seem able to "overrule" grammatically fixed control with partial success.)

#### 4. A Proposed Strategy for Research on the Semantics of Thematic Roles.

One conclusion I would like to draw from the above difficulties is that linguists may be casting their nets too wide in selecting linguistic data to identify or justify thematic roles. These role types have often been motivated and identified by correlating with various syntactic environments which admit one kind of role but not another (cf. e.g. Cruse's tests for his four kinds of Agents). But is just any correlation of a semantic distinction with a syntactic or lexical pattern evidence for a role-type?

Here is a case in point. In Dowty (1979) I pointed out that the only English stative verbs that can occur in progressive tense are *sit*, *stand*, *lie* and other verbs entailing a particular spatial orientation of an object has within its location: cf. *The book is lying on the floor*, *The umbrella is standing in the corner*, vs. *\*The book is being on the floor* or *\*There is existing iron oxide on Mars*. (I argued that this class was further semantically distinct in being the only statives that could be true or false for intervals rather than also moments in time, and the class may turn out to be those of Carlson's (1977) *stage-level* predicates which are also stative.) Watters (1985:14-17) observes that in Tepehua and other Totonacan languages a class of verbs distinguished by several morphological and syntactic properties (e.g. occurring in only certain tenses in Tepehua) are likewise those belonging to this semantic class (a superset of the English ones but plainly the same natural class, as it includes also verbs meaning "is fallen over", "is hung up on something", etc.).

Consider furthermore the closely related if not identical phenomenon that several English constructions, including the above progressive *sit-stand-lie* case, presuppose that a property being predicated of an object is temporary rather than permanent (Dowty 1975, Bolinger 1967), or in Bolinger's terms, an *accidental* rather than *essential* property. Out of the seven or more cases he mentions, three are illustrated below. The second sentence in each pair sounds odd simply because the property predicated is not a temporary one, given usual assumptions about the facts of the world:

(12) a. Clause-final adjective adjuncts:

She caught a glimpse of the dancer nude.  
#She caught a glimpse of the statue nude.

b. Complements of *with* and *without*:

They took the vote with the chairman absent.  
#They took the vote with the chairman arrogant.

d. *Sit-stand-lie* progressives:

The rowboat is lying on the riverbank.  
#New Orleans is lying at the mouth of the Mississippi River.

(*cf.* New Orleans lies at the mouth of the Mississippi River.)

For additional discussion see Bolinger (1967, 1971, 1973), Dowty (1975, 1979:173-180). The phenomenon corresponds to the familiar *estar* vs. *ser* contrast in Spanish and perhaps to contrasts in other languages as well. Does having this many syntactic manifestations of the contrast entitle us to christen this a new thematic role *Temporary/Spatially-Oriented*?

I expect many readers will agree with me that these are somehow not the kind of contrasts we want to take as identifying a "thematic role", but if they are not, then exactly why not? The variety of semantic distinctions that correlate with syntactic and lexical patterns in one way or another is surely enormous. To postulate thematic role types for each of them is, quite possibly, to dilute the notion beyond its usefulness, but what we lack is a principled way to decide what kind of data motivates a thematic role type and what does not.

Conceivably, the difficulty we have had in reaching agreement on just what a theory of thematic roles should look like is analogous to that of the blind men examining the elephant, each having hold of a different part of its body. Though we may correctly intuit that our disparate observations are related to a common phenomenon in the grand scheme of things, it is not surprising that we are frustrated when we cannot immediately fit our present observations directly together so as to construct from them a theory of the single thing which is the leg-ear-tail-trunk of the elephant.

What is the remedy? What I propose is to try to separate our various observations about putative thematic roles along natural boundaries, to the extent we can justify non-arbitrary divisions among them. Then, as a first step, we construct the

best-motivated theoretical account for the observations of each domain separately, ignoring prior conceptions of "thematic roles" based partly from data of other domains. Examples of such domains would include: (i) the argument-selection problem (see below); (ii) the rather preposition-dependent and lexical-structural observations of the Jackendoff-Gruber approach; (iii) the argument-indexing perspective, (iv) the phenomenon of lexical meaning extension across cognitive categories as in Jackendoff 1983 (e.g. from the literal locative Goal in *throw into the room* to abstract Goal in *rewrite into a journal article*); (v) roles as reflected in language acquisition (e.g. Clark and Carpenter (1989) on generalized "Source" cf. §12); (vi) reflexes of roles in language typology (e.g. Croft 1986b, cf. §12); and (vii) experiments on adult sentence processing. If two domains really do turn out to lead to the very same theory (and inventory) of roles, so much the better, as this would suggest that the observations of each domain independently reflect the same underlying phenomenon. But if two domains of observations lend themselves to quite different optimum theories, then we should not fret but conclude that at least one of these domains and its theory (or several of them) represent only the leg or the trunk of the phenomenon: not the whole elephant, but still related to it in an important way we do not yet fully understand.

In most of the remainder of this paper (§5-§9), I am going to focus solely on the *argument selection phenomenon*, and I will conclude that the best theory to describe it is quite different from, and in some ways simpler than, the usual conception of thematic role type. This phenomenon is the question of what principles languages use to determine, for each argument of an  $n$ -place relation that is denoted by a predicate, which argument (intuitively speaking) can be expressed by which grammatical relation<sup>1</sup>. Although the problem was studied a great deal in (early) Case Grammar (under the term *Subject Selection*, Fillmore 1968), Gruber and Jackendoff also seemed to have recognized the importance of thematic roles for argument selection, as have of course a number of more recent writers, who refer to the problem as *template matching* (Baker 1985) or the question of the existence of a *universal alignment principle* (Perlmutter and Postal 1984, Rosen 1984) or a *universal theta assignment principle* (Baker 1985). Note that (very concrete!) data on this problem is, relatively speaking, easy to come by: dictionaries and reference grammars for any language list the various valences for each verb.

By "cutting the data along natural class boundaries", I mean more specifically that

<sup>1</sup> I say "intuitive argument" because this sentence does not literally make sense in an extensional semantic theory, where the denotation of an  $n$ -place predicate is an  $n$ -place relation (set of  $n$ -tuples), or in a weakly-intensional theory such as Montague's (1970, 1974) in which this denotation is a function from possible worlds to such relations. Rather, in such theories the problem is described as choosing, from the permutation-set of an  $n$ -place relation (i.e. the set in which each relation is derived from another by permuting corresponding members in the  $n$ -tuples throughout the relation), which permutation(s) will be denoted by a predicate of the language and which will not.

in the present investigation:

(i) No semantic distinction will count as relevant data for our theory of roles unless it can be shown to be relevant to argument selection somewhere in some language, no matter how traditional a role it characterizes.

(ii) ANY semantic distinction that can definitely be shown relevant to argument selection should count toward defining a role type, no matter whether it relates to a traditional role or not.

## 5. Event-Dependent and Perspective-Dependent Thematic Roles

As a consequence of adopting this last methodological principle (ii), we will now see how one class of thematic roles found in the literature should be eliminated from our discussion. Certain kinds of thematic roles that can be involved in an event seem to be quite invariable across different perspectives or ways of viewing an event. Among these are Agent, Patient, Experiencer, and (with a very few possible exceptions) Theme (in the sense of "thing which moves or changes"), Source, and Goal, as well as the "adjunct" roles Instrumental, Locative, Temporal, and Benefactive. That is, if Helen carries the rock from John to the porch, then no matter whether one in some way "views" that kind of event from the point of view of Helen, the rock, John, or the porch, or whether one passivizes the sentence or otherwise alters it syntactically (by Topicalizing a NP, etc), by substituting a synonym of "carry", or by putting it in various discourse context, Helen still remains the Agent, the rock still the Patient (Theme), John the Source and the porch the Goal. That is, any truth-conditionally equivalent sentence has the same role assignments. The nature of the carrying event itself, it seems, fixes these roles.

Other proposed roles in the literature are different. An early example is Jackendoff and Gruber's use of Theme with stative predicates. As already mentioned, the grammatical subject (alone) is Theme in both *The rock is to the left of the tree* and *The tree is to the right of the rock*, according to Jackendoff (1976:94-96). These describe exactly the same state of affairs (assuming we don't change the deictic orientation for "left" and "right" between sentences), yet the rock is held to the Theme according to one but not the other (and the tree vice versa). Talmy (1978, 1985a, 1985b) has used the terms *Figure* and *Ground* for the same contrast, at least once explicitly raising the question whether these categories should be regarded as thematic roles:

- (13) a. The lamp (Figure) is over the table (Ground)  
b. The table (Figure) is under the lamp (Ground)
- c. The bicycle (Figure) is near the tree (Ground)  
d. The tree (Figure) is near the bicycle (Ground)

This kind of distinction has of course been noted by many writers under many terms (cf. e.g. Fillmore 1977), though not always under the rubric of thematic roles or semantic case. Should it be? Writers seem to agree that the meaning difference is (as the names Figure and Ground imply) a matter of asserting the location of the Figure/Theme NP with respect to the Ground/Location, putting the first NP "in perspective", making it more salient, etc. (though syntactic arguments have sometimes also been given for this kind of role assignment, e.g. by Jackendoff (1976:96-98)). Note incidentally that if this semantic contrast is a matter of

thematic role, it permits us to escape all counterexamples to thematic uniqueness mentioned earlier. That is, *x is similar to y* could differ from *y is similar to x* in which NP is the Figure or Theme, and (though now contrary to Jackendoff but with Fillmore), *buy* could differ from *sell* in that the first has the buyer as Figure and the second the seller. (Would admitting this contrast as a thematic role difference, we might also wonder, be tantamount to reducing the hypothesis of thematic uniqueness to a non-empirical question?)

I want to suggest that we rule out such perspective-dependent notions as Figure/Ground or Gruber's stative Theme as candidates for thematic roles. This is not to deny the existence of these distinctions nor their importance, but to propose only that thematic role is the wrong rubric for them.

The argument for this position is outlined as follows. Natural languages make use of a variety of grammatical means for indicating how the NP referents and other information in a sentence are related to the immediately-preceding and the not-so-immediately-preceding discourse and to the common ground of information shared by the discourse participants. It is widely agreed that in English and languages of similar typology, the grammatical relation *subject* is a weak indicator of "Topic" (LI, 1976), but in place of that much-disputed notion here, I will say simply that I assume the NP referent of a subject is weakly indicated to be "more directly connected" to the preceding discourse and common ground than those of other NPs in the same sentence, e.g. the subject referent may have been mentioned relatively recently. (By "weak indicator" I mean that it is a default that can be overridden by other indicators of givenness, e.g. presence of an anaphoric form elsewhere in the sentence, topicalization or clefting.) "Newness" vs. "Givenness" is a matter of degree, not an absolute contrast (Prince 1981), and note that I say "more connected" relative to other NPs, not that it meets any absolute criterion of topic-hood. (One consequence of this conventional association, presumably, is that existential constructions in many languages have a grammatical form that removes the NP from normal grammatical subject status, possibly displacing it with a dummy NP or locative (Clark 1978), thereby signifying that its referent is NOT connected to previous discourse in the way that subject status would otherwise indicate.) Note that we now speak of perspective-dependent notions: whether a referent is new or given varies with the discourse even for the same factually described situation. The argument for eliminating Figure/Ground from the inventory of thematic roles is thus in outline:

- (i) In an adequate linguistic description, greater relative connectedness to previous discourse, givenness, etc. must be explicitly specified as a semantic correlate of grammatical subject denotations (in English-like languages).
- (ii) All putative instances of perspective-dependent thematic roles and other "perspective-indicating" lexical entailments of words can be shown to be instances of (i) when properly analyzed.
- (iii) Therefore by Ockham's Razor, perspective-dependent thematic roles are unnecessary, and all roles are event-dependent in meaning.

Establishing (ii) would be a major undertaking far beyond the scope of this article and will have to wait for another context, but here are two sample arguments. The difficulty in distinguishing a semantic discourse correlate of lexical verbs with respect to their subject argument (which is where "Figure"-type roles always seem to be found) from that of a discourse semantic correlate of the grammatical relation subject is, of course, that these almost always involve one and the same argument. The one case where they diverge is the passive; consider (14):

- (14) a. The truck hit the tree.  
       b. The truck hit it.  
       c. The tree was hit by the truck.  
       d. It was hit by the truck.

One of the simplest and strongest ways an NP in sentence can be "connected to previous discourse" is as a direct answer to a WH-question. Imagine the examples in (14) as answers to the question *What happened to the tree?*: The most preferred answers, I believe, are (b) and (d). Example (c) is somewhat less preferred (because the normal case is to use anaphoric reference to the tree in this situation), but (a) is most clearly deviant in this context. It was already suggested that an (discourse-anchored) anaphoric NP (in contrast to non-anaphoric ones) is an indicator of connection-to-context that overrides subject as indicator, which would explain why (b) is as natural as (d) to "connect" the answer to the question (and (d) has the Gricean disadvantage of being a LONGER sentence than (b)). But without an anaphoric asymmetry in the two NPs, it is clearly better to put the answering NP in subject position, (c), than non-subject, (a). In (15), to be taken as answers to the question *What happened to the truck?*,

- (15) a. The truck hit the tree.  
       b. It hit the tree.  
       c. The tree was hit by the truck.  
       d. The tree was hit by it.

only (b) is fully natural, with (a) slightly less preferred, and both (c) and (d) deviant. As before, putting the answering NP in subject position is normal: (a) or (b). Because both (c) and (d) involve a passive (a longer and "marked" form) where the active would have had the "right" NP as subject, even the "correct" asymmetry in anaphoric forms in (d) does not override the wrong subject choice (contrast with (14b)). Attributing the source of prominence to the lexical subject-argument of *hit* could not have explained this pattern, as it corresponds in meaning to a non-subject in the passives. This paradigm of voice-shift X anaphora-shift can be repeated with other kinds of connections to prior discourse besides WH question and answer, I believe.

Another argument, which brings out more intuitively the "perspective" associated with the subject position, is to use the verb which is most certainly a true symmetric predicate: the verb *be* with two proper names or other definite referring expressions. I avoid cases like *Mary is a doctor*, where the second is indefinite, because of the now common proposal that this "predicative" NP is in some sense a predicate (Partee 1986), in contrast to the subject NP, which entails of course that this *be* is not symmetric. But though "identity statements" like *Tully is Cicero* have been subjected to much scrutiny in the philosophical literature, one aspect of the meaning of *be* here which has not been questioned, as far as I know, is that its meaning is symmetrical with two flanking names or definite descriptions. I assume the burden of proof here is on anyone who would want to claim *be* is NOT symmetrical in meaning in these cases. Now in ordinary discourse, one often finds advice given and questions asked in the following counterfactual forms:

- (16) a. If I were you, I wouldn't buy that used car.  
       b. If you were me, would you ask him for a date?

The semantics of these sentences presents many mysteries, but here I am only interested in the fact that while the two examples in (16) are common, none of the statements in (17) are completely normal, nor the questions in (18):

- (17) a. #If I were you, you wouldn't buy that used car.  
       b.?#If you were me, you wouldn't buy that used car.  
       c. #If you were me, I wouldn't buy that used car.
- (18) a. #If you were me, would I ask him for a date?  
       b.?#If I were you, would I ask him for a date?  
       c. #If I were you, would you ask him for a date?

(Parallel comments would hold for *If I were Bill, I would take the job.* vs. *#If I were Bill, he would take the job.*) A full discussion would take us too far afield, but note: (i) In the advising statement, the subject pronoun must be first-person, but in

the question it must be second person: (17b), (18b)<sup>1</sup>. This is probably because, in some sense, (16a) offers the speaker's thoughts and judgments applied to the hearer's personal situation, "the speaker's mind in the hearer's body", while the question asks for the reverse (though why identification by thoughts takes precedence over physical identity may ultimately be obscure). (ii) The subject pronoun of the antecedent clause must be the same as that of the consequent clause: (17a,c), (18a,c). Some might dismiss this as a preference for grammatical parallelism, but I think it is not. The offending (a) and (c) sentences are not bad style nor uninterpretable but are, with work, meaningful and differ from (16) in tending to suggest a bizarre "mind control" of one person over another of the science-fiction sort. The relevance to our present concerns, however, is simply that there are clear asymmetries in meaning brought about by interchanging arguments of *be*---involving a difference in "perspective"---which we otherwise have no need to attribute to *be*'s lexical meaning but may need independently to characterize the subject vs. non-subject NPs in discourse.<sup>2</sup>

<sup>1</sup> Sentences of the (17b)-form are not advice like (16a) is, though they can be retorts to counterfactual advice, e.g. as a response to your advice in (16a), I might say *If you were me, you couldn't AFFORD that car.* That is, they continue to talk about the same kind of non-actual possible world one's interlocutor has set up. Similarly, (18b) is not a request for information like (16a) but can be a rhetorical question (i.e. expecting a negative answer.)

<sup>2</sup> That is, I am assuming that at least in a counterfactual sentence, but maybe in other intensional contexts as well, the subject NP referent by itself, in addition to contributing compositionally to the proposition expressed by the antecedent clause as a whole, somehow also plays a role in determining the precise connection between that counterfactual antecedent proposition and the actual world, i.e. via how the persons in the counterfactual worlds are anchored to their real-world counterparts. Whether it does this through literal meaning or only implicature I do not know; this question is well beyond the scope of this paper and relates to much recent work, e.g. (Stalnaker 1984).

## 6. Incremental Theme

Besides narrowing the class of role data, our focus on argument selection will also require us to widen our scope to include a new role category I will call *Incremental Theme*. Though the key idea of this section was once sketched in a talk (Dowty 1987) suggested by proposals in Hinrichs (1985), this idea was independently noticed and most fully developed formally by Manfred Krifka (Krifka 1987, 1989). This proposal is that the familiar way in which the aspect of telic predicates (or *accomplishments* and *achievements*) depends on their NP arguments (Verkuyl 1972, Dowty 1979) can be captured formally by the principle that *the meaning of a telic predicate is a homomorphism from its (structured) Theme argument denotations into a (structured) domain of events*, modulo its other arguments. *Homomorphism* is a standard mathematical notion which is finding more and more applications in linguistics, e.g. Montague (1970) Keenan and Faltz (1985). Put simply, a homomorphism is a function, from its domain to its range, which preserves some structural relation defined on its domain in a similar relation defined on the range. (See Partee et al 1990 for formal discussion.) In the case of telic predicates, this relation which is preserved is the "part-of" relation: If  $x$  is part of  $y$ , then if a telic predicate maps  $y$  (as Theme) onto event  $e$ , it must map  $x$  onto an event  $e'$  which is part of  $e$ .

For example, take the telic event described by *mow the lawn*: if I tell my son to mow the lawn (right now), and then look at the lawn an hour later, I will be able to conclude something about the "aspect" of the event of his mowing the lawn from the state of the lawn, viz., that the event is either not yet begun, partly done but not finished, or completed, according to whether the grass on the lawn is all tall, partly short, or all short. On the other hand, I will not necessarily be able to inspect the state of my son and conclude anything at all about his completion of his mowing the lawn. In this event, my son is the Agent and the lawn is the Theme, in fact the Incremental Theme. The homomorphism claim means that because of the meaning of *mow*, the state of parts of the lawn and their part-whole relationships are reflected in the parts of the event of mowing it and ITS part-whole relationships.<sup>1</sup>

The hypothesis that telic predicates are homomorphisms neatly explains Verkuyl's (1972) long-standing puzzle about the way that bare plurals and mass term arguments can make a sentence with a telic predicate behave as if were "durative"

<sup>1</sup> Note that the claim is NOT that telic predicates denote functions that are also one-to-one, i.e. *isomorphisms*. A homomorphism can be a many-one function. Thus the claim that *eat* denotes a homomorphism from its object argument denotation to an event is not counterexemplified by a situation in which I eat a whole sandwich in one gulp (all parts of the sandwich mapped onto the same event) instead of the more usual one in which different parts of the sandwich are mapped by the eating event into the distinct sub-events of eating the respective parts. And, this also implies, the part-of relation is not understood as "is a proper subpart of" but rather, so as to allow a thing to count as a part of itself.

or "imperfective" in aspect, as in the familiar examples (19):

- (19) a. John drank a glass of beer. ["perfective"]  
b. John drank beer (for an hour). ["durative"]

The NP *A glass of beer* refers to an entity that has various proper subparts that are of course quantities of beer of various sizes, though no one of these is itself a GLASS of beer: if *drink*, a telic predicate, is a Theme-to-event homomorphism, it maps this argument denotation into an event of drinking a glass of beer, and maps the subparts of this quantity of beer into sub-events of drinking subquantities of that beer, but it also follows that none of these proper subevents is an event of drinking a GLASS of beer. If we follow Bennett and Partee (1972), Taylor (1977), Dowty (1979) and later writers in defining a telic sentence as one denoting a unique event, i.e. one having no proper subevents describable by the same sentence, then (19a) is telic. Example (19b) is similar, and of course has the very same homomorphic predicate mapping some quantity of beer and its subparts into a corresponding event and its subevents: the difference is that the NP *beer* does not specify a definite quantity of beer, so subquantities of the main quantity could also be referred to by this same NP *beer*. This implies that the subevents which make up the main event are describable by the same core sentence, i.e. *John drink beer*, so this sentence, unlike (19a), can simultaneously describe an event and sub-events of that same event: hence (19b) is not a telic sentence and it (but not (19a)) can be felicitously and truthfully modified by a durative adverbial *for an hour* (which requires (Dowty 1979:332-334) that there are multiple successive occasions on which its core sentence is true). See Krifka (1987:13-19, 1989) for more detailed discussion of this analysis.

Examples of Incremental Themes are traditional "effected objects", "destroyed" objects, and objects entailed to undergo a DEFINITE change of state:

- (20) build a house, write a letter, perform a sonata;  
destroy a presidential finding, eat a sandwich;  
paint a house, polish a shoe, proofread an article;

But it turns out that many traditional Themes, i.e. things entailed to "move or undergo a change of state", are not Incremental Themes. For example, the objects in *push a cart*, *raise the thermostat*, *dim the lights* move or change, but the verbs by themselves imply only an INDEFINITE change of position or state (and they are atelic). On the other hand, many achievement verbs entail a definite change in one of their arguments but are not homomorphic (*die*, *touch the finish line*, *recognize a face*), except in a trivial sense, since by normal criteria their arguments never undergo this change in distinguishable separate stages, i.e. subevents. Therefore it

would be both an undergeneralization and an overgeneralization to identify Incremental Themes with cases in which the direct object referent is "totally affected or effected". A different situation is presented by the examples in (21):

(21) walk from the bank to the post office, drive (a car) from New York to Chicago, run a mile;

grow into an adult, become an architect.

If John drives from New York to Chicago, John necessarily undergoes a definite change of location from one place to the other, but if this trip were interrupted before it was finished, we would not infer that part of John has arrived in Chicago while the rest of him is still in New York.

Upon reflection, it is clear that what is partially but not totally affected in this case, in a way parallel to the themes in (19), is the **path** John traverses in driving from NY to Chicago: if the event is started but not completed, then part of this path has been traversed by John, not all of it, but the positions of parts of JOHN of course remain intact with respect to each other. We could distinguish the Themes in (21) by a new term *Holistic Themes*: though they undergo a change of state in stages, the change is "incremental" only because of some relationship they bear to the true Incremental Theme, not because they undergo a change part by part. One interesting thing about such examples is that the "argument" with respect to which these telic predicates are homomorphisms on this hypothesis, namely the Path argument, is (like Jackendoff's "Theme" in the verb *butter*) not a syntactically realized argument at all; the prepositional phrases *from New York* and *to Chicago* refer to the beginning and end points of the Path. However, an Incremental Path Theme can be syntactically realized as a Direct Object in semantically parallel telic examples like *cross the desert*, *traverse the United States (in six days)* or *drive the Blue Ridge Skyway (from beginning to end)*. Similar observations about Holistic Themes apply to the last examples in (21), *John was becoming an architect but was interrupted before he could finish his degree*, etc., though here the "Path", if we want to call it that, is even more removed from syntactic expression---the stages that one goes through to reach the status of architect were partly but not exhaustively achieved; NOT "part of John but not all of him has become an architect".

Yet a different manifestation of Incremental Theme appears in (22), which we can call a *Representation-Source Theme*:

- (22) photograph a scene  
copy a file  
memorize a poem  
read a book

To see the point of these, compare them with their (near) paraphrases in (23):

- (23) take a photograph of a scene  
make a copy of a file  
form a memory of a poem  
acquire the information in a book (from it)

The direct objects in (23) are effected objects and Incremental Themes; they denote representations, of some kind or other, of the things mentioned in the following prepositional phrase. The prepositional objects in (23) and the direct objects in the respective paraphrases in (22) are not traditional Themes, in that they refer to things which are not effected (or affected) by the relevant action. They are however like Incremental Themes in an indirect way: since representations have parts which reflect the structures of the objects they represent, an incompletely produced representation may well be a representation of a proper part of the object to be represented, so the structure of the source object can be indirectly reflected in the event of producing the representation. For this reason, the representation-source arguments which are the sole object NPs in (22) act as de facto Incremental Themes there.

Some of the examples in the literature on aspect and aktionsart involve plural or quantified NPs in one or more positions that are like Incremental Themes in their apparent homomorphic relationship to events, yet a singular NP with the same verb does not seem to denote a (non-trivial) homomorphic semantic function:

- (24) a. John visited Atlanta  
b. John visited 25 cities (in two weeks)  
c. 2,500 tourists visited Atlanta (in two weeks).  
d. It took 15 tourists a half an hour to visit all 10 Photo Sites in the park.

That is, if we imagine the event in (b) to be interrupted without being completed, we might expect John to have visited some but not all of the 25 cities, and similarly in (c), that some but not all visitors made their visits; in (d), the completion of all 150 visits is at issue. But neither subject nor object works this way in (a). Apparently, a quantified NP argument along with almost any distributive telic verb (and some collective telics) can be understood homomorphically, because this combination generates reference to a set of individual events, one for each entity referred to by the quantifier. It is the

"meta-event" combining all these individual events that has sub-parts corresponding to the individual entities picked out by the quantifier NP. However, it turns out that only those NP arguments that are Incremental Themes even when singular will be relevant for argument selection---i.e. those cases where incremental themehood is entailed by the meaning of the predicate itself---so I will reserve the term *Incremental Theme* for that narrower class of cases, excluding ones like (24b-d).

Tenny (1987, 1988) has independently called attention to the way certain NPs "measure out the event" named by a verb, but it is clearly the same phenomenon as that discussed here, her term "aspectual delimitedness" replacing the more traditional term "telicity" used here. (Cf. also Hopper and Thompson 1980 and Rappaport and Levin 1988.) However, her description differs from the present one in that (i) she does not associate the phenomenon with thematic roles in general, but claims aspect as a semantic category is unique in this kind of syntactic association with arguments, which (ii) she maintains is exclusively with DIRECT OBJECT arguments ("internal arguments"), cf. Tenny (1987:179). I believe the latter is not correct, however. Transitive verbs like those in (25),

- (25) a. John entered the icy water (very slowly).  
b. The crowd exited the auditorium (in 21 minutes)  
c. Moving slowly but inexorably, the iceberg took several minutes to pierce the ship's hull to this depth.

and similar verbs such as *reach*, *leave*, *depart*, *abut*, *abandon*, have Incremental Theme subjects. Their meanings treat the stationary threshold or boundary traversed (and the direct object referent) as a line or plane rather than a region but allow that a space-occupying body (the subject referent) traverses it gradually, which means the subject is an Incremental Theme. Verbs like *cross* (*penetrate*, *permeate*, *pass*, *skirt*, etc.) allow either the thing traversed (as in (26a)), the moving body (26a), or both (26c), to be regions (and therefore Incremental Themes):

- (26) a. She crossed the desert in a week.  
b. At the turtle race, the winning turtle crossed the finish line in 42 seconds.  
c. It took Hurricane Archibald 3 1/2 hours to cross the Florida peninsula.

Also, as noted above, a prepositional phrase as well as a direct object can express an Incremental Theme (*She walked across the desert in a week*). Many intransitives also have Incremental Theme subjects, such as *emerge*, *submerge*, *deflate*, *bloom*, *vaporize*,

*decompose*, etc.<sup>2</sup> The phenomenon thus belongs in the realm of partial correlation of lexical meaning with argument configuration, not general compositional semantics. Despite the fact that Incremental Theme has not been counted within the traditional canon of thematic roles, I can see no good reason to exclude it if we begin from the position that any semantic factor which argument selection can be influenced by should be counted under this rubric: as will emerge even more clearly below, Incremental Theme is definitely in this category.

<sup>2</sup> I.e. the achievement verbs which entail a complex rather than simple change of state, also a subclass of the Unaccusative predicates (Rosen 1984). I call attention to the transitives with incremental subjects in (25) and not just the intransitives, because some will suggest that the subjects of the unaccusatives are derived by Unaccusative Advancement from underlying direct objects, hence at that level conform to the claim that all Incremental Themes are direct objects. This is less plausible for *John entered the water (gradually)*, which has a visible, independent direct object. Even here, of course, one can imagine it being suggested that *the water* originates as an underlying oblique and is advanced to direct object after *John* is advanced from direct object to subject. At that point, of course, one would have a right to ask whether the invariant association of Incremental Theme with syntactic direct object still had any empirical content or had been elevated from empirical hypothesis to methodological assumption, i.e. that one was in actuality prepared to postulate any syntactic abstractness necessary to maintain a uniform semantic association with a certain syntactic position---in other words, the methodology of generative semantics (and perhaps some contemporary theorists), where meaning is the decisive arbiter of the deepest underlying structure and indirect syntactic argumentation is sought post hoc to justify analyses suggested by such assumptions about semantic connotations of deep structure.

## 7. Thematic Roles as Prototypes

The hypothesis put forth here about thematic roles is suggested by the reflection that we may have had a hard time pinning down the traditional role types because role types are simply not discrete categories at all, but rather are cluster concepts, like the prototypes of Rosch and her followers (Rosch and Mervis 1975). And when we accept that arguments may have different "degrees of membership" in a role type, we can see that we really need only two role types to describe argument selection efficiently, which I will dub *Proto-Agent* and *Proto-Patient* (and below, simply *P-Agent* and *P-Patient*).

As preliminary lists of entailments<sup>1</sup> that characterize these two role types (i.e. lists of possible verbal entailments about the argument in question), I offer (27) and (28), without implying that these lists are necessarily exhaustive or could not perhaps eventually be better partitioned in some other way:

(27) Contributing Properties for the **Agent Proto-Role**:

- a. volitional involvement in the event or state
- b. sentience (and/or perception)
- c. causing an event or change of state in another participant
- d. movement (relative to the position of another participant)
- e. exists independently of the event named by the verb)

(28) Contributing Properties for the **Patient Proto-Role**:

- a. undergoes change of state
- b. incremental theme
- c. causally affected by another participant
- d. stationary relative to movement of another participant
- e. does not exist independently of the event, or not at all)

These lists bear significant resemblance to lists in Keenan (1976) and Keenan (1984) respectively but are interpreted differently here: on this see §10. I put properties (27e) and (28e), which Keenan includes, in parentheses, because I am not sure to what extent they should be attributed to the discourse associations of

<sup>1</sup> It is important here to distinguish entailments of the *predicate* from what follows from any one sentence as a whole (e.g., entailments that may arise in part from NP meanings, etc.): for example, if *Mary slapped John* is true, and John is a normal human, then slapping being the kind of action it is, we would conclude that John necessarily perceives something (and we would do likewise from the majority of sentences using *slap*). But it does not follow that the direct object of *slap* is entailed to have the P-Agent property of sentience, since we can also felicitously say *Mary slapped the table* or *Mary slapped the corpse*. However, the object of *awaken* does have the P-Agent entailment of sentience, as is revealed by the anomaly of *#Mary awakened the table/ the corpse*.

subjecthood mentioned earlier, rather than proto-role definition. On whether (28d) should be omitted from the Patient properties, leaving only its counterpart (27d), see §9.3.3.) Each of these characteristics (a)-(e) is hypothesized to be semantically independent, although of course most English transitive verbs have more than one such entailment for each argument: *build*, for example has all of (27) for subject and all of (28) for object. But English predicates can be found that I think show each Proto-Agent entailment separately (for its subject argument), thus justifying my including each separately, and illustrating in "pure" form the kind of entailment that I intend the labels (a)-(e) above to designate. (All also follow the argument selection principles to be given below.):

(29) Examples Illustrating Independence of Proto-Agent Entailments (in Subject NPs)

- a. *Volition alone*: John is being polite to Bill/ is ignoring Mary (cf. Dowty 1979)  
What he did was not eat [anything] for two days. (Cruse 1973:18)
- b. *Sentience/perception alone*: John knows/ believes/ is disappointed at the statement, sees/ fears Mary.
- c. *Causation alone*: His loneliness causes his unhappiness. Teenage unemployment causes delinquency.
- d. *Movement alone*: The rolling tumbleweed passed the rock. The bullet overtook the arrow. Water filled the boat. He accidentally fell.
- e. *Independent Existence*: John needs a new car.

Volitional action is familiar, but (29a) reminds us that deliberately REFRAINING from action is volitional also. (On occasion, being polite can mean deliberately doing nothing, remaining silent.) Sentience (which possibly should or should not be classed separately from perception) is found alone, cf. (29b), with the classic propositional attitude verbs, the stative perception verbs, and the stative psych predicates (i.e. *fear*, *be surprised at*, etc.). Sentience means more than a presupposition that an argument is a sentient being; it is rather sentience with respect to the event or state denoted by the verb: the objects of verbs like *elect*, *appoint*, *nominate* and *idolize*, *venerate* and *convict*, *acquit*, *exculpate* are necessarily human but not entailed to know or perceive the relevant event. Causation is almost always accompanied by movement, but stative causatives and perhaps generic causatives (29c) would fill this slot. On the other hand, movement is found without causation or volition (29d), viz. with inanimates or accidental movement; note that with *overtake*, *pass*, the object argument can move also and only be "stationary" from the faster first object's perspective.

"Independent existence" (29e) means the referent is *de re* (unless further embedded) rather than *de dicto*, i.e. non-specific, and is not brought into being or destroyed by the event named by the verb but is presumed to exist before and after the event. Though there are some verbs that entail subject existence but have none of (a)-(d), there are apparently with no verbs having any of (a)-(d) without entailing existence (for their subject) as well.

Proto-Patient entailments are harder to isolate entirely, but the following sentences indicate their nature reasonably well; this time the relevant entailments are for the direct object argument:

(30) Examples illustrating Proto-Patient entailments independently (in Object NP)

- a. *Change of State*: John made a mistake (*coming into being, therefore also (e) below*). John moved the rock (*indefinite change of position*) John erased the error (*ceasing to exist*)
- b. *Incremental Theme*: John crossed the driveway, filled the glass with water (*also stationary relative to other arguments*)
- c. *Causally Affected*: Smoking causes cancer.
- d. *Stationary relative to another participant*: The bullet entered the target, overtook the arrow.
- e. *Existence not independent of event*: John built a house, erased an error (*Coming into and out of existence; not independent of (a)*). This situation constitutes a major dilemma for us. John needs a car, seeks a unicorn, lacks enough money to buy it (*de dicto objects: no existence*)

Under "change of state" (30a) I intend to include coming into existence, going out of existence, and both definite and indefinite change of state. (Some but not all arguments of this type are Incremental Themes). Incremental Theme was discussed in §6. The next three entailments, (30c)-(30e), are the converses of Proto-Agentive entailments (29c)-(29e): if a verb has one of the first type for one argument, it necessarily has the corresponding one of the second type for another. (One reason for still recognizing both kinds rather than trying to collapse them somehow is to distinguish the P-Agent and P-Patient from the third argument of a three-place verb, as we will see illustrated in §9.3.) Under (30d), existence not independent of the event, I mean to include (i) verbs of creating and destroying, where this "effected" argument referent either does not exist before or will not exist after the

event denoted by the verb, and (ii) de dicto non-specific NPs, where no *particular* entity of this description need ever be fixed at all.

Is movement also to count as a change of state? If so, it seems that the above lists permit it to count as both agent and patient properties. No matter how movement entailments are to be precisely classified ultimately (a point to which we will return in §9.3.3), two things can be said. (i) Movement is an agent property only when not caused by another participant in the event named by the verb (*the cloud passed the tree, water filled the tank*), not when it is caused by something (*John threw the ball, The wall deflected the bullet*) or interrupted (*John caught the ball*): in this sense causation has priority over movement for distinguishing agents from patients. (ii) From considerations below (§9.3.3), it seems that movement usually only counts as a relevant change of state (i.e. a proto-patient property) when described as to or from a specified location (*put the book into the box, drive the hornets from the nest*).

Although we are using sets of entailments much like distinctive features to cross-classify arguments, I deliberately avoid saying "feature decomposition of roles" (as contrasted with Rozwadowska (1988), Zaenen (1988)), because I believe the boundaries of these kinds of entailments may never be entirely clearcut and I also would not rule out the desirability of "weighting" some entailments more than others for purposes of argument selection (as just mentioned with causation). Thus a cross-classification in terms of them will not be completely well-behaved in the way a true linguistic feature system will be. For example, the boundary of sentience is clouded by cases of computers or intelligent animals doing certain actions or in certain states that are stereotypically reserved for human, sentient participants, and such cases are reflected linguistically in *the machine switched itself off* (Cruse), *The dog believed you were a stranger, The program did that because it thinks you haven't saved the file first*, etc.; without delving into the philosophical questions these examples involve, I think it can be said that such language is not "wrong" or "metaphorical" for a certain kinds of sentient properties in limited situations but shows that sentience itself is something that different entities can have to different degrees. An unclear boundary of causation is the case of producing a change in a part of one's own body: in *I hurt my toe* there probably is causation, as this was an (unintended) result of some other act, but for *I raised my arm* (as compared with *I raised my glass*), the difficulties in calling this causation are well-known in the philosophical literature (what act did I do that caused my arm to rise?). The research of Talmy on "force dynamics" (Talmy 1985b) gives us further reason to worry about and maybe reanalyze the relationship of movement and causation beyond what is said in this paper.

The general point is that discrete feature decomposition has its proper place in describing syntax, morphology and phonology, because these domains are aspects of

the "coding system" of language at various levels and therefore in principle discrete. But semantic distinctions like these entailments, however, ultimately derive from distinctions in kinds of events found "out there" in the real world: they are natural (physical) classifications of events, and/or those classifications that are significant to human life. As such, there is no reason to believe all such classes must have discrete boundaries. Nor, I believe, is our cognitive ability to understand and recognize event classes limited to perceiving discrete types or those that cross-classify in some neat "grid" of semantic features or fields (which is not to say they NEVER classify this way). Much less are such classificatory schemes a preexisting universal mental mold which language forces us to categorize the world discretely and solely in terms of (pace some mentalistic linguists). If it turns out that our cognitive apparatus has evolved in such a way that something like an opposition between two Proto-Roles is a means of making a preliminary categorization of event participants for purposes of learning and organizing a grammar (a possibility explored in §11.-§13. below), then this does not affect the fact that the properties in (27) and (28) are significant because such categories of events are important to us in the first place and therefore to our cognition and our language secondarily, not vice-versa.

Furthermore, to the question once raised by Gennaro Chierchia (p.c.) whether in defining roles in terms of these entailments we would be replacing one unclear set of semantic primitives (the traditional thematic roles) by another just as unclear, I think the response is that these entailments are not any less clear, and more important, are more straightforwardly relevant to human life: while it is certainly not obvious that in ordinary reasoning and conversation people directly pay attention to or worry about whether something really was or was not a Theme or Source or Agent (in some sense of "Theme" etc. exactly as defined by Jackendoff or other linguist), but we certainly do concern ourselves all the time, both in everyday life and in courts of law, and sometimes to a painstaking degree, with whether an act was really volitional or not, with whether something really caused something or not, whether somebody was really aware of an event or state or not or had a certain emotional reaction to it, whether something was moving or stationary; and with whether something changed in a certain way or not, whether an event was finished or not, and whether an act produced something as a result or not.

## 8. Argument Selection

### 8.1 The Selection Principle and Corollaries

The way these proto-roles are involved in argument selection is given by the principle (31), which is to be understood so as to have the two corollaries (32) and (33) and the characteristics in (34):

(31) *Argument Selection Principle:*

In predicates with grammatical subject and object<sup>1</sup>, the argument for which the predicate entails the greatest number of Proto-Agent properties will be lexicalized as the subject of the predicate; the argument having the greatest number of Proto-Patient entailments will be lexicalized as the direct object.

(32) *Corollary 1:* If two arguments of a relation have (approximately) equal numbers of entailed proto-agent and proto-patient properties, then either or both may be lexicalized as the subject (and similarly for objects).

(33) *Corollary 2:* With a three-place predicate, the non- subject argument having a greater number of entailed proto-patient properties will be lexicalized as the direct object, the non-subject argument having fewer entailed proto-patient properties will be lexicalized as an oblique or prepositional object (and if two non- subject arguments have approximately equal entailed P- patient properties, either or both may be lexicalized as direct object).

(34) *Non-Discreteness:* proto-roles, obviously, neither exhaustively classify arguments (some arguments have neither role) nor uniquely (some arguments may share the same role) nor discretely (some arguments could qualify partially but equally for both Proto-roles)

Although using the traditional term "argument selection", I do not mean by "selection" a step that occurs during the derivation of a sentence (as in early Case Grammar), nor the linking-up of two different levels of representation, the syntactic level and the "thematic level" (the latter does not make any sense on the straightforward conception of monostratal syntax, and homomorphic (Montague-

style) compositional semantics assumed in this paper<sup>1</sup>). Rather, I mean a constraint on what kind of lexical predicates may exist in a natural language, out of many imaginable ones. Besides *build*, one can imagine a hypothetical basic (i.e. non-passive) verb meaning "is built by", i.e. a verb with the built as subject and the builder as object. But it is the consequence of (31) that the latter is not found while the former can be, and the phrase "be lexicalized as" is only a convenient locution for describing such constraints.

It should be noted that although I have used the term *prototype* in talking about roles, I am not suggesting that individual lexical meanings themselves are prototypes, in the way suggested in Rosch and Mervis (1975) or Lakoff (1977) or similar work. "Proto-roles", as I am using them here, are higher-order generalizations ABOUT lexical meanings (viz. "fuzzy" classifications of verbs by argument), not statements about individual lexical meanings, so the boundaries of individual word meanings can be as precise as you like, with definite criterial definitions. Note also only arguments, not adjuncts, are being classified prototypically<sup>2</sup>.

To see how these principles apply to verbs, note first that they imply that verbs in (35) should be the most stable in the lexicon in their argument pattern, since their subjects have several P-Agent entailments (volition, sentience, causation and movement) and no P-Patient entailments, while the objects have several of the latter (change, causally affected, and (mostly) incremental theme, stationary, dependent existence).

- (35) build (a house)  
write (a letter)  
murder  
eat  
wash (a plate)

Andrews (1985) calls attention to the "prototypicality" (in one sense) of these as

<sup>1</sup> To be sure, one could easily reformulate the claims of the present paper within a theory in which "semantic arguments" (or "semantic roles") of predicates were "linked" with grammatical relations in a way (partially) governed by the non-discrete role-types and selection principles of this paper, but to do so would in my view add conceptual baggage that is quite unnecessary and even obfuscating.

<sup>2</sup> If by *NP adjunct* we mean a phrase whose referent's relationship to an event is the semantically compositional result of applying that phrase's meaning to the meaning of any verb or VP (categorially, a "VP functor"), rather than a NP referent whose relationship to the event is defined by the verb's meaning itself (Dowty 1982), then any adjunct (like the instrumental *with a knife*) must have a constant meaning across every VP it occurs in. Thus there can be many kinds of meanings for "Patient", but only one for English instrumental *with*. (This view of course allows there could be different prepositions describing slightly different "kinds" of instrumentality, (cf. *with, by means of, through, etc.*), benefaction, etc. but each individually has the same meaning for every verb.)

*primary transitive verbs*, and Hopper and Thompson (1980) and others (cf. papers in Hopper and Thompson 1982) have pointed out consequences of such verbs being high on a scale of *transitivity*.

Combinations of certain P-entailments correspond to the familiar role-types (or often, to each of various conceptions of them). *Agent* is volition + causation + sentience + movement, or in some usages, just volition + causation or just volition (Dowty 1979), or, according to the ordinary language sense of "agent", causation alone. *Experiencer* is sentience without volition or causation. *Instrument* is causation + movement without volition or sentience. *Theme* (excepting Jackendoff's and Gruber's stative Theme) is most typically change + Incremental-Theme + dependent-existence + causally-affected, but causally-affected is sometimes absent (*Patient* can be distinguished from broader *Theme* by this entailment); Incremental-Theme is as we have seen sometimes absent from arguments called Themes, as is dependent-existence. But change alone is not really a sufficient criterion for this traditional role, as other participants too often move or otherwise change in events (Agents, Instruments, "Secondary" Themes), nor is any other one or group of these entailments: this points, I believe, to the traditional difficulty of tying down traditional Theme (or Patient) by any fixed criterion and the desirability of regarding this role in particular as a cluster concept instead. As this list indicates, these properties offer us instead of the traditional disjoint roles, broader and narrower semantic classes, which may be desirable for concerns like Cruse's about the need for various senses of "Agent" (cf. the four combinations above).

## 8.2 Role Hierarchies

Many of the other familiar relative ranking of the traditional role types in argument selections--as well as arguments that may fall "between the crack"---will follow: Not only do strong Agents outrank strong Patients, but both Instruments and Experiencers outrank any relatively patient-like argument for subjecthood, as in *The rock broke the window* and *John sees Mary*. In agreement with Fillmore (1968), an Agent outranks an Instrument. At least one P-Agent entailment, in the absence of any P-Patient entailments, is enough to qualify an argument for subject, and conversely with P-Patient entailments for object. The limiting cases of these situations---only one entailment of either kind---are in fact the example sentences in (29) and (30). Though the traditional *Source* and *Goal* are not really defined by any P-entailments, it nevertheless follows from the second corollary in (32) that Theme arguments will be direct objects while traditional Sources and Goals are obliques in many cases (but cf. §9.3), because Themes have more P-Patient entailments than these other arguments: cf. *John removed the lamp from the box* and *John put the lamp on the table*. The lamp undergoes a change of position and is causally affected, but the box and the table remain stationary and relatively

unaffected.<sup>3</sup> Thus the Proto-roles and their argument selection principle determine hierarchies of traditional roles such (36) (where ">" means "outranks for subject" and "outranks oblique for direct object") and additional rankings such as (37) (where "Arg" is an argument with no P-Agent and no P-Patient entailments):

(36)	Agent	>	Instrument Experiencer		Patient	>	Source Goal (usually)
(37)	causing even moving argument Experiencer		>		caused event Source, Goal, Arg Arg		

The point not to be missed here is that such hierarchies fall out of the two P-role definitions and the argument selection principle, whereas if Agent, Patient, etc. are introduced as primitives, then hierarchies must be stipulated additionally . (On the position of "Source" and "Goal" in hierarchies, see §11 and also §9.3.)

### 8.3 Argument Selection Indeterminacy

Lexical doublets like *buy* and *sell* or *like* and *please*, lexicalizations of the same relation (or almost the same) with different argument configurations, have been a puzzle for the argument selection problem. If selection rules should be formulated to give a single possible pattern for each verb, then these are counterexamples. But if selection principles are only tendencies admitting a small number of exceptions, then why do the multiple lexicalizations consistently appear in some semantic classes but never in others (e.g. never in the "primary transitive verbs")?

The selection principle (31) offers an explanation, since it permits alternate lexicalization in case of "ties" in Proto-Role entailments; arguably, this is a natural and not stipulative explanation under the Proto-Roles hypothesis: why shouldn't two lexicalizations be possible if there is nothing in the meaning of the verb to significantly distinguish the two possibilities in terms of the Agent-Patient

<sup>3</sup> It can be pointed out that the Source undergoes a change in that it no longer has the Theme in it, and the Goal undergoes a change in that it acquires the Theme in it; both these are caused as well. It may be important that the lamp here undergoes two changes (leaving its original position and assuming its new one), while the Source and Goal undergo one, or that the Source and Goal changes are otherwise less significant (on the relative importance of changes in different participants and its effect on argument selection, see §9.3.3), and it may be that insofar as the lamp's position on this path is an Incremental Theme, the lamp indirectly "counts" as one too for argument selection purposes. But in any event there are Theme-Source-Goal sentences with one added entailment that differ in argument configuration from these (cf. §9.3.1), so the difference in P-Patient values here cannot be too great.

continuum?

*Buy* and *sell* have already been mentioned as examples of verbs which do not distinguish their buyer and seller arguments by any entailments relevant to traditional roles, nor are they different in any proto-role entailments: cf. §3.2). Other such pairs are *borrow* and *lend*, and the two *rent*'s (*I rented it to her* vs. *She rented it from me*).

The psychological predicates (Postal 1970), or *mental verbs* (Croft 1986a) or *flip verbs* Rogers (1974) are another example of doublets but of an interestingly different sort. For convenience, I will call the subject of the verbs in the first column of (38) the *Experiencer* and the other argument the *Stimulus* (following Talmy's (1985b) terminology):

(38) *Psychological Predicates*

<i>Experiencer Subject</i>	<i>Stimulus Subject</i>
x likes y	y pleases x
x fear y	y frightens x
x supposes (that) S	(it) seems (to) x (that) S
x regard y (as) VP	y strikes (as) VP
x is surprised at y	y surprises x
x be disturbed at y	y disturb x

(*similarly, is astounded/dismayed/disappointed, etc.*)

What I believe sets this class of predicates off from all other natural language verbs is that (i) the predicate entails that the Experiencer has some perception of the Stimulus--thus the Experiencer is entailed to be sentient/perceiving though the Stimulus is not--and (ii) the Stimulus causes some emotional reaction or cognitive judgment in the Experiencer. The first of these is a P-Agent entailment for the Experiencer, while the second is a P-Agent entailment for the Stimulus argument.<sup>4</sup>

Moreover, these predicates have no OTHER entailments for either argument that are relevant to argument selection (with one possible exception to be discussed directly), which leaves a situation in which each argument has a weak but apparently equal claim to subjecthood. This contrasts with the *buy/sell* case in that here there are different P-entailment for each argument, but the selection

<sup>4</sup> This explanation for the occurrence of doublets in psych verbs was put forward in Dowty (1982a); Rozwadowska (1988) independently pointed to these two semantic properties of this class of verbs and used them to explain syntactic properties of the two arguments in nominalizations.

principle still give each out the same "count".

In a 1986 LSA paper, William Croft (Croft 1986) made an interesting further observation about this class of verbs: the Experiencer-Subject verbs of this class (left column in (38) above) are always stative, while the Stimulus-Subject verbs can be either stative or inchoative--i.e. describing the coming about of the perception and the consequent emotional or cognitive reaction. Even more interestingly, Croft claims that this restriction on the inchoative interpretation holds not just in English but in at least the three other languages he investigated (Russian, Lakhota and classical Nahuatl).

Note that the inchoative interpretation implies a change of state in the Experiencer (coming to experience an emotion or new mental state), but not necessarily any motion or other change in the Stimulus. (Suppose it is true that what happened was that the package in the back seat surprised John; it doesn't follow that the package did anything at all.) Therefore I would interpret the pattern Croft observed cross-linguistically as resulting from the fact that the inchoative interpretation entails a Proto-Patient property in the Experiencer that is not present in the stative: undergoing a (definite) change of state. Hence though the two arguments are still equal in Agent properties, they are unequal in that one is a "better" Patient, so it must be the direct object according to the selection principle (31).<sup>5,6</sup>

<sup>5</sup> Croft (1986a) proposed a different explanation in terms of causal chains, but perhaps the two are not really incompatible (cf. §11).

<sup>6</sup> There are of course well-known analyses of psych verbs in which the two forms of a doublet pair (e.g. Experiencer-Subject *be surprised at* and Stimulus-subject *surprise*) are derived from a common deep syntactic source and therefore not really a case of alternative lexicalizations---beginning with Chomsky's *Aspects* and Lakoff (1967), best-known in Postal (1970), many Relational Grammar analyses, and recently in Belletti and Rizzi (1986). This is not the place to make a meaningful comparison with these analyses, which are extensively developed but made in the context of specific theoretical assumptions I do not share, but perhaps two observations will be useful. First, note that it is the Experiencer-Subject form of the verb that is inevitably analyzed as "basic", the Stimulus-Subject form as derived. Cf. also Talmy's (1985b) observation that some languages (Atsugewi) have only Experience-Subject verbs as basic, the other class derived from these by lexical process. This may show that there is some sense in which sentience (Experiencer) outranks causation, even if it is not enough to block lexicalization of both forms in many languages.

Second, no matter how compelling the arguments may be that that Stimulus (= Theme) subjects of psych verbs behave like "derived subjects" (e.g. raised, passivized and non-thematic NPs) in English and Italian, while Experiencer objects are like underlying subjects, the deeper question which these accounts do not answer is why THIS particular class of lexical predicates should occur in these abstract underlying structures and appear in this surface alternation, while other classes of verbs (prototypical transitives like *kill*, statives, motion verbs, three-place verbs, etc.) never do. (Simply stipulating that it is verbs with Agent-Experiencer argument structures that have such properties is not much help; though this might identify just the right class extensionally, the traditional theory of discrete, "primitive" thematic role types in no way explains why this particular combination (rather than, say Recipient-Source) should have this constellation of syntactic properties.

The remaining question about argument selection principle indeterminacy that is of interest is whether any multiple lexicalizations are attested that are not predicted to be "ties" in argument ranking by these principles. I am not aware of any, and in §9. I will try to show that some apparent alternations of this kind ("symmetric" predicates and the *spray-load* alternations) are in fact not of this kind.

#### 8.4 Non-Standard Lexicalizations

As troublesome for the Proto-Roles Selection Hypothesis as unpredicted multiple lexicalization would be single lexicalizations that violate it. There is one relatively small group of such verbs that includes *receive*, *inherit*, *come into (an inheritance)*; *undergo*, *sustain (an injury)*, *suffer (from)*, *submit to*, *succumb to*, *tolerate* which seem to have Goals (*receive*, etc.) or Patients (*undergo*, etc.) as subjects, but Agents or causes as other arguments. Perhaps the appropriate comment is that these are in fact exceptions, but are few in number, so the selection principle is not an absolute rule but is nevertheless a strong tendency. However, it is noteworthy that almost all entail that their subject argument is sentient (for the relevant event). Of those that do not, I may be correct in sensing that their use with inanimates often sounds bookish and derivative of their animate use (*The car sustained/suffered little damage in the collision*, *The theory underwent a major reexamination*). *Receive* and *get* are other exceptions for which this is not apparently so: *The house received a new coat of paint*, *The play got a good review*. But *receive* is historically interesting in that the OED (i) lists citations for this verb which virtually all have human subjects, particularly before the 19th century, and (ii) implies that "active" senses of *receive* ("take or accept something willingly") are historically as common as "passive" senses (no volition on the part of the recipient implied). *Undergo* also has historical active senses ("submit oneself to") (and almost exclusively human, sentient subjects). *Get* has active meanings hard to disentangle from its non-volitional ones. If sentience were an actual entailment of the subject of a verb of this class, then this argument would have one P-Agent property as well as one or more P-Patient entailments. These observations may suggest (i) sentience might in some cases be a sufficient entailment to license an argument's lexicalization as

For mono-stratal syntactic theories which reject derivations altering grammatical relations, the challenge is of course to analyze the same data that appears in these arguments in terms of an account which exploits the relationship between syntax and semantics directly, e.g. one that points to the special anaphoric control properties of NPs in positions reserved for animate, sentient referents, whether they be grammatical subjects or objects, perhaps following the ideas of (Kuno 1987), and associates the anaphoric behavior directly to the semantic and pragmatic

subject, no matter how many P-patient entailments it has (in addition to other configurations possibly, cf. *receive* vs. *give*), and (ii) argument selection might be determined by a "core" use of a predicate, not entailments of its fully general meaning, and/or historical semantic drift can result in a predicate that violates selection principles.

### **8.5 Argument Selection in Ergative Languages**

A very important issue for the Proto-Role Hypothesis, which I can unfortunately make only brief mention of here, is argument selection in ergative languages. Ignoring the various kinds of "mixed" ergativity for the present, I want to focus on ergativity as found in the well-known case of Dyirbal (Dixon 1972) and in certain Mayan languages like Mam (England 1983) and Quiche (Trechsel 1982), in which the ergative-absolutive contrast is not only one of case marking or agreement but apparently the basis of syntactic organization throughout the grammar of the language, just as the subject-object contrast is for other kinds of languages. That is, absolutely-marked NPs "behave alike" in transitive and intransitive clauses for most syntactic purposes, while ergative NPs of transitive clauses (agent-like in meaning) are treated differently. Dixon (1979) has described this situation by classing absolutes as the "syntactic pivots" of such ergative languages, just as the nominatives (transitive and intransitive subjects) are the syntactic pivots of other languages.

Schmerling (1979), Dowty (1982a), and Trechsel (1982) have pointed out that if the categorial interpretation of grammatical relations suggested in Dowty (1982a, 1982b) is adopted, then there is every reason to simply identify "syntactic pivot" with the categorially-defined "subject", as the syntactic properties of these languages can then be described quite naturally. That is, an Ergative NP combines with a transitive verb to form a VP, having the syntactic and semantic properties of VPs in other languages. This means in effect treating the transitive "Patient" as a grammatical subject and the transitive "Agent" as analogous to an object (i.e. this is a form of the "inverse hypothesis" of ergative syntax, the idea of which is of course much older than this categorial interpretation).

Under this view, the argument selection principle in (31) cannot literally apply to syntactically ergative languages, but their argument pattern can be described with the same proto-roles and the same kind of principle, if we merely REVERSE the syntactic association: arguments relatively high in P-Patient entailments are syntactic pivots (categorial subject) and relatively agentive arguments are non-pivots (categorial object, here ergative NPs).

If the categorial inverse analysis of these languages is the correct way to proceed, this provides an extremely strong reason why we should not try to COLLAPSE the

notion of P-Agent with grammatical subject and P-Patient with grammatical object (or Absolutive), as Keenan (1976, 1984) has done, nor adopt a theory which necessarily correlates them in this single way. Another such reason, of course, is to properly distinguish the event-dependent role notions which are associated with lexical verbs from the discourse-dependent semantic associations of subjects (including subjects of passives, which are not Agents), as argued in §5. Rather, proto-roles and grammatical relations are distinct phenomena that languages must correlate consistently with one another, but in one of two possible patterns. Note that what we do NOT find, even in split ergativity, is "random" alignment from one verb to another, e.g. "build" with Agent absolutive, "kill" with Patient absolutive.<sup>7</sup>

considerations, not to an abstract syntactic level.

<sup>7</sup> This last hypothetical but non-occurring possibility must not however be confused with that of "active" languages like Lakhota (Boas and Deloria 1941), which make use of both nominative and accusative marking for intransitive subjects, allotting them verb by verb according to whether its (only) argument is more agent-like or

more patient-like: here the alignment of marking IS consistent in a certain way with meaning across all verbs (presumably), but intransitives are not marked like either transitive subjects or transitive objects consistently. (This situation is an instance of the "unaccusativity" phenomenon, for which see §12.)

<sup>1</sup> Note that many predicates have two arguments but the second is not a grammatical direct object but a PP, as in *rely on NP*, *suffer from NP*, *be afraid of NP*, *arrive at NP*: the selection principles apparently only govern argument selection for two-place predicates having a subject and true direct object. This will be important for understanding *Water filled the tank* vs. *the tank filled with water* or *Water poured into the tank* in § 9.3.2.

## 9. Systematic Semantic Contrasts in Multiple Argument Configurations

The most interesting data for the proto-role argument selection hypothesis, whose implications may go well beyond argument selection itself, comes from three cases of verbs that have two different possible argument configurations, correlating with a systematic semantic contrast that can be related to the argument selection hypothesis which has just been presented.

### 4.1 Partially Symmetric Interactive Predicates.

In the early days of transformational grammar, people supposed that (39) and (40) were transformational variants of the same deep structure (Gleitman 1969, Lakoff & Peters 1969): note that they seem to be synonymous (i.e. truth-conditionally so, ignoring differences in discourse function):

- |      |                       |               |
|------|-----------------------|---------------|
| (39) | This one and that one | rhyme         |
|      |                       | intersect     |
|      |                       | are similar   |
|      |                       | are alike     |
|      |                       | are equal     |
|      |                       | are different |
| (40) | This rhymes with      | that          |
|      | intersects with       |               |
|      | is similar to         |               |
|      | etc.                  |               |
|      | is different from     |               |

And this analysis was at first assumed to extend to cases like *John and Mary agreed* vs. *John agreed with Mary*, and, implicitly, *John and Mary kissed* vs. *John kissed Mary*.<sup>1</sup> These might not seem any less plausible than the cases above at first, but then Chomsky called attention to the example (41)<sup>2</sup>:

<sup>1</sup> I infer this from the absence of any mention of the agentivity problem in Gleitman (1965) or Lakoff and Peters (1969). Without comment, Gleitman mentions *collide* and *separate*, and Lakoff and Peters mention *agree*, verbs that, while not in the *kiss*-class, exhibit a similar asymmetry problem, as described below.

<sup>2</sup> The example is attributed to Chomsky (personal communication) by Quang (1970:33).

- (41) a. The drunk embraced the lamppost.  
 b. #The drunk and the lamppost embraced.

The oddness in (41b) is of course that it implies that the lamppost somehow took part in the act of embracing. Once we see this, it suddenly becomes quite apparent that *John and Mary kissed* is not really synonymous with *John kissed Mary* either: the same asymmetry in who is responsible for the action appears there too (though I think it is interesting that, in my experience, people do not usually notice this fact until one points out (41) to them). It was soon discovered that this difference in agency was found with a whole set of verbs which Fillmore once called verbs of *partially symmetric human interaction* (Fillmore 1966, Quang 1970, Dowty 1972, 1979):

- (42) Kim and Sandy      hugged  
    embraced  
    kissed  
    made love  
    fucked<sup>3</sup>  
    talked  
    disagreed (?)  
    shook hands (?)

- (43) Kim                      hugged                      Sandy  
    embraced  
    kissed  
    made love to  
    fucked  
    talked to  
    disagreed with (?)  
    shook hands with (?)

Now although the symmetrical examples in (39) and (40) are all stative, it should not be assumed that all agentive, NON-stative verbs do have the asymmetry: for example those in (44) are agentive, but there is no (truth-conditional) asymmetry in agency between the sentences in (44a) and their counterparts in (44b) above.

<sup>3</sup> Special semantic properties of this verb and its synonyms have been examined in a celebrated study by Quang (1970).

- (44) a. Kim and Sandy    married  
    played chess  
    debated  
    discussed the matter
- b. Kim                    married<sup>4</sup>                                    Sandy  
    played chess with  
    debated  
    discussed the matter with

The relationship among the three classes seems to be as follows. Marrying, playing chess, debating and other such activities (e.g. *fighting*) are actions that by their nature require the volitional involvement of two parties: one can't understand the essential nature of these actions without knowing that. By the same token, volition is irrelevant to whether the stative relations in (39)-(40) obtain. The relations in (42)-(43) denote actions that differ from both in that most of their criterial properties by which they are recognized are symmetrical with respect to the two participants (e.g. being in a certain kind of body position with respect to the other), yet the relation may involve volition on the part of either one or of both parties, without the language, as it were, feeling the need for "independent" (more neutrally, "unrelated") lexemes to distinguish such subcases.

As volition is a P-Agent entailment, all three of these patterns are syntactically consistent with the selection principle: if volition is entailed at all, it is entailed for the subject argument; there is no verb that entails volition for object but not subject, (nor, as far as I can tell, one that entails that at least one of the participants is volitionally involved but does not indicate which, either in the transitive or the collective intransitive form). And---though this claim is perhaps harder to verify---it seems that every verb describing a kind of relation that COULD sensibly be understood as volitional for either one or both participants but is otherwise symmetrical in meaning DOES exhibit this alternation.

A different situation is presented by (45):

- (45) a. The truck collided with the lamppost.  
       b. (#) The truck and the lamppost collided.

(45b) might seem like a bizarre sentence but in fact would be perfectly natural to describe a situation where a new lamppost was being carried to the top of a hill, came loose from its moorings, rolled down the hill, and intersected the path of a

<sup>4</sup> The relevant reading here is the one in which Kim is a marriage partner, not the official who performs the ceremony.

moving truck at the bottom. Thus the difference here is that (45a) entails only that the truck was in motion in the event of collision, while (45b) entails that both the truck and the lamppost were, though the nature of the event is otherwise similar, e.g. entailing forceful impact between the two and suggesting damage to one or both. The pattern is like (42)-(43), but the entailment that distinguishes subject from object of an otherwise symmetric predicate is not volition, or any standard concomitant of traditional Agency, but rather motion. (Note that neither truck nor lamppost is being "personified" here, as would be the case, for contrast, in the agency-imputing "active *be*" that occurs with adjectives, as in #*The truck is being dangerous*, #*The lamppost is being collision-prone*.) Further examples are:

- (46) a. The ship passed the lighthouse in the night.  
 The snake separated from its skin.  
 The ivy gradually intertwined with the trellis.
- b. (#) The ship and the lighthouse passed in the night.  
 (#) The snake and its skin separated.  
 (#) The ivy and the trellis gradually intertwined.

Therefore, one cannot try to analyze this relationship in (45)-(46), as did Quang (1970) and Dowty (1972, 1979 ch. 2) for examples in (43), by postulating an abstract operator DO of "Agency" that takes scope over both NPs in the (a) sentence or only one in the (b) sentence: this would get the semantics of (46) wrong. Nor does any one traditional thematic role unite these two cases, as the subjects of the second group are presumably Themes, not Agents.

If there IS a single linguistic generalization that covers both examples like (43) and ones like (45)-(46), then it seems that only something like a Proto-role hypothesis can provide: though there are not necessarily any differences between the entailments of a collective-subject predicate (i.e. with conjoined NP or plural NP as subject) and the two-place lexically identical version of the same predicate (cf. (41), (44)), if there is a difference it will apparently be that the collective subject version has some Proto-Agent entailment for both (all) the subject-denotations that the two-place version lacks for its object-denotation.<sup>5</sup>

<sup>5</sup> Note incidentally that it is not necessarily the case that a collective-subject predicate must always entail exactly the same thing about all the members of its subject-denotation. For example, *The students in my class voted to adopt the proposal* (an example due to William Ladusaw, cf. Dowty 1986) entails that at least 51% of the individual students cast votes for the proposal but does not say how the other 49% might have voted or indicate which were the affirmative voters. Thus it seems conceivable that *John and Mary kissed* might have meant only that at least one of the two was volitionally responsible for the kissing event. But no verb of this class has such a meaning, as far as I know. Similarly, it is not the case that literally every collective-subject verb entailing motion entails that all members of the subject denotation must move: *All the students gathered in the*

The systematicity of these patterns, combined with their subtlety, raises the question whether the etiology of this phenomenon is slightly deeper than a range of diverse lexical items conforming to a universal lexicalization tendency. To put it in terms of the learning of lexical meanings, do speakers of English really learn the semantic difference between the (a) versus (b) patterns of a dozen or more verbs like (39)-(40) individually, by observing the semantic difference between uses of the two forms for each of a dozen or so verbs, and similarly for a group of motion verbs like those in (45)-(46)? Or does the Proto-Role alignment principle play an active, causal role in this learning: that is, when confronted with a predicate denoting a kind of event that sensibly CAN be understood as either symmetrically or asymmetrically volitional (or motional), does the learner AUTOMATICALLY assume that the collective-subject version is symmetrically volitional (or motional), the two-place version asymmetrically volitional, without requiring any specific empirical data to that effect (and similarly for symmetrically and asymmetrically motional)? If so, then the Proto-roles and their alignment principle would be functioning as a kind of "semantic default" for the learning of lexical meaning. We will return to such this question in section §11.

*hall after the class ended*, for example, could be true if some of the students were already in the hall before class ended and simply stayed in place, while the rest came there, and *The crowd dispersed* is true when enough individuals have left that the people remaining no longer constitute a crowd. Therefore, to try to explain away the generalization by saying the child learns these cases by assuming there is a lexical rule deriving a collective intransitive verb  $V_2$  from a homophonous transitive  $V_1$  such that  $A \text{ and } B V_2$  means the same as  $A V_1 B \text{ and } B V_1 A$ , and all collective intransitives of this class involve this rule, is to beg the deeper question why the child should not instead assume a rule giving  $A \text{ and } B V_2$  the slightly weaker meaning *One of A and B V1 the other*.



## 9.2. Inchoative Interpretations in Stimulus-Subject Experiencer Verbs.

The second case of lexical pattern conforming to the selection principle in a subtle way has already introduced: this is Croft's generalization (cf. §3.4. above) that an inchoative interpretation is possible in a Stimulus-subject psychological verb but never in an Experiencer-subject verb: the progressives and clefts in (47) and (48) are diagnostics for such an interpretation (Dowty 1979:163-165):

(47) *Stimulus-subject psych verbs (non-stative contexts)*

- a. The birthday party is surprising/pleasing Mary (right now)
- b. What happened to Mary was that the birthday party surprised/pleased her.

(48) *Experiencer-subject psych verbs (non-stative contexts)*

- a. \*Mary is being surprised at/ is liking the birthday party (right now)
- b. \*What happened to Mary was that she was surprised at/ liked the birthday party.

Recall that this was observed to agree with the Proto-Role selection hypothesis, since change-of-state (which the Experiencer undergoes in these cases, not the Stimulus) is the Proto-Patient property which tips the scale in the direction of the Stimulus-Subject/Experiencer-Object form.

Once again it seems natural to ask whether a difference which occurs systematically across all the psychological verbs in the lexicon, and in all the four languages Croft observed, is the result of (and is perpetuated solely through) independent learning for each such verb, i.e. is coincidental, or whether the Proto-Role selection principle could somehow act as a semantic acquisition default to facilitate conformity to this pattern.

### 9.3 Alternations in Direct versus Oblique Objects

To see how syntactic patterns of alternation between direct and oblique object relate to the proto-roles hypothesis, we will distinguish among four semantic subtypes: alternating *load*-type verbs, non-alternating *fill*, *hit*-type verbs, and representation-source predicates (*photograph a landscape*).

#### 9.3.1 The *spray/load* cases.

The venerable examples involving alternations of direct and prepositional objects with the verbs *spray*, *load*, *smear*, etc. have a long history in modern linguistics (beginning at least with Hall 1965). As early as 1971 it was pointed out by Anderson (1971), though also hinted at by Fillmore (1971:368) as well, that the pairs in (49) and (50) are not complete paraphrases, but rather the (a) sentence suggests the total supply of hay or paint is affected, while the (b) sentences suggest that the cart is completely filled or that the wall is fully covered with paint:<sup>1</sup>

- (49) a. Mary loaded the hay onto the truck.  
b. Mary loaded the truck with (the) hay.
- (50) a. Mary sprayed (the) paint onto the wall.  
b. Mary sprayed the wall with (the) paint.

If this claim is correct, then such examples are another case of semantic variation across multiple argument configurations of the "same" predicates that is consistent with the proto-role hypothesis and argument selection principle: the proto-patient entailment of Incremental Theme is always an entailment of the actual direct object in these cases (as some would say, the "surface" direct object), no matter which of the two NP appears in this syntactic position. An Incremental Theme, it will be recalled, is a NP that can determine the aspect of the sentence, since the parts of the event correspond to parts of the NP referent that are affected by the action; the event is "complete" only if all parts of the NP referent is so affected (or effected). The event of loading the truck with hay is partially or completely done, according to whether the truck is partially or completely full of hay, but the event of loading the hay onto the truck is partially or completely done according to whether the quantity of hay in question is partly or completely on the truck (regardless, in the last instance, of whether this completely fills the truck or not).

Again, this is a candidate for a semantic default phenomenon: The difference

<sup>1</sup> Recent articles that have also been concerned with this difference include Hopper and Thompson 1980, Rappaport and Levin (1988), Tenny (1987, 1988).

seems both subtle and systematic across verbs, thus perhaps unlikely to have been learned individually verb by verb, and it is a difference consonant with the proto-role definitions and selection principles.

Notice the difference between this way of talking about roles in (49)-(50) and the more traditional one: in Jackendoff's and Gruber's terms, *the hay* is the Theme in both (49a) and (49b), presumably because it is "the thing which moves." Still, another traditional sufficient criterion for Theme-hood is "thing which undergo a change of state", e.g. *the house* in *John painted the house* is counted as Theme, as do many effected and affected objects, though the house doesn't move. Thus in this class of events, TWO things undergo the kinds of change of state that are, at least sometimes, sufficient to qualify them as Themes: there is an ambiguity even in the traditional assignment criteria for this class of verbs.

Note that an assumption in my discussion is that the two different subcategorizations for such a verb correspond to different meanings for the verb that are recorded as independent items in the lexicon (or, as distinct though related ones, perhaps connected by lexical rules<sup>2</sup>), **not** to two different surface structures derived from the same deep structure. If one wanted to adopt the latter approach, it would be necessary to postulate a surface structure interpretation rule to determine the aspect of the sentence. (Anderson (1971), in fact, proposed such an analysis.) The main reason for rejecting that approach is that not all verbs which show the alternation in syntactic configuration have such a difference in aspectual meaning (as we will see with verbs like *hit* in §9.3.3. below), so the phenomenon cannot be a general compositional semantic one associated with direct objects of three-place verbs. Conversely, not all incremental themes are direct objects: as mentioned in §6, subjects, pairs of PPs, or verbs alone sometimes verbs alone can "encode" incremental themehood. The additional fact

<sup>2</sup> That is, I am assuming that the relationship between the verb in *load the truck with hay* and that in *load hay onto the truck*, like that in the collective-subject-alternation (§9.1) and the psych-movement (§9.2), is to be described by a lexical rule in the sense of Dowty (1978, 1979: chapter 6). See these works for a full discussion, but briefly, lexical rule in this theory is one which supplies a hypothetical derived lexical item and a (rule-predicted) hypothetical meaning for it for each word in its domain; some of these possible lexical items are (individually) learned to be actual ones by a speaker, and the speaker learns an actual meaning for each which is usually similar to but can differ unpredictably from the meaning given by the lexical rule (e.g. *readable* means more than "capable of being read"). I assume such rules include not only word-derivation cases (*decision* from *decide*) and zero-derivations (noun *walk* from verb *walk*) but also "lexical" phrases (*egg on* or *hammer flat*) and changes in valence, including detransitivizations and the changes in argument configuration discussed in this paper. I will assume for purposes of this paper that the lexical rule for *load*, etc. would itself probably give a meaning for derived *load* such that *load the truck with hay* is indeed the same as *load hay onto the truck*, and that such semantic differences as are noted below for some but N.B. not all instances of this pattern arise because speakers often acquire an actual meaning for the derived lexical item that differs slightly from the lexical-rule predicted interpretation but on the other hand conforms more closely to entailment patterns fitting the Proto-Role selection principles.

which is of course suggestive of a lexical phenomenon is that not all verbs which do have the Incremental Theme interpretation of the direct object participate in the syntactic alternation (cf. e.g. *cover* and *fill*, discussed below).

All of this, of course, assumes the aspect/aktionsart difference Anderson and Fillmore intuitively felt in these examples is correct. Today, aspect and aktionsart are better understood than in 1971, so we should be able to back up their intuitive observations, if correct, with known semantic diagnostics for aspect. One's first intuition about such aspect differences is, after all, not always reliable.<sup>3</sup> I believe this aspectual claim IS correct, though there are a number of complications to be dealt with in order to demonstrate that this is so.

First of all, the examples one often sees have a bare plural or mass term in one of the relevant NPs, or the determiner parenthesized in the examples comes and goes sporadically and without comment. Yet it is independently known (Verkuyl 1972) that such a NP can make an otherwise telic sentence behave like an atelic (or durative); cf. §6 above. In order not to be manipulating two variables at once (the *with* vs. *onto* alternation and the definite/bare mass noun distinction), let us avoid bare plurals and bare mass nouns for a moment and use only definite NPs.

Consider first sentences with *complete* or *finish* and what one can conclude from them (cf. Dowty 1979, pp. 57, 181); these of course entail a perfective interpretation of some kind or other:

- (51) a. Mary completely loaded the hay onto the truck.  
b. Mary completely loaded the truck with the hay.

Suppose we ask in each of these two cases, the questions in (52):

<sup>3</sup> For example, Tenny (1987:156) asserts, without applying any of the standard aspectual diagnostics or giving any other semantic justification, that *John shaved himself, dressed himself, bathed himself* are telic (in her term "describe a delimited event"), while *John shaved, dressed, bathed* are atelic ("non-delimited"). But I can find no corroboration for such a difference by the usual tests, and when I tried putting such predicates in standard diagnostic frames and querying several English speakers about this data, I found no consistent judgment that the reflexives are interpreted more telically than the intransitives. Conceivably there really is some kind of ephemeral aspectual distinction here, but it is apparently not the normal telic vs. atelic one: the point is that raw intuitions about an isolated example are not a satisfactory guide to aspectual analysis, particularly in view of the familiar problem that in English most lexical predicates are themselves ambiguous (or indeterminate) in telicity, with pragmatics often making one or the other possibilities prominent (Dowty 1979:60-62). (NB that Tenny employs reliable diagnostic tests elsewhere in Tenny 1987.)

- (52) a. Was all the hay put onto the truck?  
 b. Was the whole truck full of hay?

For statement (51a), the answer to the first question is "yes" and to the second question is "not necessarily." This implies the NP *the hay* is an Incremental Theme and the NP *the truck* is not. Conversely, if the NP *the truck* really is an Incremental Theme in (51b), then the answers for that statement should be "not necessarily" for question (52a) and "yes, definitely" for (52b). But in fact, the answers here are not so clear. Some people are inclined to say that (51b) is really only appropriate if the quantity of hay is such as to fill the truck exactly, so no hay or truck space is left over. (Others say 52b is an odd question here.) Notice, though, that if we change the statement by reintroducing a mass term in the non-direct-object position, as in (51b'),

- (51) b'. Mary completely loaded the truck with hay.

then the sentence is more natural, the answer to (52b) is yes, and question (52a) now makes little sense, since no particular quantity of hay seems to have been referred to (except, maybe, just the quantity that DID end up on the truck, but (52a) has a totally trivial yes answer on that interpretation). This situation IS perfectly consistent with the hypothesis that *the truck* is the only Incremental Theme in the sentence, since an Incremental Theme but not necessarily other arguments must be definite for a sentence to be understood in perfective aspect (as *completely* requires). By contrast, (51a') is anomalous, as it should be if the direct object is Incremental Theme and if we follow Krifka (1987, 1989) in treating telics as homomorphisms from such arguments into events:

- (51) a'. #Mary completely loaded hay onto the truck.

What I think complicates the situation with the original (51b) is that one inevitably takes into account not just the literal meaning of the sentence but also interprets them in light of the purposes people have in performing the actions they do. The purpose of trucks and carts is to move stuff around (we don't generally acquire the stuff just to get the carts and trucks filled), and since (51b) mentions a definite quantity of stuff, it is natural to take the overall purpose of Mary's action to be transporting that quantity of stuff somewhere. If the truck is full but part of the stuff is left over, then in a broader sense Mary's work probably is not finished. On the other hand, we are not generally so bothered with extra space left over in a cart or truck if all the stuff we want to move is loaded inside. Contrast the above with (53):

- (53) a. Mary completely sprayed the wall with this can of paint.  
 b. Mary completely sprayed this can of paint on the wall.

- (54) a. Was the wall completely covered?  
 b. Was all the paint used up?

In (53a), we have little reluctance to say the task is complete if the wall is covered but there is still paint left. But a difference between painting walls and loading trucks is that the purpose of the former is most usually to get the wall covered---not just to move the paint around---hence having paint left over is generally not the problem that having stuff left over after the truck is loaded often is. Still, (53a) might be understood another way: imagine that the paint has a chemical in it which repels termites if the paint is applied in sufficient thickness, that Mary's purpose is not to achieve a particular color or appearance on her basement wall but to achieve adequate termite resistance in it. Then if the paint was just the quantity needed for adequate termite protection, we might well not regard the action done until all the paint was used up, even if the whole surface had been covered by at least some paint. This interpretation is actually also consistent with the proto-role hypothesis, I believe, because the wall still undergoes a definite change of state (becoming sufficiently protected), but this points up another difference between loading trucks and painting walls: though one normally does stop painting a wall after it is completely covered, one actually can go on putting paint on it indefinitely---in a way that one can't keep on loading a truck after it's fully loaded. In other words, *spray paint on the wall* (as well as *spray the wall with paint*) can have an atelic (or *activity*) sense as well as a telic (*accomplishment*) sense.

This observation is relevant to applying other aspectual tests to these sentences, such as *for an hour* vs. *in an hour*. The former is a durative adverbial and is only intelligible with a predicate that can have an atelic (activity) reading. The latter by contrast only occurs with a telic reading of a predicate (i.e. accomplishment, achievement or inchoative). Cf. Dowty 1979:56-64,332-336, 340-348) for discussion of these tests. Now Verkuyl's observation was that a bare plural or mass term put in certain syntactic positions makes a telic predicate into an atelic one. So if our hypothesis is correct that the direct object is always the Incremental Theme in such examples, then changing THIS NP from definite to bare plural or mass should alter the telicity and hence the adverbial possibilities, while altering the OTHER object NP should not. First, the definite NP *this wall* is alternated with mass term *paint*:

- (55) a. John sprayed this wall with paint in an hour.  
 b. (#)John sprayed this wall with paint for an hour.  
 c. #John sprayed paint onto this wall in an hour.  
 d. John sprayed paint onto this wall for an hour.

We predict that (b) and (c) should be bad, while (a) and (d) are good. This is

borne out in (a) and (d), as well as (c) (though (c) does have a marginal but irrelevant inchoative reading "it took an hour for John to start spraying the paint"), which is in accord with the hypothesis because an inchoative is a kind of telic reading). But (b), which should be bad, sounds OK. The reading it has, however, is the aforementioned atelic or activity one: it's not possible to understand the event to be described here as having an inherent completion point, either in terms of getting the wall to a definite state or the paint used up.

We now test examples where the NP referring to the paint mentions a quantity but the NP referring to the place does not: recall that bare plurals---here *subway cars*--- are just like bare mass terms in their effect on aspect:

- (56) a. #John sprayed subway cars with this can of paint in an hour.  
 b. John sprayed subway cars with this can of paint for an hour.
- c. John sprayed this (whole) can of paint onto subway cars in an hour.
- d. #John sprayed this (whole) can of paint onto subway cars for an hour.

As predicted, it is again the direct object NP that controls the acceptability of the time adverbial, never the other NP. Notice that example (d), which is parallel to the problematic example (51b) above, is clearly anomalous and does not admit an atelic (activity) sense instead, as (51b) did: while one can keep putting more paint on the same wall, even after it's covered, there's no (normal) way to keep putting the same quantity of paint on something over and over again. ((56a) does have the marginal inchoative reading on which it makes sense.)

Incidentally, one should not be misled by the absence of detectable aspectual difference between the ablative and abstrument examples like the following (*Abstrument* is a term Hook (1983) coins for the role of the NP marked with *of* below, by analogy to "ablative" and "instrument"; an abstrument is an oblique NP denoting the thing or substance removed from a space):

- (57) a. John stripped the bark from the tree.  
 b. John stripped the tree of (its) bark.

(similarly *wash sand off the beach/wash the beach of sand, empty water from the tank, empty the tank of water*, etc.)

In these cases, the quantity of space is originally occupied by a certain quantity of stuff: removing all the stuff from the space entails vacating all the space, and conversely, vacating all the space entails removing all the stuff. Only with converging predicates (like *load, spray*) can one sensibly use up all of a

preexisting quantity of filler without filling up all the space--or conversely--, and thereby detect an aspectual difference between the two forms. This difference is a consequence of the physics of space (e.g. one could not have started with a space "occupied" by a quantity of stuff larger than the space would actually hold, and then vacate all the space without removing all the stuff), not anything exceptional that needs to be explained about the semantics of English. (Here again, a failure to recognize the contribution that facts about the world make to "meaning" could make the linguist's task seem harder than it is.)

### 9.3.2 Lack of alternation in objects with *fill* and *cover*

As Mellema (1974), Fillmore (1977), and others have observed, the verb *cover* does not participate in this same direct object alternation that *spray*, *load*, etc. do:

- (58) a. Bill filled the tank (with water).  
b. Bill covered the ground (with a tarpaulin)
- (59) a. \*Bill filled water (into the tank).  
b. \*Bill covered a tarpaulin (over the ground).

Aspectual tests will confirm that the direct object, not the prepositional object, is an Incremental Theme in (58), so these verbs are in accord with the Argument Selection Principles.

One might wonder WHY (if indeed for any reason) these verbs do not alternate in this pattern, as *spray* and *load* do. It is tempting to respond that intuitively, the notion of producing a completely occupied space or a completely overlaid surface or opening seems fundamental, a "core" part of the meanings of these verbs, in the way that a completely affected space is not for *spray* or *load*. But then a deeper question is why, if for any reason, these particular verbs should differ from others in this respect. It is apparently not because of an inherent semantic incompatibility, for English-speaking children have been observed to temporarily produce examples like *I filled water into the glass* (cf. Bowerman 1982, Pinker 1989:25,26), so a theory should not predict that the other syntactic form (and meaning, presumably) is impossible. One might speculate that the existence of the morphologically related adjective *full* (and historically the root of *fill*, which is a causative) and noun *cover*, both entailing complete occupancy/coverage of the space

in question<sup>4</sup>, help maintain the restriction of these verb meanings to a locative Incremental Theme and have prevented the child's temporary innovation from surviving into the adult language, over the many centuries these forms have existed in English.

The two-place forms of these verbs, (60)-(61),

- (60) a. Water filled the tank.<sup>5</sup>  
a'. \*Water filled into the tank.  
b. Snow covered the ground.
- (61) a. The tank filled with water.  
a'. \*The tank filled water.  
b. \*The ground covered (with) snow.

illustrate not only that a subject can sometimes be an incremental theme, as in (61a), but also that the subject can ONLY be the Incremental Theme when the verb is intransitive (*The tank filled*) or expresses its other argument via a prepositional phrase rather than a grammatical direct object, (61a). (Conversely, the Incremental Theme is not happy in a prepositional phrase, as (60a') shows.) It is cases like this that show that the argument selection principles must be formulated to require only that grammatical direct objects have more proto-patient entailments than subjects, not that any non-subject argument be more patient-like than the subject.

<sup>4</sup> By contrast, the cognate noun *load* need not always refer to stuff filling a fixed, predetermined amount of space, as shown by *She carried a load of books with her*, so any implication to that effect, e.g. in, *We need three loads of gravel for this job*, is probably implicature. A *cover*, on the other hand, is always something completely overlaying (or surrounding) some relevant other object, or something originally constructed or intended to do so. The suggestion here, put in terms of the hypothesis mentioned in §9.1 and discussed in §10 that argument selection principles can act as defaults in language acquisition, is that the association of Incremental Theme entailment with the locative argument might be made so vivid by *full* and noun *cover* that this association is individually learned for these verbs and thereafter immutable, while for the *spray/load* class it is not individually learned but supplied as a default entailment of direct objects in each of the two syntactic configurations the verbs occur in, giving rise to the slight alternation in meaning between the two configurations. But once the entailment of *fill* and *cover* with respect to their locative arguments is "frozen", the selection principles would then permit them to occur only in the form in which the direct object is the locative argument.

<sup>5</sup> This example looks similar to cases like *The crowd entered the auditorium*, for which I earlier claimed that the subject *the crowd* could be the Incremental Theme. But there is a subtle difference, as can be seen by comparing *The crowd entered the auditorium halfway/partly* with *The water filled the tank halfway/partly*: from the former, we can conclude that some percentage of the crowd has entered, but we cannot conclude anything about the percentage of the auditorium that is occupied. On the other hand, from the latter we can conclude that some percentage of the tank is occupied, not that a certain percentage of some quantity of water is in it; thus *the tank* in (60a) is the Incremental Theme.

### 9.3.3 Hitting versus Breaking

In another classic article, Fillmore (1970) pointed out that there are a number of verbs of physical contact such as *hit* which yield (truth-conditionally) synonymous alternations of their direct objects with their prepositional objects as in (62),

- (62) a. John hit the fence with the stick.  
b. John hit the stick against the fence. (= 62a)

while on the other hand there are verbs like *break* that do NOT yield synonymous alternations:

- (63) a. John broke the fence with the stick.  
b. John broke the stick against the fence (≠ 63a)

Fillmore observed that (i) *break* entails a visible and permanent change of state in its direct object argument (while *hit* and similar verbs do not), and (ii) this change of state is entailed for *the fence* in (63a) but for *the stick* in (63b).

What is of interest to us is the negative generalization (not explicitly drawn in the early literature, as far as I know): there are no verbs which are like *break* in entailing a change of state for only one of its non-subject arguments but which produces a synonymous alternation between (a) and (b) forms like those above. Rather, such a change-of-state entailment argument is always entailed for the direct object argument, never for an oblique object argument. This generalization would in fact be mandated by the proto-role and argument selection principle: assuming the number of other proto-patient arguments for the two non-subject arguments is otherwise equal, a change-of-state entailment for one argument but not the other would, according to the selection principle, make the first outrank the other for direct object status.

But what of the *spray/load* class, which does allow both syntactic configurations? Isn't a change of state entailment involved here? The important difference, I believe, is that with these verbs BOTH non-subject arguments are entailed to undergo significant changes of state: in loading a truck with hay, the hay changes location, but the truck also changes from an unloaded to an loaded state.<sup>6</sup>

<sup>6</sup> By "significant", I mean that this change of state is different from, for example, the (semantically well-defined) change that a place necessarily undergoes when another object is moved from or to it, as for example the locative argument in *put the book on the table* or *arrive at the city*: we do not ordinarily classify tables according to whether an object has moved onto it or not, or cities as to whether someone has arrived at them or not, in the

Note that the *hit* class verbs denote events that are not aspectually subdividable vis-à-vis either object argument: If the act of hitting the fence (once) with the stick is interrupted but not completed, it's can only be because the stick has not yet contacted the fence at all, not because only part of the stick has hit the fence or because only part of the fence has been hit. That is, neither object can be a (non-trivial) Incremental Theme, so there is no semantic alternation here in Incremental Theme interpretation of the kind seen with *spray/load*.

Thus the properties of the three classes of verbs discussed are as follows:

- (64) I. *spray/load* class
- a. entail change of state in both arguments (N.B. different changes of state), either could potentially be Incremental Theme (the "measure" of the event)
  - b. appear in both syntactic patterns, but with slight change of meaning, viz. in Incremental Theme, which is always direct object argument; other entailments "alternate" with change in syntactic pattern
- II. *break* class
- a. entails change of state (and Incremental Themehood) in only one argument
  - b. radical change in meaning from one pattern to the other: change-of-state is fixed with direct object, and other entailments alternate.
- III. *hit* class
- a. No difference in proto-role entailments between arguments (but concerning motion, see below)
  - b. complete synonymy between two patterns: all entailments alternate.

same way as we classify trucks as to whether a load has been put in them or not or not, walls as to whether paint has been applied to them. See the discussion of relative "significance" of changes of state in (62) vs. (65) vs. (66) below.

These patterns are all in accord with the selection principles. Note also that it is difficult to see how a (semantically non-ad-hoc) classification in terms of "atomic" thematic roles could combine with an argument selection principle to describe these classes economically: My description crucially relies on the fact that traditional "Theme" is decomposed into several properties (change of state, incremental theme, versus other verbal entailments) and that argument selection depends on the weighting of these entailments, though can often be "floating" where arguments do not differ in these particular entailments. The advantages of eschewing atomic roles in favor of a count of individual entailments for argument may be even greater if as suggested below the relative IMPORTANCE of each entailment in a verb's meaning is a factor in selection.

The *hit*-class of verbs are also relevant to the question, mentioned earlier (§7), of the status of motion entailments in argument selection. If motion, as a change-of-state entailment, counted as a proto-patient property, it would seem that *hit the fence with the stick* entails movement for its prepositional argument but not its direct object argument. Since there are no entailments of change of state to distinguish or equalize the non-subject entailments (as we saw with *load* or *break*), *hit* would, if motion counted as a proto-entailment, violate the argument selection corollary that requires the argument ranking higher in patient properties to always be the direct object.

Possibly this shows, as suggested earlier, that motion should be treated as irrelevant for object selection altogether; I actually have no reason to reject this hypothesis. But a more interesting possibility arises when we compare a list of verbs that do alternate as *hit* does, cf. (62'), with similar verbs that fail to alternate and instead allow only the "Instrument" (65), or "Location" (66), as direct object:

- (62') a. John hit the fence with the stick  
 b. John hit the stick against the fence.

likewise: *strike, slap, swat, bash, whack, bang, pound, tap, bump, ?push* (different meaning?), *tamp, beat, hammer, flail* (with inanimate locative argument), *batter*

- (65) a. swat the boy with a stick  
 b. \*swat the stick at/against the boy

likewise: *smack, wallop, swat, clobber, smite, fell, bust, swipe, thump, pellet, stone, bunt, bat, poke, jab, flail, thresh, buffet, batter, pummel, pelt, drum, club, cudgel, bludgeon, truncheon, lambaste, whisk, strap, belt, baste, flog, spank, paddle, paddywhack, flog, cane, thrash, flail* (with animate locative argument)

- (66) a. \*dash the wall with the water  
b. dash the water against the wall

likewise: *throw, slam, bat, lob, loft, bounce, tip, crash* (note this does not behave like *break!*), *heave, hurl, fling, thrust, impel, sling*

(The above classes represent my judgments and will probably differ with the reader's on a few items.) Several verbs in (65) are derived from nouns referring to instruments, e.g. *club, belt, bat*, hence cannot always take a prepositionally-marked instrument phrase at all without creating redundancy. Of the rest, many verbs are typically or necessarily restricted to human or other animate beings as their "Locative" argument and imply a pain-inflicting or punishing action. While such actions do not inflict a readily observable change of state like the *break* class does, they do of course typically effect at least a certain mental state in the victim, and producing this effect is typically the motivation for the agent's performing the action; it is of more concern than the movement in the Instrument argument per se. Thus I am suggesting the verbs in (65) are actually like Fillmore's *break* in entailing a significant if less visible change of state in their direct object argument, although they cannot alternate in pattern to indicate that the change is in the "Instrument" NP instead as *break* does.

The meanings of the non-alternating verbs in (66) on the other hand most often differ from each other in the manner in which an object is caused to move through space, and I suggest they are typically used in contexts where it is the change of position in the thing moved (a ball or projectile) that is important, not any effect of the action upon the location where the object ends up.

The alternating *hit*-type verbs in (62') are in a sense intermediate between the other two classes: They more characteristically take inanimate "Location" arguments rather than animate ones, and although they are sometimes used when the agent's motivation is to achieve an effect on the "Location" rather than an effect on the Instrument, it can also be the reverse (*beat the rug against the wall*), or it can be the effect of the event in general rather than on either of the objects that is of interest (e.g. the noise it produces---these are also the verbs for describing the production of various sounds by percussive means).

In other words, I believe that (62)-(65) in general suggest that the characteristic *significance* of a change of state entailments in the context of the verb's overall meaning in part determines how it is counted (or weighted): only the more important change entailments count toward the Proto-Patient entailments of the argument in question, as they are added to other patient entailments to determine the allowable syntactic configuration(s). It tends to be verbs for which these change entailments are equally significant (or equally insignificant)

for both arguments that alternate like *hit* does.<sup>7</sup>

#### 9.3.4 Representation-Source Themes and Transformation verbs

I have discussed cases like (67) earlier,

- (67) take a nice picture of a scene  
make a superior recording of a live performance

which have effected objects (and Incremental Themes) as direct objects and contrast with cases like (68), where the object is what I called a representation-source theme:

- (68) photograph a scene  
record a conversation

I propose that it is also a consequence of the hypotheses under discussion that there are no examples like (69),

- (69) \*photograph a scene into a nice picture.  
\*record a live performance into a superior recording

where the representation-source is direct object and the Incremental Theme is a prepositional object: the obliques in (69) would outrank the direct objects in P-Patient entailments (change of state, incremental theme).

This claim immediately requires further justification because of examples like (70):

- (70) turn a live performance into a superior recording  
copy a file onto a disk  
commit the book to memory

Are these counterexamples? Not if these actually assert that the information in the file, viewed abstractly, "moves" from one place to another; exactly the same

<sup>7</sup> Also relevant to the argument selection problem for direct objects is of course the dative alternation (*give the book to John*, *give John the book*), but as this is a complicated case and as so much has been written about it recently, I will defer discussion of it for another context. With imagination, the interested reader can probably construct a position which would be compatible with the present approach from my comments about about the

information ends up in the resulting "copy" as in its source.<sup>8</sup> And in fact we do NOT seem to get this kind of sentence when it's clear that there is distinctly more information in the original than the copy:

- (71) a. \*copy the landscape into a painting;  
b. make a sketch of the landscape  
(#)make the landscape into a sketch  
c. make a summary of the lectures  
(#)make the lecture into a summary  
d. make a rubbing of the tombstone  
(#)make the tombstone into a rubbing

The parenthesized #-signs acknowledge the fact that the sentences are acceptable if the original object is not copied but actually converted into the second object. It is relevant to note for comparison that examples involving true physical transformation do have alternate argument configurations:

- (72) a. make a bird feeder out of the coffee pot  
b. convert the coffee pot into a bird feeder

This is predictable, since "both" arguments (rather, the "same" object under two descriptions) qualify as Incremental Themes---a change in one in fact literally is a change in the "other". Thus my interpretation of the examples in (70) is that for purposes of satisfying the Incremental Theme requirement, they are abstract "conversion of information" sentences analogous to (72b) rather than to (69).

<sup>8</sup> The first example in (70) is of course not literally that, but as a bit of hyperbole, or metaphor if one prefers, suggests that much has been successfully transferred; alternatively, it may be only that all the information wanted is transferred.

## 10. Related Proposals

This is a good point at which to pause to acknowledge other proposals in the literature which are similar to the present one and at the same time point out differences.

The proto-role hypothesis agrees with Jackendoff (1976, 1987) and Foley and van Valin (1984) in claiming thematic roles are "not primitives", but by that phrase they mean that roles should be defined in terms of configurations of semantic representations (which are made up of other primitives), not that role-types are prototypical, "fuzzy" notions. The "parts" of their definitions are not entailments like volition, sentience, etc. but DO, CAUSE, STAY, etc. I prefer to remain open here on the question whether all lexical meanings can be broken down into semantic structures of these kinds, as the present proposal is to an extent compatible with each answer to that question (though for arguments that aspectual distinctions cannot all be represented in a decompositional system, see Dowty (1979: ch. 3)). Likewise, even if lexical meanings are finitely decomposable in this way, it is a separate question whether each of the proto-entailment properties in (29) and (30) would semantically correspond exactly to a configuration of structures involving CAUSE, STAY, etc.

Rather, this hypothesis is most like that of Rozwadowska (1988) and Zaenen (1988), who have argued for a description of role-types in terms of "semantic features". As mentioned, I avoid this phraseology because I believe some entailments may have unclear boundaries and others may need to be "weighted"; dividing up the "features" into two opposing proto-categories is also not found in these works.

Foley and van Valin (1984:28ff) however do group roles into two "macro-roles", *Actor* and *Undergoer*, which are like my P-Agent and P-Patient. But theirs are two discrete categories, to which each role-type (and therefore each argument) does or does not belong: I see virtue in saying there are DEGREES of membership in the two P-categories, and I omit the traditional role-types as intermediate categories of special status. Jackendoff adds Actor and Instigator to the traditional Agent as role types, and he assigns arguments to multiple role-types (e.g. simultaneously Agent and Experiencer), and in these ways parallels my recognition of various combinations of entailments defining kinds of subcategory (e.g. volition + causation + sentience); But I have different "subcategories" from his, at least in the patient domain, and as mentioned, arrive at, conceptualize, and use roles differently. Culicover and Wilkins (1986) partition role-types into two groups, the *action tier* (Agent, Patient) and a *motion tier* (roles involving movement--Theme, Source Goal); to these Jackendoff (1983) adds a *temporal tier* (aspect and other time adverbials), but these divisions are orthogonal to any of the classes of arguments shown relevant to argument selection here. As already mentioned, the lists of entailments I have used to characterize the two proto-roles turn out to be quite

similar to those proposed by Keenan to universally characterize "Subject" (Keenan 1976) and "Absolutive" (Keenan 1984) respectively, but I have also explained why it is crucially a part of my proposal that Proto-Roles cannot be collapsed with grammatical relations in that way.

I agree with both Lakoff (1977) and the work of Hopper and Thompson (1980) in emphasizing the *continuous* nature of the distinctions one wants to draw in this domain. Lakoff views agency as a prototype and a psychological "Gestalt" characterized by a great variety of properties, though he also adopts a prototype view of individual lexical meanings (which I do not), and seems to view Patienthood as simply the complement or absence of Agentivity (which I cannot do). As already noted, the latter authors view *transitivity* as a property a **clause** can possess to a greater or lesser degree, whereas I think the transitivity of a clause can be derived by summing the independently needed agentivity and patientivity counts of the arguments.

Finally, there are similarities to Fillmore's later writings on case grammar (Fillmore 1977), which gives a different and more complex account of argument selection than his earlier theory did. This account involves both a "hierarchy of deep cases", as in his earlier work, to determine subject versus object, and also a "saliency hierarchy" to separate primaries--- subjects and objects--from obliques, the latter hierarchy being partly similar to the proto-role account in that it gives preference to humanness, change of state or location, and "definiteness and totality", but differs in being viewed as "saliency", a matter of importance or perspective. While I have also suggested some cases where object vs. oblique argument selection reflects the relative "importance" of entailments, I doubt that this argument selection case can really be REDUCED to saliency (cf. §9.3), and I think subject-versus-object selection need not appeal to a fundamentally different kind of selection principle from that for obliques, but my proposal is like this one in recognizing that a collection of distinct semantic contrasts is involved in argument selection. This is not the place to attempt a complete and necessarily very complicated comparison of the two approaches.

Thus most of the ingredients of the present proposal can be found independently in one place or another (which I take to be an encouraging sign), though never put together in such a way as to give a proto-role theory of argument selection like that offered here.

## **11. Some Questions for Psycholinguistics**

To the extent that the proposals made up to now have linguistic justification, they naturally suggest certain questions for psycholinguistics and the psychology of language. These will not be examined in depth here but only briefly noted for the sake of possible future study. Likewise, it should be well observed that these now transfer the level of discussion from the argument-selection problem alone to other domains for which the notion of thematic role has been invoked, so the cautions suggested in §1. should be kept in mind: though the possibility that a common notion of role-type applies across many domains is of interest, we should not rush to conclude this too quickly.

### **11.1 Argument Selection Principles**

First, one might ask whether there is any psychological and/or practical reason why languages should have argument selection principles at all. One answer that has suggested itself to several people is that such lexical patterns must surely make the task of acquisition of a (first-language) grammar more straightforward: see Grimshaw (1981), Marantz (1982), Macnamara (1982), and Pinker, who has termed it (one of the forms of) "bootstrapping". This idea may be described as follows.

Consider the dilemma of the child acquiring her native language, at the stage at which she first begins to figure out how grammatical relations are marked in her language. In a sentence with a verb and two nouns, how will she determine how the syntax indicates which is the grammatical subject and which is the grammatical object? It could turn out that word order marks this, or else that not word order but case affixes in NPs indicate grammatical relations (and children do recognize case immediately as signaling subject and object in such cases: cf. Slobin (1982)), or possibly that agreement affixes on the verb are the only signals of grammatical relations (in which situation the categories of nouns relevant for agreement, as well as the verbal morphology, must be decoded). The child will have to determine her language's system by implicitly comparing a number of different sentences in order to discern patterns. Obviously, this task is more straightforward if there are independent clues to guessing, when presented with a sentence and the situation which the sentence is used to describe, which of the two nouns IS the grammatical subject and which is the object. If consistent argument selection principles exist that must hold for some important class of verbs, i.e. principles relating grammar to meaning, these are the clue that the child can exploit in learning the morphological and syntactic coding of grammatical relations. Then the child can go on to use grammatical cues, in turn, to correctly learn the lexicalization of other classes of verbs for which semantic cues are not reliable (hence, "bootstrapping"). This hypothesis assumes, of course, that the child can independently infer at least parts of the intended meaning of an utterance from

the context in which it is used, at least some of the time.

The present account of selection principles makes slightly different predictions from other versions of the "bootstrapping hypothesis." First, it naturally explains why thematic-role-related entailments (causal and Agent-like entailments vs. Theme/Patient entailments) are the relevant semantic categories for children to pay attention to for the initial step in order to learn the grammatical codings (as opposed to, say, Experiencer and Location), rather than leave us to merely stipulate them, as e.g. Pinker (1984:40) does. Likewise, by giving the clearest argument-selection status to "highly transitive" verbs (high number of P-Agent and P-patient entailments that are harmonic with the principles), it offers a natural account of why children might fix on THOSE verbs as clues to grammar but so not try to infer marking of grammatical relations from statives, psych predicates, or verbs like *receive* or *undergo* and thus become confused by the sometimes idiosyncratic lexicalization of the latter---even though some of these verbs DO have, N.B., some degree of agency or causation or change of state involved in their meanings. Finally, given the conclusion of § 8.5 about syntactically ergative languages, this version differs in entailing that the direction of correlation (P-Agent with subject, P-Patient is object) is not really universal but that the converse association is also permitted; thus the child is predicted to have the further task of learning (from independent grammatical facts, such as coordination) WHICH of the sets of coding features she is identifying actually mark subject and which mark object.<sup>1</sup>

### **11.2 Why the Categories Proto-Agent and Proto-Patient?**

At the most general level, one might ask whether there is any reason why the particular selection of entailments involved in the proto-roles (intention, causation, change-of-state, etc.) should appear rather than other entailments (e.g. is a round object, etc.). But I assume there are fairly obvious answers to this question both in the world as well as in what is known about human cognition: distinguishing these properties is on the one hand an ability with obvious advantages to human survival and on the other, a well-studied cognitive ability that emerges at an early age (cf. e.g. Leslie and Keeble (1987) on the ability of infants to perceive causation as early as 27 weeks). I do not see anything that separates the present proposal from many

<sup>1</sup>It is also interesting to compare this hypothesis with a procedure actually recommended in a recent textbook for linguists in the field as they begin to analyze the grammatical system of an unknown language (Andrews 1985, 68-69). Andrews recommends that the linguist first try to elicit from the native informant a representative set of what he calls "primary transitive sentences", sentences with transitive verbs that have clear instances of Agents and Patients. From this list, the linguist should be able to discover how the grammar distinguishes subjects from objects generally (and whether the language is accusative or ergative). Only then is the linguist advised to go on to verbs such as psychological predicates, which can be examined, using the grammatical criteria already established, to see which arguments the language treats as subjects and objects in these less predictable cases.

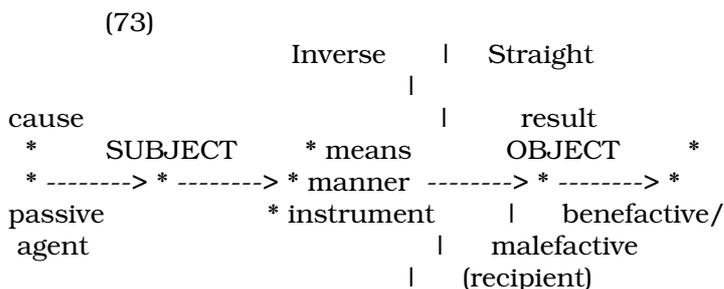
others at this level of generality.

What is more distinctive about the present view is the grouping of these entailments not into disjoint role types (Agent, Experiencer, Theme) but into two and only two super-groups of entailments. Natural questions to ask here are: (1) whether the child does not distinguish at all at the time of early language acquisition between, say, intention and movement and causation, or between causal affectedness and uncaused change of state, or (ii) though the child may be cognitively capable of distinguishing between these various individual categories (causation, volition, etc.), whether on at least on some occasions they are grouped together as a significant cognitive supercategory for the child (and similarly for Proto-Patient), and when faced with the difficult task of learning the first language, it is easier to first single out the supercategories linguistically than the finer ones. It does seem that in the environment of the very young human such categories coincide empirically in the majority of cases: most of the events that are described linguistically to a young child probably have a human "agent" that is a causal force AND a sentient and volitional participant AND an entity that moves (and a preexisting entity) simultaneously, and similarly for Proto-Patient categories.

Though I believe such cognitive hypotheses are ones that only cognitive science and psychology, not linguistics by itself, can adequately evaluate, I mention here two independent motivations I know for semantic "supercategories" very similar to those proposed here. One is a language acquisition study by Clark and Carpenter (1989) bringing evidence that "Children have a category of source that encompasses not only locations but also agents, causes, possessors, standards of comparison, and prior events" (p. 2). That is, their generalized "Source" category is evidently not exactly the "Proto-Agent" category I have discussed here but a supercategory of it. But note that the linguistic cases they treat which make it such are precisely those I have NOT discussed here, namely, cases where this generalized Source appears in some other grammatical form besides normal subject--- usually, as object of preposition *from*, as in the child's utterance *I took my temperature from the doctor*, meaning roughly that I had my temperature taken by the doctor). Their observed generalizations show that there is a common cognitive category here, even when expressed in DIFFERENT grammatical forms (subject and obliques). Clark and Carpenter introduce the term *emergent category* for categories similar to 'covert categories' but that "reflect the conceptual similarities perceived by children among paradigms of structures, even where these similarities are obscured by the conventional categories," (generalized) Source being one such category. Though the (traditional) Source does NOT move, as many Proto-Agents do, note that there is a conceptual connection between Agent and Source: in some prototypical causation events such as throwing something or handing an object to someone, the Agent, though it causes the event and makes a small local movement, stays behind while the object, the "Theme", moves away from it; in Clark and Carpenter's terms (p. 21)

the Agent is seen as the "starting point" for the action. In a non-causal event likewise, the Theme also moves away from the Source (its original location) while this Source remains stationary. (The question of how many and which cognitive super-categories of Proto-Agent and Proto-Patient might exist of course goes well beyond the scope of this paper.)

A different kind of evidence for proto-roles, which is like the foregoing in involving obliques rather than subject and object selection, comes from Croft's (1986b) cross-linguistic study of which syncretisms in cases (i.e. morphologically represented cases or adpositions, not "abstract case") representing the same thematic role are found natural languages and which are not. According to Croft, the best theory of this syncretism distribution is based on an analysis of events into "causal chains" whose organization is indicated by the chart in (73) (Croft 1986b:177):



Here, points marked with asterisks represent event participants having certain thematic roles, "SUBJECT" and "OBJECT" label the participants named by these two grammatically-identified positions (presumably, Croft assumes these are more or less typical Agent and Patient respectively), and the rightward-pointing arrows indicate the chain of causal relationships that Croft believes to hold among participants in a complex event. (Whether it is really correct to call all of these CAUSAL relationships and if so whether the ordering of them should be exactly as Croft diagrams it is actually not crucial to our concern here: the relevance to the proto-roles hypothesis is that roles to the left of the vertical line have proto-agent entailments, those to the right have proto-patient entailments.) Croft's observation is that case "syncretisms" within a language (the same morphological case or adposition used to indicate two different thematic role-types, e.g. as English *by* indicates passive agent as well as manner and instrument) clearly tend to occur within the set of "straight" oblique roles (those causally "downstream" from the direct object argument) or within the set of "inverse" oblique roles (those causally "upstream" from the object argument), but not ACROSS these two sets: in a typologically-balanced sample of 40 languages, he finds 39 instances of syncretisms among straight roles, 30 syncretisms among inverses, but only 5 cases of "non-directionality" (languages with such an impoverished case system to make the straight/inverse distinction meaningless) and only 2 outright exceptions.

By appealing to a familiar metaphorical association between movement and causation, what Croft (1986b:188) calls the OBJECT-LOCATION metaphor (in which the moving object continues to serve as the grammatical object as above, the point of origin is viewed as having a causal relation to it---cf. remarks on Clark and Carpenter' SOURCE above---and the point of destination as having a caused relationship), the allative (or Goal) oblique role may added to the set of straight roles and the ablative (or Source) role is added to the inverses. Examining syncretisms of one of these locative/directional roles with the causally-identified oblique roles in the above table, Croft finds an additional 13 syncretisms among inverse roles, 15 syncretisms among straight roles, but only 3 or 4 syncretisms across this division.<sup>2</sup> Croft's hypothesis about the role of causal change has many interesting implications which go far beyond the scope of this paper (it could be viewed as offering at least a partial explanation of what proto-role properties have in common), but its primary relevance is simply that it provides data from a domain quite different from acquisition for a non-discrete classification of role-types centering around two super-categories, but data that involves oblique arguments, and encompasses the role-types Source and Goal, as well as Subject and Object selection.

One interesting side observation to be made here is that the traditional roles Source and Goal make their appearance in Clark and Carpenter's and in Croft's studies, but neither these two roles nor sets of proto-entailments corresponding to them were found to be needed to describe subject and object selection. This might (or of course might not) turn out to be an illustration of my earlier suggestion that different notions of "thematic role" may emerge depending on which domain of questions ones tries to answer. (There is perhaps a natural taxonomic prejudice in the linguist to suspect that at the finest-grained level of analysis, underneath all this semantic variation, there is some universal Lockean semantic vocabulary into which all attested "thematic-role-related" generalizations can all eventually be decomposed. But is it time to give this idea up and look for universal principles instead in, say, the general PROCESS by which "big" semantic categories are linguistically (and cognitively) divided up into progressively finer ones as grammar and cognition develop, rather than for limits on the finest cognitive-linguistic categories that can supposedly result? What if, after all, there are no such ultimate limits?)

### **11.3 Proto-Roles as Defaults in the Acquisition of Lexical Meaning**

<sup>2</sup> Croft takes the position that this OBJECT-LOCATION metaphor is only one of several relevant but linguistically distinct associations that can be found between causation and change of position (and therefore that causal and spatial relations cannot be equated, as some theories have attempted to do). See Croft (1988:120-264) for discussion of these and several other varieties of typological syncretism generalizations.

By far the most interesting psycholinguistic suggestion posed by the proto-role hypothesis is, I believe, the possibility (already alluded to at several points) that proto-roles could serve the language learner as defaults for details of meanings of individual verbs where the learning context does not actually provide enough information to determine these details. This was suggested for the subject-volitionality entailments characteristic of intransitive vs. transitive collectives like *kiss* (cf. (42)-(44)), the subject-motion entailment for two forms of *collide* (cf. (45), (46)), the object-change-of-state entailment in Experiencer-object *please* vs. stative *like* (cf. (47)-(48)), the difference in incremental theme entailments of *load the truck with the hay* vs. *load the hay onto the truck* and other ditransitive patterns (cf. § 9.3.1).

On the one hand, these patterns looked too widespread and systematic to be the result of chance. But on the other hand, there were individual lexical exceptions to most every pattern (lack of either movement or volition in intransitive *be similar*, symmetric volition in *debate*, lack of incremental theme entailment alternation for the syntactic alternation with *hit*, contrasting with that in *spray* and *load*, the different pattern in change entailment for the alternation for *break*, lack of syntactic object alternation altogether in *fill*, *cover*, etc.), so the patterns cannot be attributable to compositional semantic rules associated with the constructions or to general constraints across all lexical meanings of a certain type. Most of these "exceptions" of course turned out to look quite sensible when one considered the nature of the types of events in the real-world that are important to humans (the necessarily symmetric versus possibly asymmetric volitionality in debating versus kissing, for instance.) Pending some other explanation of these apparently significant but partial semantic regularities, the fact that the semantic distinctions involved here all turn out to be among those we have postulated as defining ones for proto-roles motivates the hypothesis that learners may pick up such details of verb meaning by "semantic default", i.e. by taking it for granted that the subject and object arguments have the full complement of possible proto-role entailments appropriate to each of these grammatical relations, whenever the learning environment in which this word is encountered does not contradict this explicitly.

## 12. Proto-Roles and Intransitives: The Unaccusative Hypothesis<sup>1</sup>

An enormously influential idea which has now been adopted into several syntactic theories is the *unaccusative hypothesis* of Perlmutter (1978) (actually earlier put forth in Barbara Hall Partee's dissertation, Hall 1965): the thesis that some surface intransitive clauses, the so-called *unaccusative* ones, derive from underlying clauses with grammatical objects but no subjects, while others, the *unergatives*, derive from underlying clauses with grammatical subjects but no objects.

A fact which was I think was probably quite significant to the prima facie plausibility of this hypothesis, though not always explicitly emphasized as one of its features, is that the intransitive predicates argued to be unaccusative on syntactic grounds usually turned out to entail relatively patient-like meanings for their arguments (e.g. *arrive, die, fall*), while those argued to be syntactically unergative were usually agentive in meaning (*smile, walk, talk, etc.*). That is, given the assumption that direct objects are somehow inherently patient-like (and/or vice versa), it is plausible that the former but not the latter are "underlying" objects promoted to subjects.

However, as Rosen (1981, 1984) first clearly pointed out, it is apparently not possible to predict easily (if at all) from a given intransitive verb's meaning whether it will turn out to be unaccusative or unergative, in a given language, for no single semantic criterion (volition, agentivity, presentational meaning) or combination of them seems to determine this correctly for all verbs. Much less can the classes be defined semantically across all languages--certain pairs of corresponding verbs with meanings like *bleed, suffer, be afraid, talk in a delirium* can be observed to behave as syntactic unaccusatives in one language, unergatives in another. Some examples cited by Rosen from various sources are (Rosen 1984:61-67):

<sup>1</sup> This section owes much to the verbal conference presentation Bach (1988), which in turn commented on an earlier version of the present paper, in that Bach proposed that the proto-role hypothesis was relevant to unaccusativity phenomena, a topic I had not by that time addressed. My conclusions here are not necessarily the same as Bach's, however. Bach suggested a parallel (undoubtedly worth further pursuit) between unaccusativity and Whorf's (1945) notion of a *covert category* but proposing that grammaticized unaccusativity contrasts might be a case of a covert category becoming overt, not overt becoming covert. Cf. also Clark and Carpenter's (1989) views on *emergent* categories.

(74)	Unergative	Unaccusative
die	Choctaw	Italian
sweat	Italian	Choctaw
bleed	Italian	Turkish, Eastern Pomo
suffer	Italian	Choctaw
be hungry	Lakhota	Choctaw
sneeze	Italian, Dutch, Choctaw	Eastern Pomo, Choctaw

Another problem that has been observed with the unaccusative hypothesis is that sometimes different syntactic phenomena or "tests" which supposedly distinguish unaccusative from unergative verbs within a single language actually draw the boundary in different places. For Dutch, Perlmutter (1978) had originally claimed that (i) unergative but not unaccusative intransitives can appear in impersonal passives, and to this Zaenen (1988) adds, (ii) unergatives select the perfect auxiliary *zijn* ("be") while unaccusatives select *hebben* ("have"), and (iii) only unaccusatives allow for a past participle in prenominal position. However, Levin and Rappaport (1986) and Zaenen (1988) point out that the correlation does not hold up well; many verbs have no impersonal passive but take *hebben*.

I propose that the thematic proto-role hypothesis offers both an explanation of why a semantic distinction among intransitives of the unergative-vs.-unaccusative sort should have a grammatical correlation in the first place, but more importantly, that it goes a long way toward explaining both problems of "variation" in membership in the two classes.

As may have already occurred to the reader at this point, Proto-Agent and Proto-Patient are arguably the two (fuzzy) categories of arguments that semantically characterize unergatives versus unaccusatives, to the extent that the distinction has any clear semantic characterization. If, as I have proposed in this paper, these proto-categories describe argument selection, possibly play a role in language acquisition (in learning the coding of grammatical relations and as lexical semantic defaults), may correspond to something like Clark and Carpenter's "emergent categories", and have typological reflexes, then they must be important factors in the semantics-syntax interface and are probably cognitively salient at the time of language acquisition. Hence the fact that languages can make a bifurcation along this line among intransitives (according to their SINGLE argument), parallel to the use of the contrast to distinguish AMONG the arguments of individual transitives and ditransitives, should not be surprising but is rather almost to be predicted. (It for reasons of pointing this out, of course, that I delayed the discussion of unaccusatives until after the preceding sections.)

But just as we saw that the Proto-Agent/Proto-Patient distinction was not a

discrete one in argument selection but one of degree, we can see that the "cut" between unergative and unaccusative arguments is indeterminate, varying, I argue, according to the same parameters. From the list of Proto-Agent properties, the most important for the unergative/unaccusative contrast seems to be volition (or "protagonist control" in the unaccusativity literature, i.e. a presupposition that volition is possible for this type of action); volition necessarily involves sentence, and verbs with both these entailment are ALWAYS unergative, it seems. The slightly broader presupposition of "being predicatable of a human being" sometimes but not always put a verb in this class, as does movement. As Rosen (1984:65-66) points out, predicates which are restricted to humans and involve some movement but in which volitionality can either be present or absent (or for which it is "marginal") seem to vary from one language to another as to which class they belong to: this class includes *sneeze, bleed, vomit, snore, blush*. That is, whether sneezing is a volitional activity is less clear than whether singing or dancing is (or, on the other hand, whether being six feet tall is): while one does not normally deliberately decide to sneeze, one can if one wishes deliberately perform an action that is outwardly indistinguishable from ordinary sneezing, and one can sometimes avoid sneezing by trying not to do it (whereas in most cases acts or states that are not volitional are also not really "avoidable").

Among the Proto-Patient entailments, incremental themehood (or slightly more generally, whether the argument is an incremental OR holistic theme, i.e. is telic) seems to be highly significant for the distinction between unaccusatives and unergatives, just as it was often found to be the most significant proto-patient entailment for object selection of transitives (cf. §9). But the appeal to two entailments each from a different proto-role gives us two possible loci for a semantic boundary (in addition to any vagueness in the criteria individually, e.g. that in "Agency" already alluded to):

(75)	Atelic	Telic
Agentive	1. definitely unergative	2.  ?
Non-Agentive	3.  ?	4. definitely unaccusative

If the most important distinction is between agentivity and lack of it, then verbs in square 1. and 2. are unergative (2. would include e.g. *stand up, retire*), and those in 4. and (most) in 3. and 4. are unaccusative (3. includes e.g. statives like *exist, be in the room*). But if the distinction between telicity and lack of it is primary, then

verbs in 2. and 4. are unaccusative, while most in 1. and 3. are unergative. (Perhaps "active" languages like Lakhota, where the notable grammatical realization of the contrast is in case or agreement marking for the subject NP, are closer to exemplifying the former, while unaccusativity as manifested in Italian is more like the latter.) A prediction made by associating the Proto-Role hypothesis with unaccusativity is that in any language with manifests unaccusativity, predicates that are "high" in agentivity AND "low" in patient properties are invariably unergative, while those low in agent properties and high in patient properties are invariably unaccusative; only those high in both kinds of entailments, or low in both, should be unstable. As far as I am aware, this is correct.

Before going any further, we must distinguish two different ways that an unergative/unaccusative distinction could enter into a grammar:

(i) the distinction is a grammatical one between two classes of intransitive verbs, having a correspondence with some partitioning of the continuum from Proto-Agent to Proto-Patient (though possibly only a rough correspondence). But each individual verb is assigned once and for all to one of the two grammatical classes (in each language).<sup>2</sup>

(ii) Certain grammatical constructions have certain meanings associated with them (entailments or conventional implicatures) involving Proto-Agent or Proto-Patient properties, hence a given intransitive verb is appropriate in such a construction only if it has the right kind of meaning. The set of grammatical rules/constructions appropriate to one semantic class, versus that appropriate to the other class, thus isolate two classes of verbs, but via semantic constraints originating in the rules themselves.

The difference between (i) and (ii) is the same as that between grammatical and semantic gender: in grammatical gender, each noun is permanently assigned to one gender, but gender does not (synchronically) play a role in semantics, though it may reveal its historical semantic roots in a partial correlation with semantics. In the case of semantic gender, gender does make a real contribution to meaning, and certain distributional facts, for example, that *he* cannot normally be co-indexed with *she* in *He thinks she is intelligent* can be given an explanation in terms of

<sup>2</sup> As Rosen (1984) mentions, there is of course the possibility of assigning certain verbs to both classes in a language, e.g. *fall* and *sneeze* in Choctaw (as mentioned in (74)). If however this should turn out to be necessary for more than a few verbs, and if the occurrence of the verb in the two syntactically diagnostic frames should turn out to correlate with a difference in the verb's interpretation of an agent-vs.-patient sort, then I would argue the correct analysis is of the second, "semantic", type below, instead of or in addition to the syntactic type. Hypothesizing that a large semantically coherent group of verbs have duplicate categorization in unaccusative and unergative syntactic classes (and with corresponding different semantics in the two frames) would be, I argue, missing the point.

meaning, not syntax. Likewise, if the difference between the syntactic distributions of ergative vs. accusative predicates is of the second sort, then it is not necessarily a syntactic distinction at all. The possibility of an analysis like (ii) of course presupposes that a grammatical construction (or some morpheme serving as head of the construction) can be analyzed as having a meaning (and/or conventional implicature) of its own, but it is a feature of compositional semantic theories since Montague (1974) that they permit constructional as well as lexical meaning.

One way of distinguishing the latter way of drawing the distinction from the former way is that the class of predicates permitted to appear in constructions specific to one class can be extended beyond the normal class in certain contexts, to convey some fanciful, metaphorical humorous, or otherwise non-literal effect. For example, English *is being* ADJ presupposes that the property ADJ is under volitional control of the subject (cf. *Mary is being quiet* vs. #*Mary is being pregnant*), and yet a speaker may utter *This Xerox copier is being stubborn again* without either violating a grammatical categorization of *stubborn* or being taken to seriously believe that the Xerox machine has malevolent intentions. The classes of adjectives permissible in *is being* ADJ is thus semantically delimited, not syntactically determined.

Zaenen (1988) argues that the Dutch impersonal passive construction is just such a case, conveying that the property denoted by the verb is an intentionally controllable one. "Abnormal" sentences like (76) (Zaenen 1988: 14) and (77) (Perlmutter 1978) can be and are uttered by native speakers of Dutch, though understood as conveying an atypical assumption as to what actions can be intentional and are therefore a kind of joke:

(76) Er werd door de krenge gestonken  
"There is stunk by the nasty women"

(77) In het tweed dedrijf werd er door de nieuwe acteur op het juiste ogenblik  
"In the second act there was fallen by the new actor on cue"

German impersonal passives likewise have a volitional implicature and permit extended use with non-agentive verbs for humorous effect (John Nerbonne, p.c.); in fact, Nerbonne (1982, 1984) analyzed German impersonal passives as having such an implicature. By contrast, a syntactic account of unaccusativity does not immediately predict that such "violations" should be any more permissible than any other kind of syntactic ill-formedness.

If an "unergative/unaccusative" contrast effect arises from an implicature of a syntactic rule rather than a syntactic/lexical categorization of intransitive

predicates, then insofar as different syntactic rules each have their own interpretation rule, "semantic" unaccusative/unergative contrasts isolated by different rules could in principle have slightly different implicatures (at least, under the Proto-Roles hypothesis, where there are predicted to be various semantic properties for demarcating the classes) and so in principle isolate different "unaccusative" verb classes in the same language. And Zaenen (1988) argues this possibility is the source of the above-mentioned bifurcation in Dutch: while she claims the impersonal passive construction has an (atelic) volitionality implicature (thus distinguishing verbs in cell 1. above from the other cells), telic but not atelic predicates are argued to select *hebben* rather than *zijn* (i.e. distinguishing verbs in cells 3. and 4. from the others). Of course, telicity in predicates is determined not only by the lexical class of the verb but also the aspectual adverbials that accompany it (Verkuyl 1972, Dowty 1979), and this is Zaenen's argument that a sensitivity to aspect in the semantics of *hebben* vs. *zijn*, rather than a fixed syntactic category of unaccusative predicates, is at issue in this second distinction in Dutch as well as the first. The broader point is of course that the semantic analysis explains how two "different" divisions between unaccusatives and unergatives are made in the same language, while the position that the distinction is a necessarily syntactic one cannot readily accommodate such a situation. <sup>3</sup>

Rosen (1984) clearly takes the position that unaccusativity as a syntactic phenomenon really exists, as have a number subsequent writers. While I think the question deserves to be examined more closely via a closer semantic scrutiny of the "tests" observed in each language, I will assume for purposes of this article that syntactic unaccusativity occurs, and address briefly the implications of the Proto-Role hypothesis for it. (The above discussion does not in any way rule out the possibility that both syntactic and semantic unaccusativity could be found in the same language, and this may well occur: for example, Holisky (1987) cites the case of Tsova-Tush, where case marking interacts with unaccusativity in a complicated way involving markedness, but a way that appears to include both syntactic categorization of verbs and semantic entailments of constructions, neither reducible to the other.)

In this paper, I have been at pains to argue that while the way the Proto-Agent/Proto-Patient opposition is CONNECTED to the grammatical opposition between subject and object, neither opposition is REDUCIBLE to the other, nor is the association of subject with (Proto-)Agent and object with (Proto-)Patient a necessary one. There were three reasons for this:

<sup>3</sup>(Though on a semantic analysis of the distinction it does not literally follow that entailments attached to particular syntactic rules or items taking verbs as complements would tend to cluster even loosely around the "same" two semantic groups of predicates at all, this clustering would seem likely to arise, given the assumption that, as an empirical fact about common intransitive verb meanings, most cluster around one of the end of the Agent-Patient continuum or the other.)

(1) The correlation of proto-roles with grammatical relations in English-like languages is only a TENDENCY, not an absolute, and admits of quasi-violations (under relatively predictable circumstances) like the lexicalization of "conflicting" pairs like *like* vs. *please* and "counterexamples" like *receive* and *undergo*.

(2) Some languages, namely syntactically ergative languages, have the INVERSE correlation between subject/object and Proto-Agent/Proto-Patient from that of English.

(3) there is already some evidence that Proto-Agent (and possibly Proto-Patient) exist as operative categories in language independently of subject and object, namely in Croft's observations about two classes of "oblique roles" (i.e., grammatical positions exclusive of subjects and objects) that are similar to Proto-Agent and Proto-Patient, and Clark and Carpenter's generalized Source category, a super-category of my Proto-Agent realized through grammatical obliques.

To the extent these arguments are correct, then I believe there is much less naturalness (much less necessity) than is often assumed in identifying a "surface" syntactic or lexical category of unaccusative intransitives with "underlying objects that have been promoted to subject". That is, if we know independently that "Proto-Patient" is an influential semantic category that can manifest itself in various ways in language besides merely direct object (and if direct objects are not necessarily Proto-Patients) then why not identify Proto-Patients directly with a syntactic subcategory of intransitive verbs, without the intermediate assumption that because they are Patient-like they must also be, in some sense, direct objects? After all, it is often desirable to distinguish some subcategory of verbs for syntactic or morpho-syntactic purposes, even where the subcategory has partial or complete semantic correlation, by means no more complicated than a syntactic feature: for example, stative predicates are morphologically or syntactically distinct from other verbs in many languages (cf. Comrie 1976:50 and Watters 1985:14, discussed earlier), though the semantic correlation of the syntactic stative class varies not only from one language to another but is apparently inconsistent within languages.

Of course, the hypothesis that unaccusative clauses are derived from sentences with grammatical objects but no subjects has been defended at length with ostensibly SYNTACTIC arguments ---arguments that the overall grammar of a language is improved by this kind of derivation, in spite of the price paid for the additional step of advancing the object to subject in most situations. My purpose here is not to take issue with these arguments (an undertaking far beyond the scope of this paper in any event), but to maintain that they always need to be evaluated in light

of three points: (i) the unaccusative advancement hypothesis must stand or fall on the SIMPLICITY of its syntactic analyses per se; it should not really gain any support (explicitly or implicitly) from the "naturalness" of associating patient-like intransitive subjects with grammatical objects, and (ii) in arguing for an unaccusative derivation, it does not suffice merely to accumulate a variety of ways in which unaccusative predicates behave alike and ways in which they are different from unergatives: the Proto-Role hypothesis, if it can indeed have effects in various aspects of grammar acquisition, would be abundant reason in itself why grammar-learners might tend to be attuned to intransitives with patient-like arguments as a class and hence regularize (and over time increase) coincidental differences between them and agent-like intransitives; distinguishing unaccusatives by a feature [unaccusative] might describe this situation perfectly well. Rather, the only successful arguments for unaccusative advancement will crucially involve grammatical parallels across unaccusative verbs and transitives (and/or unaccusative subjects and objects of transitives) that can be exploited to simplify the grammar without introducing additional complications through the unaccusative derivation. For example, a language in which in which a grammatical rule, say passive, applied to both transitives and unaccusative intransitives but not to unergatives would be such a case.<sup>4</sup> (By "additional complications" I mean for example, that we must prevent it from following from our unaccusative analysis and the fact that verbs can "share" an argument bearing the same grammatical relation, e.g. *Mary caught and John ate the big fish*, that we predict *\*Mary saw and arrived the tall stranger* is grammatical.) Above all, (iii) it is necessary to be sure one is dealing with syntactic unaccusativity rather than solely semantic unaccusativity, for semantic unaccusativity does not motivate an unaccusative advancement analysis.

How persuasive such syntactic arguments are is, unfortunately, a question that will almost surely depend on one's grammatical theory. If one favors a multi-stratal theory in which "advancement" derivations are already frequently used for other purposes (as in Relational Grammar, Arc Pair Grammar, or Government Binding Theory) and in which mechanisms are already in place to trigger and constrain such advancements, then the unaccusative advancement analysis is relatively "cheap", and even slight simplifications achieved by that analysis would easily justify it. If however one believes that monostratal syntactic theories with structured grammatical categories such as Generalized Phrase Structure Grammar

<sup>4</sup> In languages like Dutch and German, of course, the facts are the other way around: impersonal passives are found with unergatives, not unaccusatives. Relational Grammar of course has an interesting account of this situation, where Unaccusative Advancement in effect precludes the possibility of passive thereafter, though this involves theory-specific assumptions which one could imagine being otherwise (passive might have had a chance to apply BEFORE unaccusative advancement, etc.). My point is not to quibble with that analysis but simply to point out that the more directly a set of facts reveals the simplifying power of an unaccusative analysis, and the less theory-specific the assumptions that are required to cash in on the simplification, the more persuasively the case for the unaccusative hypothesis itself is made.

(Gazdar et al 1985), Head-Driven Phrase Structure Grammar (Pollard and Sag 1987), or one of various versions of Categorical Grammar (cf. e.g. Oehrle et al 1988), when combined with an explicit semantics, have provided a fundamentally adequate description of natural language syntax, then I suggest that what is known about the unaccusativity phenomenon fails to provide any good reason for rejecting monostratal frameworks. This will be so in particular if the two opposing proto-roles have the importance in various aspects of language that I have suggested here, for they offer a reasonable motivation for the semantic parallel between the existence of two subcategories of intransitives in natural languages and the criteria for selecting direct objects for multi-place verbs, but without invoking "grammatical object" in the description of intransitives at all.<sup>5</sup> To be sure, the extensively-argued advancement analyses such as those of Rosen (1984), Burzio (1986) and others for Italian demand to be answered in detail (either to argue that each unaccusativity phenomenon is semantic or to provide a plausible monostratal alternative for any grammatical unaccusativity) to follow through on this suggestion

<sup>5</sup> It has been suggested that certain Italian sentences with unaccusative verbs may be best analyzed as having SURFACE direct objects but no subjects (or having dummy subjects). Such an analysis is not however an ADVANCEMENT analysis, and it is not in conflict with a monostratal syntactic theory or with the proto-roles hypothesis, as far as I can see. One would treat such predicates, like existential dummy-subject verbs, as having meanings which are technically two-place functions but only trivially so, having denotations which give the same values for any subject argument (or alternatively, are well-defined only for the "dummy" argument), so they are equivalent to one-place predicates in semantic effect. If a language makes use of such a possibility to "expand" one-place predicates to two-place, then given the proto-role hypothesis it is not at all surprising to see unaccusatives but not unergative predicates given this treatment, since it permits the association between proto-patients and grammatical objects to be made more widespread, though at the price of a slightly more complex syntax than simple intransitives have --- but not as complex as a multi-stratal analysis. (If the same verb appears sometimes with a subject, sometimes with an object plus dummy, then a lexical rule would be required to relate one subcategorization frame to the other.) Compare this with the case of dummy subjects of existentials, where as suggested in § 5, languages "expand" a one-place to a vacuous two-place relation in order to remove a NP having a newly-introduced referent from the (weak) association of "topic" that the grammatical position of subject would otherwise give it. In both cases, the correlation of grammatical position with a semantic/discourse property that is achieved is a widespread but not a necessary one; indeed, the former conflicts with the latter (the non-subject arguments of existentials created this way, which in English examples like *It's me!* are quite object-like, are not patients).

