

Prototypical Predicate Constructions: Evidence from Slavic

Patrick M. Murphy

University of North Carolina at Chapel Hill

patrick_murphy@unc.edu

While attempting to capture systematic relations between form and meaning, transformational grammar has relied upon several tools for analysis. Though the generative enterprise has promised to capture linguistically significant generalizations and in doing so provide an explanatory account of linguistic phenomena, some of these generalizations are impossible to capture in terms of these tools. This is primarily due to a fundamental shortcoming of the mechanisms employed most often in transformational approaches; that is, that many linguistically significant generalizations cannot be expressed by means of a derivational relationship. This paper attempts to lay the groundwork for the analysis of prototypical transitive predicates, in which syntactic and functional equivalency of constructions is captured in terms of direct surface relationships as opposed to derived equivalencies. This is accomplished by means of a ‘fuzzy’ evaluation metric on the topicality of signs, based upon data from Polish word order in Siewierska’s (1993) study.

Tools of the Trade

Generative grammar has employed several mechanisms to capture the systematic relationship between linguistic constructions. Take, for example, passive sentences (e.g., *The house was built by the workers*) and their active

counterparts (e.g., *The workers built the house*). Chomsky utilizes transformations to capture the relationship between such sentences, e.g., both the active and passive sentence are derived from the same thematically relevant deep structure by transformations, although these sentences have differing surface structures. The functional similarity is that there is one verb in the lexicon whose argument structure is expressed at D-structure; a transformational rule does not apply in the active case but a transformational rule derives the passive S-structure¹.

Another means by which generative grammar has sought to relate such constructions is by means of the lexical rule, whereby lexical entries are derived from other lexical entries. Much as transformations relate one tree structure to another, lexical rules relate lexical heads to one another in an attempt to capture generalizations. Instead of relying upon a passive transformation to express the relationship between active verbs and their passive correlates, one may propose a passive lexical rule which accepts as input active verbs and which outputs passive verbs by altering the verb's argument structure.

In Government and Binding theory, transformations alone no longer account for passives; the passive verb is the same as an active verb but it does not have an external θ -role to assign; the internal argument can move to the subject position because it will not be assigned two θ -roles. In the active sentence, movement of the 'object' to the 'subject' position is ruled out because the object

¹ Though the transformations of 'classical' transformational grammar have been replaced with the more general notion of movement (i.e., 'Move-') and constraints on such movement within Government and Binding theory (GB), the relationship between entire sentences or tree structures is still codified in such a derivational relationship.

would be assigned two θ -roles². While no longer the product of a transformational rule, this relationship is still derivational in terms of a lexical rule deriving a passive argument structure from that of an active predicate.

The attractiveness of these approaches aside, they run into problems with descriptive adequacy let alone explanatory adequacy. In addition, many of the valuable insights of functional analyses are not easily codified in terms of these devices, and therefore often ignored by generative grammarians. Although they may have a different research agenda, in ignoring functional analyses, many generative linguists miss the linguistically significant generalizations one would certainly like to capture. One such example would be the topicality of arguments as outlined below.

To this end, we will use a ‘fuzzy’ evaluation metric to capture the functional similarity between these structural differences, and appeal to another mechanism upon which Head-Driven Phrase Structure Grammar (HPSG, Pollard and Sag 1994) has relied to capture generalizations: type theory and multiple inheritance. In HPSG, signs are represented as typed feature structures and types are arranged in a hierarchical fashion. Subtypes add their own independent information to that which they inherit from their supertype. In this

² Recall that in GB theory, NPs cannot receive more than one θ -role, and NPs inherit θ -roles from their traces. Thus in the passive cases the internal argument will receive a θ -role through its governed trace, but will not receive an external θ -role (the passive verb has none to assign), while in the active case, movement of the internal argument to the subject’s A-position would mean that the internal argument would receive both the internal θ -role (through its trace) and the external θ -role (through its position).

way, generalizations are expressed through the arrangement the types that are postulated for a given grammar and the grammatical information they contain.

There are any number of phenomena where it seems a 'fuzzy' or weighted metric is needed. While grammatical study often centers around rules which are absolute (grammatical versus ungrammatical), it is clear that many of these black-and-white rules must somehow be gray. Take, for instance, the large amount of situational, dialectal, and even idiolectal variation in usage; different speakers have different preferences which suggest 'fuzzy' evaluations. For evidence against the parametric analysis of such phenomena, let us consider Abney's (1996b) arguments.

Abney first questions how parameters or discrete grammatical rules can be applied to known facts of language acquisition. After all, advocates of the Principles and Parameters approach have long professed to have as their goal an explanation of how children acquire language, and this supposedly hinges upon the universal principles and parameters of Universal Grammar. Yet Abney notes that during language acquisition children do not abruptly change their grammars in the fashion that parameters would seem to suggest, but rather they gradually change the relative frequencies of structures over months or more, possibly weighing one structure versus another until assigning one structure a probability of zero. As Abney notes, "an aspect of this...that bears emphasizing is that the probabilities are added to a grammar of the usual sort" (1996b: 2).

Similarly, language change is certainly not as non-continuous as the parameters of GB would seem to predict; it is clear that the gradual and subtle changes we would like to account for are easier stated in preferential terms as opposed to one-or-the-other rule of grammar. It is more likely that grammatical changes occur as the preference for certain constructions within speech

communities shift. This leads us naturally to the issue of dialectal and idiolectal variation as well as register. In order to explain dialect continua and the *Sprachbund* Abney suggests “the continuousness of changes with geographic distance is consistent with the picture of a speech community with grammatical variance” (3).

Even linguistic universals themselves seem to beg for a preferential analysis. Most universals are given in terms of language preferences, and not as absolutes. Phenomena which are ruled out by widely accepted universal constraints are probably not truly universal. Double passivization and violation of Ross’ Complex NP Constraint may be so rare that we want to consider them absolute universals. Yet their apparent existence would seem to suggest we must instead make them possible constructions which, measured in terms of a probabilistic or weighted grammar, are simply so unlikely or marked that we predict they appear with almost no frequency.

These are perhaps the strongest arguments in favor of stochastic models of linguistic data; gradations in grammaticality judgements, performance issues, semantic drift, etc., may also argue for weighted models of many parts of grammar as opposed to absolute rule systems.

Topicality and Word Order Preferences

One area that strongly suggests evaluation in terms of such weighted grammars is that of word order variation. Clearly, choice of word order is influenced by several factors which may be modeled stochastically; most ‘naturally’ occurring order, pragmatic factors, etc. One choice that a large amount of evidence seems to suggest seriously affects word order choice is that of the topicality of subject and object.

In the chapter of his book on word order and discourse factors, Myhill (1992) looks at several languages with quite different basic word orders and attempts to make some generalizations. Most of his analysis involves study of Referential Distance and Topic Persistence counts of NPs in each language, with the NPs broken down into certain categories such as definite versus indefinite, pronoun versus full NP, etc.

These terms may be understood as follows:

Referential Distance (RD)

This measurement assesses the gap between the previous occurrence in the discourse of a referent/topic and its current occurrence in a clause, where it is marked by a particular grammatical coding device. The gap is thus expressed in terms of number of clauses to the left. The minimal value that can be assigned is thus 1 clause, which is maximally continuous.

Topic Persistence (TP)

We measure persistence in terms of the number of clauses to the right — i.e. in subsequent discourse from the measured clause — in which the topic/participant continues as an uninterrupted presence as a semantic argument of the clause, an argument of whatever role and marked by whatever grammatical means. The minimal value that can be assigned is thus zero, signifying an argument that decays immediately, i.e. of the lowest persistence.

(Givón 1983a: 15)

Restricting ourselves to examination of only one language for the sake of brevity, let us consider Myhill's analysis of Givón's (1983a) study of Ute (basic order for Ute is described as 59% SV and 78% OV). This study found that preverbal NPs have high Referential Distances, whereas low RDs were associated with postverbal NPs. Topic Persistence also correlates to word order, but is different for subjects and objects; subjects have high TP if preverbal and low TP if postverbal whereas objects were just the opposite. This led Myhill to propose the following generalizations about Ute NPs:

Function of preverbal and postverbal position in Ute

Preverbal subject — switch attention to a character not on the scene before who will become the center of attention

Postverbal subject — keep attention on a character who will not be continued for long as the center of attention

Preverbal object — refer to an entity not mentioned in the preceding or following discourse

Postverbal object — refer to an entity mentioned in the preceding and following discourse

(Myhill 1992: 171)

In looking at several languages with different basic word orders, Myhill suggests a universal pragmatic word order pattern: marked or contrastive topics (or 'shifted topics') tend to appear in initial position. Other tendencies seem to depend on the language's basic word order type (SVO, SOV, etc). For SVO languages it was found that preverbal subjects are likely to be pronominal or definite, while postverbal subjects were likely to be marked topics (those which have not been mentioned for a while) or presentatives (where a new topic is introduced and focused upon).

Myhill also looked at passives as a word order phenomenon, paying special attention to the topicality of the verbs' arguments. While he found that several different factors influenced choice of active verbs or passive verbs, of primary importance was the relative topicality of the agents and patients. He found that generally speaking, subjects of active verbs are higher in topicality than the direct objects of active verbs, with such subjects usually being human and objects being non-human. Passives reverse this pattern and are associated with logical subjects which are low in topicality and logical objects high in topicality. Estival and Myhill (1988) examined the role of topicality and choice of word order in English. As passives are used when the logical subject is less topical and the logical object is more topical (they cite Svartvik's (1966) text count

study which found inanimate logical subjects in 81% of English passives but only in 27% of actives), the following table shows some of the features affecting topicality³:

	Active	Passive	% Passive
Nominal A	350	18	5%
Pronominal A	1760	1	0%
Non-human A	100	8	7%
Human A	2010	11	1%
Indefinite A	180	8	4%
Definite A	1930	11	1%

(Estival and Myhill 1988: 457)

The logical subjects are often not even overtly expressed in passive constructions, which “can be seen as the most extreme type of detopicalization” (Myhill 1992: 102). This seems to be the most basic type of passive in the languages of the world that employ passive constructions; Keenan’s (1985) generalization is that a language that has passives with agent phrases will have passives without them (249).

Word Order Preferences in Polish

Word order has long been a topic of discussion among Slavicists. Perhaps the most famous example of this is the study of word order from the Functional

³ Here A is the subject of a transitive verb or the corresponding oblique of the passive, the “logical subject.”

Sentence Perspective of the Prague School, whereby old information (the *themē*) is placed before new information within the sentence (the *rhemē*) in non-emotive speech. Thus word order is pragmatically determined⁴.

Still, it has been argued that in fact the universal preference might be the other way around, with new information being presented at the end of the sentence. Whether this is true or not, here again we seem to have a case of a statistical universal, not an absolute one. Even within the theme/rheme analysis it is clear again that in 'emotive' cases word order can be changed and there are other stylistic/dialectal/idiolectal variations. It would seem that though there might be a universal preference for one order or the other, we need to account for variation.

Siewierska (1993) investigates Polish word order in terms of both 'syntactic weight' (length and complexity of syntactic constituents) and information structure (topicality in the terms discussed for Myhill and Givón above). Her findings regarding topicality in Polish (remembering that high Topic Persistence was taken to be indicative of topicality) are as follows:

Topic Persistence in Polish		
<u>Word Order</u>	<u>Subject TP</u>	<u>Object TP</u>
SOV	1.69	0.65
SVO	1.02	0.68
VSO	0.56	1.11
VOS	0.94	0.50
OVS	0.88	0.49
OSV	0.92	0.37

(Siewierska 1993: 257, 261)

⁴ For an analysis of the theme/rheme distinction within HPSG, where type-theoretic constraints are utilized to capture word order constraints based on file change semantics, see Murphy (1995a).

Note that in terms of topicality *a la* Myhill or Givón, Polish word order seems fairly flexible for “normal” topicality. That is to say, on average for the word order types, we see topic persistence for subjects hovering around 1.00, whereas object topic persistence averages 0.63; this follows Myhill’s generalization that subjects are more topical than objects, but that inverted word order is a strategy of languages used more commonly as topicality of the subject drops, the topicality of the object increases, or both.

Two word orders seem to encode predicates whose subject’s and object’s topicalities are marginal. In SOV word order, the topicality of the subject argument is considerably higher than the average topicality for subjects (1.69). In VSO word order, not only is the topicality of the subject considerably below its average topicality (0.56), but the topicality of the object is considerably above its own norm (1.11). Apparently, these two word orders are utilized by Polish speakers to encode unusual topicality of subject and/or object. While not necessarily inverted word order as per Myhill’s generalization, it would seem a different word order at least encodes the unusual topicality of arguments. It is interesting to note that by Siewierska’s count, of the six possible word orders, these two types represented but 16.2% combined of the transitive orders, which suggests that their relative rareness reflects their more marked nature.⁵

⁵ Throughout I use the term ‘marked’ in a not uncontroversial way: to refer to a sign’s relative complexity, predictability, or productivity. In essence, I use ‘marked’ to mean divergence from a prototype, with more typical/unmarked signs sharing more features with a prototype and more atypical/marked signs sharing fewer features. While I do not wish to ignore Jakobsonian notions of opposition and invariance as part of markedness, I will continue to use the term in this manner throughout this paper.

Modeling Prototypical Predicates

In trying to provide a descriptively adequate model of the Polish data, we will model generalizations about what predicates are and are not prototypical in terms of constraints. This can be envisioned as a metagrammar which, while not ruling out ill-formed signs as does the grammar, evaluates signs in terms of appropriateness conditions. I will formulate these conditions in terms of Prolog-style clauses. First, we need a mechanism by which to measure the topicality of a given nominal object. With a measure of its topicality, we can now define what it means to be of 'neutral' topicality for an agent or patient.

neutral-topicality-agent (\square index) :-
 topicality (\square , 0.65, 0.85).

increased-topicality-agent (\square index) :-
 topicality (\square , 0.85, 0.90).

neutral-topicality-patient (\square index) :-
 topicality (\square , 0.15, 0.35).

...

That is to say, an agent is considered by the metagrammar to be of neutral topicality if its topicality falls within normal bounds, being a function of those features of the sign which seem to be indicative of topicality (person, number, humanness, animacy, definiteness), and represented here as being between 0.65 and 0.85, on a scale from 0 to 1 where 0 represents total lack of topicality and 1 represents a most topical sign. Similarly, neutral topicality for a patient can be defined, its bounds here being lower to conform to Myhill's generalization. Further statements can be made for increased topicality of agents and patients, decreased topicality, etc.

What is then needed is a way to state what a prototypical transitive verb looks like. Again, we can represent an appropriateness condition in the metagrammar as a clause like the following:

central-trans ($\boxed{1}$, word) :-

$$\boxed{1} \left[\text{SYNSEM|LOCAL} \left[\text{CONT} \left[\begin{array}{l} \text{AG } \boxed{2} \\ \text{PT } \boxed{3} \end{array} \right] \right] \right],$$

neutral-top-agent ($\boxed{2}$),

neutral-top-patient ($\boxed{3}$).

marginal-trans ($\boxed{1}$, word) :-

not (central-trans ($\boxed{1}$)).

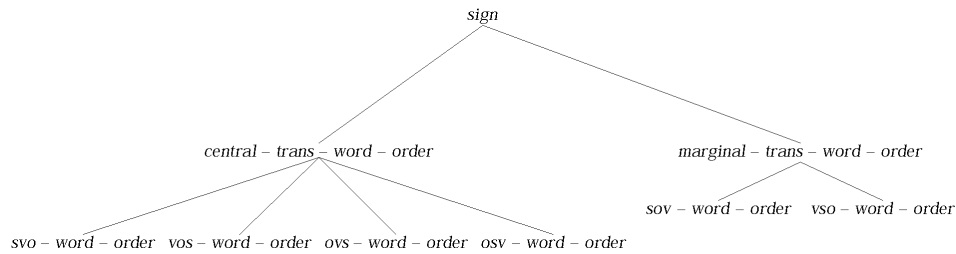
Central transitivity for a predicate is then defined as requiring both agents and patients of neutral topicality, as defined above. For those predicates whose agents and patients are not of neutral topicality, they will be evaluated as being of marginal transitivity.

Now the Polish data can be modeled in terms of these appropriateness conditions. In HPSG terms, we could partition the type *sign* to reflect the possible word orders, and then place again an appropriateness condition on the subtypes of *marginal-trans-word-order* (SOV and VSO orders) such that they should have arguments of non-neutral topicality.

Type Partitions:

central-trans-word-order (subtype of *sign*): *svo-word-order*, *vos-word-order*, *ovs-word-order*, *ovsv-word-order*

marginal-trans-word-order (subtype of *sign*): *sov-word-order*, *vso-word-order*



central-trans-word-order

predicate ([1]),
 central-trans ([1]).

sov-word-order

predicate ([1]),
 $[1] \left[\text{SYNSEM|LOCAL} \left[\text{CONT} \left[\text{AG} [2] \right] \right] \right]$,
 increased-topicality ([2]).

Here an example of an appropriateness condition on word order of type SOV is expressed: the predicate's agent ([2]) should be of increased topicality. Similar appropriateness conditions could then be placed on other subtypes.

It should be emphasized that these are appropriateness conditions and rule out nothing. With increased topicality of agents, Polish speakers could use an OVS word order, but Siewierska's data suggests that most often they employ an SOV word order which the metagrammar deems most appropriate. This can lead to statistical prediction but not absolute prediction.

While we have modeled Siewierska's data within evaluation of this metagrammar, I believe this theory has explanatory power as well.

Principle of Construction Evaluation

Let C be a construction. Then C is valued highly by the evaluation metric if C is headed by a predicate which is either a central or a marginal member of the transitive prototype.

These types are subject to two constraints: one affecting the sign's PHON feature (a matter of observable form), and another affecting SYNSEM (the syntactic-semantic properties of the sign). We would formalize this by placing an ordering constraint on the phonologies on signs of each word order type (in terms of linear precedence constraints or Kathol-style domains), while SYNSEM is constrained by the metagrammar's appropriateness conditions.

The result is that our model predicts that though topicality may be universally important, this fact may be encoded in different ways by different languages. Polish has apparently opted to use word order to encode this aspect of meaning, but the arbitrariness of the sign is reflected in the fact that speakers must learn to associate which grammatical features (such as ordering of the PHON features here in Polish and other word order constraints) with which meanings through appropriateness conditions. In addition, other functional information (such as voice) could be encoded in a similar fashion as outlined above. One language might encode a given meaning through word order, another by means of voice alternation. Like topicality, we want to propose appropriateness conditions on the use of voice but not disallow outright those uses of active or passive voice which are non-prototypical.

Conclusion

This paper has argued that linguistic generalizations such as those of relative transitivity and topicality cannot be captured by transformations, lexical rules, or other derivations, but should rather be modeled as signs weighed by an evaluation metric. Siewierska's (1993) quantitative analysis of word order in Polish suggests that Polish speakers encode unusual topicality for subjects and/or objects by means of two word order patterns. This is formalized by

specifying constraints on a predicate such that the prototypical transitive construction's arguments must be of neutral topicality for their type (subject or object), and then describing these two Polish word order patterns as subtypes of a type that requires their predicates to not be of a more marginal type. While this particular choice may be specific to Polish, it is proposed that other languages would encode marginal topicality in other features of the sign (word order, phonology, morphology) in a manner similar to that outlined above.

Further investigation would examine whether the data for Polish above is reflected in other Slavic languages and in other language families. More study should perhaps shed more light on exactly what the numbers are for evaluating topicality: which features are most important, how it can be computed, etc. Implementation of such a metagrammar in Prolog seems an obvious next step; evaluation of a computational grammar and lexicon should provide insight to these numbers, and perhaps an artificially intelligent routine can make more precise estimates as to the average speaker's topicality thresholds for each of the word orders.

Works Cited

- Abney, Steven. (1996b) "Statistical Methods and Linguistics." In Klavans, Judith and Philip Resnik [eds.] *The Balancing Act* Cambridge: MIT Press.
- Estival, Dominique and John Myhill. (1988) "Formal and functional aspect of the development from passive to ergative systems." In Shibatani, Masayoshi [ed.] *Passive and Voice* Amsterdam: John Benjamins.
- Givón, Talmy. (1983) *Topic Continuity in Discourse: A Quantitative Cross-Language Study* Amsterdam: John Benjamins.
- Hopper, Paul J. and Sandra A. Thompson. (1980) "Transitivity in grammar and discourse." *Language* Volume 56, Number 2, 251-299.
- Keenan, Edward L. (1985) "Passive in the world's languages." In Shopen, Timothy [ed.] *Language Typology and Syntactic Description* Cambridge: Cambridge University Press.
- Murphy, Patrick M. (1995a) "The Theme Construction: Discourse in Information-Based Syntax and Semantics." Presentation to the Spring Linguistics Colloquium, University of North Carolina at Chapel Hill, 1 April 1995.
- Myhill, John. (1992) *Typological Discourse Analysis: Quantitative Approaches to the Study of Linguistic Function* Oxford: Blackwell.
- Pollard, Carl and Ivan A. Sag. (1987) *Information-Based Syntax and Semantics, Volume 1: Fundamentals* CSLI Lecture Notes no. 13. Stanford: Center for the Study of Language and Information (distributed by the University of Chicago Press).
- Pollard, Carl and Ivan A. Sag. (1994) *Head-Driven Phrase Structure Grammar* Stanford: Center for the Study of Language and Information (distributed by the University of Chicago Press).

Siewierska, Anna. (1993) "Syntactic weight vs information structure and word order variation in Polish." *Journal of Linguistics* 29, 233-265.

Svartvik, Jan. (1966) *On Voice in the English Verb* The Hague: Mouton.