

An Architecture for a Universal Lexicon: A Case Study on Shared Syntactic Information in Japanese, Hindi, Bengali, Greek, and English

Naoyuki Nomura, Douglas A. Jones, Robert C. Berwick
Massachusetts Institute of Technology, NE43-802
Artificial Intelligence Laboratory
nomura@ai.mit.edu

Introduction.

Given the prominence of the lexicon in most current linguistic theories (LFG, HPSG, GB), the inventory of language particular information left in the lexicon deserves special attention. Constructing large computerized lexicons remains a difficult problem, building a large array of apparently arbitrary information. This paper shows that this arbitrariness can be constrained more than might have been previously thought. In particular, arbitrariness of argument structure, word sense, and paraphrasability will be shown not only to be constrained, but also to be integrally related. Our (radical) view is that variation of lexical behavior *across* languages is exactly like lexical variation *within* languages, specifically, the difference lies in the presence or absence of certain morphemes. For example, the fact that Japanese has richer possibilities in certain verbal patterns is derived solely from its morphological inventory.¹ Put another way, language parameters simply *are* the presence or absence of lexical material in the morphological component. Observed language variation patterns reflect *morphological* systematicity. The generative machinery for producing argument structure positions is fixed across languages.

Linguistic Motivation.

A striking example underscoring universality of argument structure is the familiar Spray/Load alternation², shown in (1). Despite the many surface differences in these data across languages, they share several essential properties.

- (1) a. John loaded the hay on the wagon.
b. John loaded the wagon with the hay.

Japanese

- (2) a. 太郎はテープを棒に'巻いた。
taroo-wa teepu-o boo-ni maita.
Taro-NOM tape-ACC stick-DAT wrap-PRF
'Taro wrapped the tape around the stick.'
b. 太郎は棒をテープで巻いた。
taroo-wa boo-o teepu-de maita.
Taro-NOM stick-ACC tape-WITH wrap-PRF
'Taro wrapped the stick with the tape.'

¹See Miyagawa, Fukui, and Tenny 1985 for a discussion of this effect. Also see Martin 1975, pp 441-455, for 56 such morphemes. See below for additional discussion of these alternations and for an alternative analysis.

²See, e.g., Levin 1993 and sources cited there, for example, Jackendoff 1990 and Emonds 1991.

Hindi

- (4) a. shyam lathi-ko kagaz-se lape-ta
Shyam stick-ACC paper-with wrap.PRF
'Shyam wrapped the stick with paper.'
b. shyam lathi-par kagaz lape-ta
Shyam stick-on paper wrap.PRF
'Shyam wrapped paper around the stick'

Bengali

- (5) a. jodu lathi-ta kaagajh-die murieche
Jodu stick-DET paper-with wrap.PST
'Jodu wrapped the stick with the paper.'
b. jodu lathi-te kaagaj murieche
Jodu stick-on paper wrap.PST
'Jodu wrapped the paper around the stick'

Greek

- (6) a. ο γιαννης φορτωσε σανο στο βαγονι
o janis fortwse sano sto bagoni
Janis DET load.PST hay on.the wagon
'Janis loaded the hay on the wagon'
b. ο γιαννης φορτωσε το βαγονι με σανο
o janis fortwse to bagoni me sano
Janis DET load.PST the wagon with hay
'Janis loaded the wagon with hay'

All of these languages have exactly the same alternation type. Why? Let us focus on the role of *figure* and *ground*³ in these examples. By "alternation", we mean that in each language the *figure*, as expressed as a direct object in the (a) cases, alternates with the *figure* as expressed in an equivalent PP in the (b) cases. Despite the differences in word order and case marking, all share the property that the direct object is subject to a so-called Holistic Effect:⁴ Crucially, the (a) sentences differ from the (b) sentences in exactly the same way in each of these languages. In (1b), where John loads the wagon with hay, the wagon is understood to have a whole load of hay, likewise if John smears the wall with paint, the

³By 'ground', we mean the surface background involved in the action represented by the verb. By 'figure', we mean the object that is brought into contact with the ground. For example, in (1), the hay is the figure which is brought into contact with the wagon, in this case, the ground. See Talmy 1978 and Emonds 1991 for discussion of figure and ground in this connection.

⁴Note that this property is not overtly grammatically marked, as, say, the case of the direct object is. (see Levin 1993 and the references there for additional discussion of the Holistic Effect.

wall is understood to have more paint than if John merely smears paint on the wall. Thus we may assume throughout that the word sense of the verbs as used in the (a) and (b) cases are essentially identical.⁵ The goal of the remainder of this paper is to analyze and implement this insight in a particular representation given by both linguistic and computational theory, and apply it to MT.

Basic Building Blocks: The Syntax of Word Formation

We propose to replace idiosyncratic language particular information with a new generative component that links universal abstract lexical structures with the surface forms of words for each language. This generative machinery is based on work by Hale and Keyser 1993 and Pustejovsky 1991. The basic architecture is shown in Figure 1.

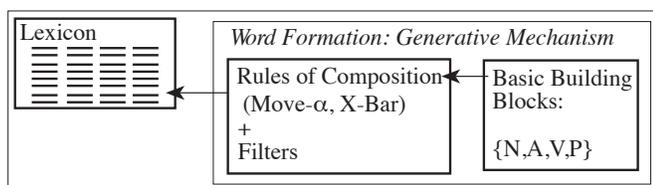


Fig. 1. Generative Syntax of Word Formation

Crucially, only a restricted number of argument structures can be generated. The basic idea is that lexical X-bar structures are *composed* from the lexical categories N, A, V, and P (see fig.2), into trees whose Spec(ifier) and Comp(lement) positions after movement yield the range of possible argument structures. The lexical entries are subject to a series of filters, as follows.

Basic Building Blocks	Rules of Composition
Lexical Categories	• X-bar Projection
N (Noun) = entity	• Move-Alpha
A (Adjective) = state	(including Head Movement)
V (Verb) = event	
P (Preposition) = relation	

Filters (in Hale and Keyser 1993)

- HMC: Head Movement Constraint (Baker 1988)
- ECP: Empty Category Principle (Chomsky 1981)
- FI: Full Interpretation (Chomsky 1986)
- UP: Unambiguous Projection (Kayne 1984)
- PL: Predication Locality (Williams 1980)

To give a concrete example of the system, we derive the thematic properties of the denominal verb *shelve* from compositional machinery operative in the lexicon by composing the noun form *shelf* with an empty preposition and an empty verb to yield the form *shelve*. The structures are as shown in Figure 2. In short, argument structure is produced by syntax operative in the lexicon

that derives words from a small set of primitives.⁶ These structures are assumed to be *identical* across languages.

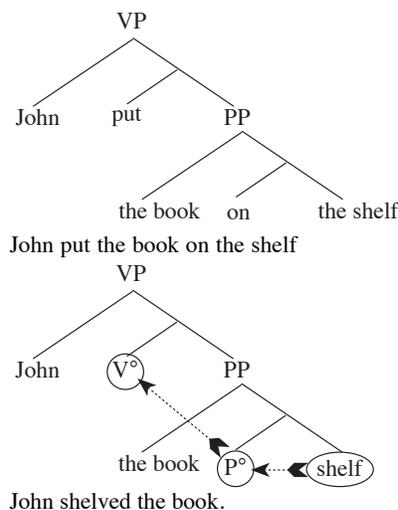


Fig. 2. Derivation of denominal verb pattern.

Figure 3 shows the detailed schema for producing lexical entries for verbs. We can only note briefly here a few important properties of the system in Figure 3. First of all, the main verb V is formed by Head Movement of X, and Y if it exists, to V. The categorial value of X and Y are selected from the set {N,A,V,P}. For example, the denominal verb *shelve* is built as shown above in

⁶ One can view the work by Dorr (1993) and, previously, other attempts at lexical decomposition ranging back through Schank as essentially the same in spirit, but without the detailed constraints provided by Figure 2. We regard similar proposals regarding “promotion” and “demotion” of arguments as essentially a reflection of Move-alpha. The novelty of our proposal is that it is not ad hoc: that is, the same constraints independently justified in syntax also appear in lexical construction. The need for a non-arbitrary, i.e., an explanatory, account of lexical argument structure should be apparent. If lexical entries varied arbitrarily, we would logically expect at least the following space of lexical possibilities, requiring upwards of a quarter-million diacritics. Let m be the number of semantic/thematic roles, such as Agent, Patient, Theme, Range, Duration, and so on, and let n be the number of grammatical functions. Then, when n=4 and m=50 (a typical number for traditional NLP systems) we have 251,176 different

$$\text{types of lexical entries } \left(\sum_{i=0}^n \frac{m!}{i!(m-i)!} \right).$$

While some theories might propose this many distinctions, it seems clear that this imposes a very considerable learning and engineering burden. Many researchers, including Makino 1991 and others, have noticed the drawbacks of encoding thematic roles, but in the absence of a clear alternative, still require them for representing lexical information. Furthermore, it is common knowledge that verbs pattern into certain equivalence classes within languages (e.g., Levin 1993), but arbitrary verb classes would imply arbitrary variation across languages.

⁵ If an interlingua-based system does not constrain the number of word senses, it faces some serious computational problems as is shown in section 3.

Figure 2, by selecting P as X and N as Y. The deadjectival verb *redden* is built by selecting A as X and selecting nothing as Y. The ditransitive verb *give* selects V for both X and Y, following Larson 1988. NP1, if it exists, is the agent of the action, and NP2, if it exists, is the affected object of the verb, following Hale and Keyser (1993).

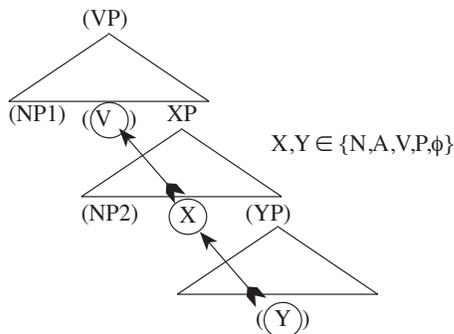


Fig. 3. Schema for producing verb lexical entries.

From the point of view of lexical representation and MT, the key constraint is that the entry for *shelve* has elements that correspond directly to the verb *put* and the preposition *on* in its representation. These elements then become available for interpretation and for translation. We show below that this is also part of the difference between English, Japanese, Hindi, and Greek verbs.

Analysis and Applications for the Universal Lexicon Computational Motivation.

There have been some controversies about the merits and demerits of transfer-based MT and interlingua-based MT. Typical transfer-based MTs prepare completely different sets of word senses for component languages so that mapping among the word senses is completely arbitrary, i.e., the complexity may be calculated as bipartite graph matching. We will assume an interlingua-based MT, which supposedly makes all the component languages share common word senses or so-called concepts and thus is constrained regarding word senses. However, interlingua-based MT still has substantial problems in making up word senses. The number of word senses, their well-definedness, and the problem about linking surface words depend on excellent lexicographers. To give just one example here, the bilingual dictionary Sanseedoo (1990) lists all the following English translations for a Japanese verb *kazaru* (*decorate*).

かざる:

kazaru: ornament; decorate; adorn; dress; embellish; exhibit; display; put articles on show; affect; be affected; grace a ;use fair words; write an inflated style; mince; not being plain or flat

Clustering these into well-defined word senses is not an easy task; thus, it is hard to answer the word sense question. Suppose we have a symbol to represent the core meaning of *kazaru*, which is shared by the English

counterpart *decorate*. Since *kazaru* has the syntactic nature of a Spray/Load type alternation, the lexicon of a typical interlingua-based MT essentially provides the information described below.

word sense:

KAZARU-DECORATE

syntactic information

Verb Alternation Type 1:

{[AGENT]NOM(が),[THEME]ACC(を),
[MATERIAL]WITH(で)}

Verb Alternation Type 2

{[AGENT]NOM(が), [THEME]ACC(を),
[GOAL]DAT/ON/OVER(に)}

AGENT, THEME, GOAL and MATERIAL are thematic roles that are the key elements in the interlingua. NOM, ACC, WITH and DAT/ON/OVER are case-marking functions mapped to the surface case markers 'が', 'を', 'で' and 'に'. Assuming a self-contained thematic role system and case-marking system, these markings are to be evaluated on the corresponding example sentences and be decided independent of each other. However, the two argument structures in the above diagram are actually incompatible with each other because the same thematic role THEME is assigned to different referents: *ground*, the patient to be decorated, and *figure*, the decoration to be attached to the patient. In effect, the MT system makes serious errors due to its confusion of thematic roles derived from the lexicon.

Example Input: 太郎が花を飾った。

Taro-ga hana-o kazatta.

(6) Taro-NOM flower-ACC decorate-PRF

Output: Taro decorated the flower.

In most contexts, the default reading of the input sentence above should be interpreted as 'Taro decorated something with flowers.' This error was caused by the semantic clash in the lexicon. In order to avoid such errors, lexicographers could overwrite some thematic roles disregarding semantic criteria, but thus would spoil the interlingual foundation. The remaining possible solution for this problem is to artificially divide the word sense (WS) into two symbols: *WSwith* and *WSon*, two completely artificial word senses.

Those two artificial word senses are essentially very similar to each other, if not identical, and will pose difficulties for lexicographers because they will have to put arbitrary links among word senses for similar words in two languages or within the same language. The two word senses put the two different argument structures in complete isolation once the analysis is completed and the interlingua is fixed; the only thing the generation module of MT can do is accept the given word sense (*WSwith* or *WSon*) and generate only one argument structure. This rigidity has a potential to generate

are unacceptably unnatural.

The result for MT and lexicon construction is that the computational machinery will stay fixed across languages and thus uniformly constrain the complexity of argument structures eliminating most of the related arbitrariness.

It is well known that word-for-word translations are not the paradigmatic case. The architecture we propose entails a significant improvement in isomorphic mappings between languages. However, the isomorphism is not at the level of words, but rather, at the level of morphological elements that enter into the lexical syntactic formation of words. Thus it is no accident that ‘put the book on the shelf’ is a near paraphrase of ‘shelve the book’, and it is no accident that ‘put the book on the shelf’ is a closer isomorphic map for the Japanese translations of ‘John shelved the book’ shown in (1). The entailment is that ‘shelve the book’ has the same morphological material as ‘put the book on the shelf’, but the former has an empty preposition and an empty verb that incorporates a noun.

In particular, we show how to replace thematic roles with the lexical syntax proposed in Hale and Keyser (1993) and augmented by work in Pustejovsky (1991).⁷ This technique yields several potential benefits: (i) robustness of the lexicon, (ii) greater flexibility in selecting more natural renditions of target language structures in translation, as in (7) below. Let us consider each of these in turn with specific examples.

(7)a. John-wa hon-o tana-no ue-ni oi-ta.
John-TOP book-ACC shelf-GEN upper place-AT put-PRF
John put the book on the shelf.

b. John-wa hon-o tana-ni simatta. (more accurate)
John-TOP book-ACC shelf-DAT put.away-PRF
John put away / stored the book on the shelf.

If paraphrasability and translation must conform to the lexical syntactic structures in Figures 2 through 4, we have a natural method for producing a constrained space of possible translations, namely, the only structures that are allowed are those produced by the mechanism outlined in Figure 4. To highlight the relationship between paraphrasability and translatability, consider the alternation behavior of several verbs within English, shown in (2). Notice in particular that different verbs participate in one or both halves of the alternations, or in neither half. As we will see, the same facts hold across distinct languages.

Nonalternation:

Not all verbs that participate in one half of the Spray/Load alternation participate in the other half, as (8) and (9) and (10) from English, Japanese, and Hindi illustrate. Interestingly, in these cases verbs across

languages also pattern alike in terms of *nonalternation*⁸. This gives additional support for our representation proposal

English

- (8) a. John covered the baby with the blanket.
b. * John covered the blanket {over,onto,...} the baby.

Japanese

- (9) a. 太郎は赤ん坊を毛布で覆った。
taroo-wa akanboo-o moofu-de oot-ta
Taroo-TOP baby-ACC blanket-WITH cover-PRF
'Taro covered the baby with the blanket.'
b. * 太郎は毛布を赤ん坊に覆った。
taroo-wa moofu-o akanboo-ni oot-ta
Taroo-TOP blanket-ACC baby-DAT cover-PRF
'Taro covered the blanket over the baby.'

Hindi

- (10) a. jOn-ne kapre-se bacce-ko dhaka
John-ERG cloth-WITH child-ACC cover
'John covered the baby with a cloth'
* jOn-ne kapre-ko bacce-ke-upar dhaka
John-ERG cloth-ACC baby-ON.TOP.OF cover

Thus there are *four* logical possibilities for alternation and nonalternation, as illustrated in Figure 4.

Type (a)

	<i>Holistic</i>		
Subject	V	NP ₁	[with NP ₂]
	<i>Ground</i>	<i>Figure</i>	

Type (b)

	<i>Holistic</i>		
Subject	V	NP ₂	[[onto/into/...] NP ₁]
	<i>Figure</i>	<i>Ground</i>	

Verb Class i.

- (a) John loaded the wagon with hay.
(b) John loaded the hay onto the wagon.

Verb Class ii.

- (a) * John poured the glass with water.
(b) John poured the water into the glass.

Verb Class iii.

- (a) John covered the baby with the blanket.
(b) * John covered the blanket onto the baby.

Verb Class iv. *

- (a) * John gurgled the glass with water.
(b) * John gurgled the water into the glass.

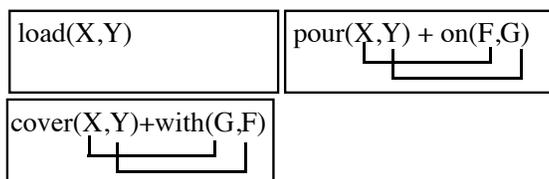
Figure 4. The Spray-Load alternation.

The nonalternations are critical clues to discovering

⁷Along with Emonds (1992), we propose shifting a significant portion of the machinery out of the semantics and into syntax, in the form of syntactically enriched lexical entries.

⁸See Levin 1993 for an extensive survey of such phenomena.

the correct lexical representation for the verbs and hence the structures that translate a more direct mapping across languages. We propose that the key to the solution is that in the non-alternating cases, prepositions are *incorporated* into the verb in lexical syntax just as *shelve* in Figure 3. For example, the lexical entries for *pour*, and *cover* contain prepositions as shown in (10). We assume that the lexical representation for the prepositions encodes specifications for figure and ground, represented as F and G in (10). In essence, what *X on Y* means is that *X* is a *figure* on the *ground* *Y*. In fact, this is the essence of what prepositions "mean", at an abstract level. Thus the prepositions are the primitives in our system, and concepts such as *figure* and *ground* are derivative. This entailment is illustrated by the linking lines in (10).



The lexical entry for *load* does *not* contain a preposition and hence is free to alternate. Thus the reason why a verb does not participate in part of an alternation is that it incorporates lexical material which clashes with potential complements. The reason that 'cover the blanket on the baby' is bad is that *cover* already encodes the figure and ground relationship by incorporating *with* into its representation—that is, the preposition *with* is frozen into the verb's representation. Adding an overt PP with *on* creates a clash in figure and ground relations. Adding a compatible PP to a verbal structure with an incorporated P introduces redundancy, but the structure is still well-formed.

Alternation Mismatches Across Languages

So far, then, we have seen only that verbs across different languages pattern alike. Surely there must be differences or else MT efforts would have succeeded long ago. Figure 6 shows all of the logically possible relationship between verbs across two languages.

Types A, F, and K verbs behave identically in a language pair. For example, the introductory sentences (1)–(6) illustrated Type A correspondences, where $L_1 = \text{English}$ and $L_2 \in \{\text{Japanese, Hindi, Bengali, Greek}\}$. Nonalternating examples (6)–(9) exhibit a Type F correspondence.⁹

We have found (B,C,G,E,I,H) —the unshaded cells in the Figure 5—to be the richest source of cross-linguistic information, namely, the verbs that do not correspond directly are the most informative regarding the nature of word formation, given the model that we have adopted. Not accidentally, they are precisely the ones that are not only of particular interest for our framework, they are potentially very difficult for machine translation, simply because, for some of them, there is no way to take advantage of the surface similarity of argument structures. Recall that we assume that the word sense for these verbs is fixed across languages. Consider, then the Type C alternation correspondence below, where $L_1 = \text{English}$ and $L_2 \in \{\text{Japanese, Hindi, Bengali}\}$:

⁹Clearly, type P verbs should constitute most of the verbal vocabulary, since most verbs do not participate in the Spray/Load Alternation, or in any given alternation, for that matter. Types (D,H,L,M,N,O) do not correspond at all — actually, we expect that these verbs do not exist, given the considerations regarding likely candidates across languages.

Cross-Linguistic Alternation Model: 16 Logical Types for The Spray-Load Alternation				
Classification of patterns (9i-iv) for a given verb in languages L_1 and L_2 into the following types A-P: (Shaded cells are exact crosslinguistic correspondences.)				
Type:	$L_1 \Rightarrow i$ Obj=Ground PP=Figure	$L_1 \Rightarrow ii$ *Obj=Ground PP=Figure	$L_1 \Rightarrow iii$ Obj=Ground *PP=Figure	$L_1 \Rightarrow iv$ *Obj=Ground *PP=Figure
$L_2 \Rightarrow i$: Obj=Ground, PP=Figure	A	B	C	D
$L_2 \Rightarrow ii$: *Obj=Ground, PP=Figure	E	F	G	H
$L_2 \Rightarrow iii$: Obj=Ground, *PP=Figure	I	H	K	L
$L_2 \Rightarrow iv$: *Obj=Ground, *PP=Figure	M	N	O	P
Figure 6. Cross-linguistic alternation model: 16 logical types for the Spray-Load alternation. Classification of patterns in Fig. 4 for a given verb in languages L_1 and L_2 into the following types A-P: (Shaded cells are exact crosslinguistic correspondences.)				

English

- (a) John decorated the wall with posters.
 (b) * John decorated the posters {over,onto,...} the wall.

Japanese

- (a) ジョンは壁をポスターで飾った。
 John-wa kabe-o posutaa-de kazat-ta.
 John-TOP wall-ACC poster-WITH decorate-PRF
 'John decorated the wall with posters.'
 (b) ジョンはポスターを壁に飾った。
 John-wa posutaa-o kabe-ni kazat-ta.
 John-TOP poster-ACC wall-DAT decorate-PRF

Bengali

- (a) raam ghar-e phul sajieche
 Ram room-ON flowers decorate
 'Ram decorated the room with flowers'
 (b) raam phul-die ghar sajieche
 Ram flowers-WITH room decorate
 'Ram decorated the room with flowers'.

Hindi

- (a) raam phul-se ghar sajaya
 Ram flower-WITH room decorate
 'Ram decorated the room with flowers'
 (b) raam ghar-me phul sajaya
 Ram room-WITH flowers decorate
 'Ram decorated the room with flowers'

As in the case of *cover* discussed above, the explanation is quite simple within our framework. For the English verb *decorate*, there is an incorporated preposition, namely *with*, in its lexical representation. There is a type clash because the direct object cannot be both *figure* and *ground*, in the case of “*John decorated posters on the wall”, see Fig. 7.

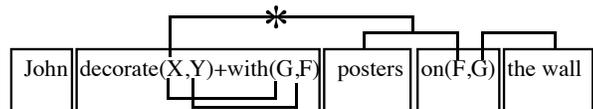


Fig. 7. Type clash with "decorate posters on the wall".

To address one of our main points, cases as in Fig. 7 pose special problems for machine translation because the translation of the word sense of English *decorate* into its Japanese counterpart *kazeru* does not have a similar type clash. Thus the problem is that the English verb has a preposition frozen into its lexical representation, though it is not visible at the surface level. Let us consider another case for MT:

Japanese to English:

この部屋は壁に花を飾ろう。

kono heya-wa kabe-ni hana-o kazar-oo

this room-TOP wall-DAT flower-ACC decorate-LET'S
 (coarsely) translated into:

"As for this room, let's decorate the wall with flowers," but with UL techniques might be more flyently translated as "Let's decorate this room by putting flowers on the wall." Even more broadly, we would like to suggest a promising direction for the development of our system. Consider the well-known difficulty of translating Japanese *-wa* phrases into English. Given the ubiquity of *-wa* phrases as compared with the

relative rarity of English *as-for* phrases, we can conclude that rendering *-wa* as *as-for* is not the best translation.

Let us now consider cases in Japanese and Hindi in which the preposition type element is visible, and which overtly affects the alternation type.

Alternation Type Change:

There are additional crosslinguistic differences, which may be observed in the surface form of the verbal structure. In Japanese, one can add the verbal morpheme *kake* to *oou*. This change, from alternation type F to type H is shown below.

- (a) 太郎は赤ん坊を毛布で覆った。
 taroo-wa akanboo-o moofu-de oot-ta
 Taroo-TOP baby-ACC blanket-WITH cover-PRF
 'Taro covered the baby with the blanket.'
 (b) * 太郎は毛布を赤ん坊に覆った。
 * taroo-wa moofu-o akanboo-ni oot-ta
 Taro-TOP blanket-ACC baby-DAT cover-PRF
 'Taro covered the blanket over the baby.'

- (a) * 太郎は赤ん坊を毛布で覆いかけた。
 * taroo-wa moofu-o akanboo-o ooi-kake-ta
 Taro-TOP blanket-ACC baby-DAT cover-OVER-PRF
 'Taro covered the blanket over the baby.'
 (b) 太郎は毛布を赤ん坊に覆いかけた。
 taroo-wa moofu-o akanboo-ni ooi-kake-ta
 Taro-TOP blanket-ACC baby-DAT cover-OVER-PRF
 'Taro covered the blanket over the baby.'

In Japanese, *kake* adds an aspect of 'trajectory' to the verb sense. More precisely, as the gloss 'over' suggests, *kake* rescues *oou* from its type clash just as the preposition *with* does in English. However, 'cover' cannot be so rescued in English simply because there is no morphological life raft. Note further that the example in (a) also behaves as expected with respect to the Holistic effect. In (a) *akanboo-o* 'baby' is the

direct object, and the baby is understood to be wholly covered. Sentence b' has no such effect regarding the baby.⁹

In Hindi, one can replace *dhaknaa* ('cover') with *dakh-dena* ('give cover'). This morphological changes turn a type F alternation contrast into type B, as described in Fig. 6.

- | | | | | |
|------|--------------------------------------|------------|----------------|----------------|
| (a) | jOn-ne | kapre-se | bacce-ko | dhaka |
| | John-ERG | cloth-WITH | child-ACC | cover |
| | 'John covered the baby with a cloth' | | | |
| (b) | *jOn-ne | kapre-ko | bacce-ke-upar | dhaka |
| | John-ERG | cloth-ACC | baby-ON.TOP.OF | cover |
| (a') | jOn-ne | kapre-se | bacce-ko | dhak-di-yaa |
| | John-ERG | cloth-WITH | child-ACC | cover-give-PRF |
| | 'John covered the baby with a cloth' | | | |
| (b') | jOn-ne | kapre-ko | bacce-ke-upar | dhak-di-yaa |
| | John-ERG | cloth-ACC | baby-ON.TOP.OF | cover-give-PRF |

Put briefly, our view is that variation of lexical behavior *across* languages is exactly like lexical variation *within* languages, specifically, the difference lies in the presence or absence of certain morphemes. Ontologically speaking, then, what language parameters *are* is the presence or absence of lexical material in the morphological component. The observed patterns in language variation is then reflected in *morphological* systematicity. For example, the fact that Japanese has richer possibilities in certain verbal patterns is derived from its morphological inventory. In specific, the reason that it is impossible for English verbs to behave like certain corresponding Japanese verbs is that English lacks an equivalent of the Japanese aspectual morphemes *tsukusu* 'exhaust', *kakeru* 'trajectory verb', etc.¹⁰ But recall, we find that *load*, for example *does* behave precisely like its corresponding verbs in Japanese, Hindi, Bengali, and Greek. In cases where verbs do *not* appear to behave alike, apparent differences are resolved by a process of language particular morphological behavior: for example, the verbal suffixes (and prefixes) of Japanese such as *-tsukusu* 'exhaust' alter verb argument structure enough to bring them into correspondence with their former English non-counterparts.

Conclusion

⁹It might be understood pragmatically to entail that perhaps the parents were worried about covering the baby too much, and wanted to allow the baby to breathe easily by allowing its head, for example, to remain uncovered. In brief, here it is the compositional behavior of morphemes that yields different alternation paradigms.

¹⁰See Miyagawa, Fukui, and Tenny 1985 for a discussion of this effect. Also see Martin 1975, pp 441-455, for 56 such morphemes. See below for additional discussion of these alternations and for an alternative analysis.

We believe that our approach is applicable universally. Future work to be done is to complete our survey of the approximately 150 types of verbal alternations of Levin 1993, and augment our analysis with further ideas from Hale and Keyser 1993, Pustejovsky 1990, 1991, and others, and to extend the coverage to Japanese and other languages. Our highly constrained system should also provide highly desirable circumscription of computational lexicons. Given the universal aspects of our lexical representations, we also expect manageable applications to machine translation, along the lines that we have suggested.

References

- Emonds, Joseph E. 1991. "Subcategorization and Syntax-Based Theta-Role Assignment". *Natural Language and Linguistic Theory*. 9.
- Hale, Kenneth and Keyser, Samuel Jay. 1993. "On Argument Structure and the Lexical Expression of Syntactic Relations". In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Kenneth Hale and Samuel Jay Keyser (eds), MIT Press.
- Jackendoff, Ray. 1993. "On the Role of Conceptual Structure in Argument Selection: A Reply to Emonds". *Natural Language and Linguistic Theory*, 11.
- Levin, Beth. 1993. "English Verb Classes and Alternations ~ A Preliminary Investigation", The University of Chicago Press.
- Martin, Samuel E. 1975. "A Reference Grammar of Japanese", Yale University Press.
- Makino, Takanori. 1991. "Natural Language Processing (in Japanese)", pp.84, Oomura-sha Ltd., ISBN4-274-07658-X
- Fukui, Naoki, Miyagawa, Shigeru, and Tenny, Carol. 1985. "Verb Classes in English and Japanese: A Case Study in the Interaction of Syntax, Morphology and Semantics". *Lexicon Project Working Papers #3*, MIT.
- Pustejovsky, James. 1991a. "The Generative Lexicon" *Computational Linguistics*, 17.4.
- Pustejovsky, James. 1991b. "The Syntax of Event Structure" *Cognition*. 41.3.
- Sanseedoo. 1990 *New Crown Japanese-English Dictionary* Rev.5, Sanseedoo Ltd. Japan
- Talmy, Leonard. 1978. "Figure and Ground in Complex Sentences", in Joseph Greenberg (ed.) *Universals of Human Language: Syntax*, Vol 4. Stanford.