【Forum】

On the Type-wise Productivity of Lexical V-V Compounds in Japanese: A Thematic Proto-role Approach

Kazuhiko Fukushima
Kansai Gaidai University

Abstract: The type-wise productivity of lexical V-V compounds in Japanese is investigated systematically from the perspective of thematic proto-roles. It is shown that the degrees of productivity are sensitive not only to the quantitative advantage of transitive verbs but also to the complexities encountered in the process of argument matching between two verbs to be combined via compounding. The correlation between such complexities and productivity is captured by an optimality-oriented approach that takes advantage of a set of markedness constraints. The constraints are not simply capable of selecting the most optimal candidate of a single comparison, but useful for comparing different winning candidates from separate and mutually independent comparisons as well. The current approach distinguishes itself from those based on the traditional concept of transitivity of verbs, which merely describe—but are not capable of predicting—the observed patterns of productivity.*

Key words: V-V compound, productivity, argument structure, thematic proto-role, optimality theory

1. Introduction

Lexical V-V compounds in Japanese have been investigated by various researchers from various empirical/conceptual perspectives. One of the issues is the matter of productivity of different types of such compounds. Even from a simplistic combinatory perspective based on transitivity of verbs (i.e. TV-TV, IV-IV, TV-IV, and IV-TV combinations[1]), it is known that type-wise produc-

---

* Parts of this paper were presented at the 24th Meeting of the English Linguistic Society of Japan. I would like to thank those who offered suggestions, criticism, and comments at the meeting. In particular, thanks go to Danièle Godard, Shosuke Haraguchi, Nobuko Hasegawa, Hiroto Hoshi, Kazuko Inoue, Taro Kageyama, Peter Sells, Yoko Yumoto, and two anonymous reviewers for *Gengo Kenkyu*. Not all the suggestions, however, have been incorporated into the current version, leaving them to the future and further development of the current project. Any remaining shortcomings are attributed to the current author.

[1] The following abbreviational notations are employed for linguistic terms. NOM: nominative; ACC: accusative; IV: intransitive verb; TV: transitive verb; DTV: ditransitive verb; ARG-ST: argument structure (as used in HPSG); PRES: present tense; PAST: past tense.

tivity of the compounds fluctuates (see section 3). However, other than simply counting the number of such compounds in type-wise fashion, previous studies have not adequately investigated the *factors* giving rise to the observed patterns of productivity. In this paper the expression 'type' is reserved for *the way according to which argument matching between two verbs is carried out* to form a V-V compound (see section 4.2). Sometimes it is used informally (albeit imprecisely) to indicate traditional transitivity-based combinatory patterns. However, the expression will not be used to refer to a semantic 'type' (like 'manner') of a compound verb.

The study reported here opens up discussion about such factors contributing to—and envisages possible explanations for—type-wise productivity fluctuation of lexical V-V compounds in Japanese. This study is intended to be preliminary as well as rather descriptive with its emphasis on reporting the result of a survey covering a large amount of data with both lexical V-V compounds and basic monomorphemic native verbs that are the ingredients for the compounds.

A new perspective is presented (sections 3 and 4) drawing on the recent account of the compounds found in Fukushima (2005) that breaks with more traditional approaches to the phenomenon (section 2). Though Fukushima has presented a comprehensive and far-reaching account for the phenomenon based on the notion of thematic proto-role of Dowty (1991), his account fails to shed any light on the productivity issues. In the current study, this 'deficiency' is ameliorated by employing perspectives of optimality theory (OT) that are suitable for capturing gradience and variation. The central intuition behind the current account for the productivity problem (section 4) is that the more complex the process of argument matching between the component verbs becomes, the more marked a given V-V compound will be.[2]

2. The Point of Departure

For the rest of this paper, I assume (albeit non-exhaustively) Fukushima's (2005) mechanisms for lexical V-V compound formation. His approach does not assume, as the central explanatory ingredients, either (a) the traditional thematic roles, (b) the hierarchy involving such thematic roles, (c) syntactic grammatical relations, or (d) unaccusativity. His approach constitutes the central perspective throughout this paper.

(1) Lexical V-V Compound Formation (adapted from Fukushima (2005: 583))
  a. Compounding: Combine two verbs.
  b. Argument matching:
    • Based on the head's argument structure, find matching arguments from the head and non-head verbs in terms of shared/non-conflicting proto-role entailments (Dowty 1991). The head is the first argument (i.e. '*Q*')

[2] We ignore the matters of lexicalization and accidental gaps (see Bauer 2001 and Lyons 1977).

of the respective semantic schemata in (2b, c). (In dvandva compounds (2a), both first and second arguments are the heads and the two heads are compared.)

- More specifically, take the least oblique argument of the head and find a match starting from the least oblique argument of the non-head verb. If they do not match, dismiss the argument of the non-head verb and move to the next least oblique argument of the non-head. Existentially quantify the variables corresponding to dismissed arguments. The Obliqueness Hierarchy assumed here is that of HPSG (Pollard and Sag 1987 and 1994) with the subject—the leftmost on ARG-ST—is the least oblique element followed by the direct object, indirect object, and other oblique arguments.
- If a match is found, move to the next arguments of the head and non-head and find a match as done above. Continue likewise until all the head's arguments are exhausted. Keep track of all the matches with indexing (with a structure-sharing tag like '[1]').
- After the first match is found, do not dismiss any more arguments from the non-head verb, though a given argument of the non-head verb may not share any proto-role entailment with an argument of the head.

c. Argument structure synthesis: Form a new syntactic argument structure by merging all the arguments of the head and non-head in the order of obliqueness. Arguments sharing the identical index are represented by a single argument with the (default) case-marker value taken from the head's ARG-ST. Do not merge in dismissed arguments.

d. Argument abstraction: Using the semantic schemata in (2), abstract over the (unbound) variables corresponding to arguments (of the head and non-head). Arguments sharing an identical index are abstracted and bound by the same λ-operator.

(2) Skeletal semantics of V-V compounds

a. Dvandva (coordinate) compounds: $\lambda Q \lambda P.(P(x_1...x_n)\ \&\ Q(x_1...x_n))$

b. Manner/means compounds: $\lambda Q \lambda P.(\text{VIA}'(P(x_1...x_m))(Q(x_1...x_n)))$
[N.B.: '$\text{VIA}'(X)$' is a propositional modifier indicating how its argument proposition ('$Q(x_1...x_n)$' here) comes about.]

c. Cause compounds: $\lambda Q \lambda P.(\text{CAUSE}'(P(x_1...x_m), Q(x_1...x_n)))$
[N.B.: The function CAUSE′ takes two propositional arguments, the first of which is designated as the cause for the second.]

(1a) is straightforward. The important feature of (1b) is that a match between arguments of the component verbs is decided based on shared proto-role entailments furnished by the verbs. Arguments of a non-head verb can be dismissed (or ignored) if no match is found in terms of such proto-role entailments. Argument dismissal continues *until* the first match becomes available. When that happens, no more arguments are dismissed regardless of whether or not any other of the head's arguments shares proto-role entailments with those of the non-head verb.

This means that (a) the resulting compound argument structure (ARG-ST) can include arguments that are *not* found within the head's ARG-ST (e.g. *ture-sar* 'take-go.away'), and (b) a non-head verb can possibly 'lose' all its arguments (e.g. *ne-sizumar* 'sleep-become.quite'). These two points are important when we consider the issues of productivity later on.

The syntactic ARG-ST for the entire compound is constructed by (1c). The arguments with the same tag are merged into a single argument and the case-marker values are determined employing the head's case array as the default.

Finally, (1d) handles the semantics of lexical V-V compounds with the help of the semantic schemata (2). Since dismissed arguments are existentially quantified, the remaining unbound variables (arguments) need to be abstracted. A variable in the schemata that correspond to syntactically merged arguments are bound by a single lambda operator.

Here are some non-exhaustive but representative examples/varieties of V-V compounds. The varieties (not types) here mostly reflect semantic properties of the compounds. As it becomes clear below, the current account in section 4 does not reflect these semantic varieties directly. The reason for this is that even within a *single* semantic variety (like 'manner'), different degrees of productivity are observed depending on how argument matching is carried out (e.g. (4) and (5) below).

The first example (3) is the dvandva variety with two heads. The respective arguments of the heads (two NP_{ga}s with proto-agent entailments) are matched as in (3b, c) giving rise to a new ARG-ST (3d) and lambda abstraction yields (3e).

(3) a. Taroo-ga naki-saken-da
Taro-NOM cry-scream-PAST
'Taro cried and screamed.'
b. *nak* (head—the schema (2a)): ARG-ST$\langle \mathrm{NP}_{ga\boxed{1}} \rangle$
c. *sakeb* (head—the schema (2a)): ARG-ST$\langle \mathrm{NP}_{ga\boxed{1}} \rangle$
d. *naki-sakeb*: ARG-ST$\langle \mathrm{NP}_{ga} \rangle$
e. $\lambda x.(\mathbf{nak}'(x)\ \&\ \mathbf{sakeb}'(x))$

Example (4) belongs to the manner variety with two TVs that is, informally speaking, quite common along with (3). Both matching of the arguments (two NP_{ga}s with proto-agent entailments and two NP_{o}s with proto-patient entailments) and abstraction are straightforward as above.

(4) a. Hanako-ga sara-o tataki-wat-ta
Hanako-NOM plate-ACC hit-break-PAST
'Hanako broke a plate by hitting it.'
b. *tatak* (non-head—the second argument of the schema (2b)): ARG-ST$\langle \mathrm{NP}_{ga\boxed{1}}, \mathrm{NP}_{o\boxed{2}} \rangle$
c. *war* (head—the first argument of the schema (2b)): ARG-ST$\langle \mathrm{NP}_{ga\boxed{1}}, \mathrm{NP}_{o\boxed{2}} \rangle$
d. *tataki-war*: ARG-ST$\langle \mathrm{NP}_{ga}, \mathrm{NP}_{o} \rangle$
e. $\lambda y \lambda x.(\mathrm{VIA}'(\mathbf{tatak}'(x, y))(\mathbf{war}'(x, y)))$

What is illustrated by (5) is rather rare (informally speaking again). This particular example is of the manner variety constructed with a TV and an IV. We note that the head IV's ARG-ST is modified in the following way. First, the two NP$_{ga}$s are matched. Second, since a match is found, the remaining NP$_{o}$ argument of the non-head TV is inherited into the ARG-ST of the resulting compound.[3]

(5) a. Ziroo-ga kodomo-o ture-sat-ta
Jiro-NOM child-ACC take-leave-PAST
'Jiro left taking a child with him.'
b. *ture* (non-head—the second argument of the schema (2b)): ARG-ST<NP$_{ga\boxed{1}}$, NP$_{o}$>
c. *sar* (head—the first argument of the schema (2b)): ARG-ST<NP$_{ga\boxed{1}}$>
d. *ture-sar*: ARG-ST<NP$_{ga}$, NP$_{o}$>
e. $\lambda y \lambda x.(\text{VIA}'(\mathbf{ture}'(x, y))(\mathbf{sar}'(x)))$

The next example (6) is classified as the cause variety consisting of a TV and an IV, and demonstrates yet a different way to form a new ARG-ST. This time we witness the dismissal of the non-head's NP$_{ga}$ argument, which does not share any proto-role entailments with the head's NP$_{ga}$ argument. On the other hand, the non-head's NP$_{o}$ and the head's NP$_{ga}$ arguments have some proto-patient entailments in common—a match. The dismissed argument is existentially quantified entailing the existence of some unspecified 'agent' of wearing. Examples like this are very rare compared to other cause variety compounds like *odori-tukare* 'dance-get.tired'.[4]

(6) a. Sebiro-ga ki-kuzure-ta
suit.jacket-NOM wear-get.out.of.shape-PAST
'The suit jacket lost its original shape due to (someone's) wearing it.'
b. *ki* (non-head—the second argument of the schema (2c)): ARG-ST<NP$_{ga}$, NP$_{o\boxed{1}}$>
c. *kuzure* (head—the first argument of the schema (2c)): ARG-ST<NP$_{ga\boxed{1}}$>
d. *ki-kuzure*: ARG-ST<NP$_{ga}$>
e. $\lambda y.\text{CAUSE}'(\exists x.\mathbf{ki}'(x, y), \mathbf{kuzure}'(y))$

Finally, an extremely rare example is found in (7a), which I named 'the mir-

[3] The verb *sar* 'leave' here is treated as intransitive but could arguably be transitive with a 'source' argument like *koko-kara* 'here-from'. If that is the case, the argument structure for the compound will be ARG-ST<NP$_{ga}$, NP$_{o}$, NP$_{kara}$>. This is what is predicted to happen by (1) since the NP$_{o}$ argument of *ture* and the NP$_{kara}$ argument of *sar* do not share any proto-role entailment. They are inherited by the compound separately.

[4] A reviewer suggests that the verb *ki-kuzure* is created by 'back-formation' from its nominal counterpart (as in *ki-kuzure-su*), because one large-scale dictionary (*Kojien*) does not list it as a verb. However, it is indeed listed as a verb by other large-scale dictionaries (at least *Daijisen* and *Daijirin*). To determine if back-formation is applicable, then, it is not sufficient to simply check whether dictionaries list it or not.

ror image' variety. What is noteworthy here is that the arguments of the head and non-head are matched up in the *reverse* order of each other as in (8a, b) instead of (7b, c). Normally (7b, c) is what we would expect. All the examples seen above would turn out to be ill-formed, if reverse argument matching like (8a, b) were chosen. However, it *is* the regular/expected way of matching (7b, c) that turns out to be problematic in this case—the whole compound would end up signifying a state of affairs where Taro yields a book to himself and Hanako to herself, which is a semantic anomaly!

(7) a. Taroo-ga hon-o Hanako-kara yuzuri-uke-ta
Taro-NOM book-ACC Hanako-from yield-receive-PAST
'Taro received a book from Hanako who yielded it to him.'
b. *yuzur* (non-head—the second argument of the schema (2b)): ARG-ST<NP^1_{ga}, NP^2_{o}, NP^3_{ni}>
c. *uke* (head—the first argument of the schema (2b)): ARG-ST<NP^1_{ga}, NP^2_{o}, NP^3_{kara}>

(8) a. *yuzur* (non-head): ARG-ST<$NP_{ga\boxed{3}}$, $NP_{o\boxed{2}}$, $NP_{ni\boxed{1}}$>
b. *uke* (head): ARG-ST<$NP_{ga\boxed{1}}$, $NP_{o\boxed{2}}$, $NP_{kara\boxed{3}}$>
c. *yuzuri-uke*: ARG-ST<NP_{ga}, NP_{o}, NP_{kara}>
d. $\lambda z \lambda y \lambda x.(\textsc{via}'(\mathbf{yuzur}'(z, y, x))(\mathbf{uke}'(x, y, z)))$

A way out is employing the concept of 'alternative lexicalization' which is one of the consequences of Dowty's (1991) Argument Selection Principle. According to that, for example, *buy* and *sell* are an alternative lexicalization of each other. Intuitively speaking, they are two different 'surface' realizations of a single 'abstract' predicate. If the ARG-ST of *buy* is 'reversed', it would become that of *sell*. Both subject and indirect object arguments of these verbs command proto-agent entailments, which allows either of them to surface as a syntactic subject.

The same story is told about the non-head *yuzur* 'yield' that can be viewed as an alternative lexicalization of *moraw* 'be.given'. Taking advantage of this special ARG-ST 'reversal' and to avoid a semantic anomaly, reverse matching of arguments as in (8a, b) becomes possible. *Yuzuri-uke* here is treated like *morai-uke* 'be.given-receive', so to speak. However, we have to note that this is a very special/limited circumstance (involving alternative lexicalization) and happens very infrequently. This point is significant regarding the productivity issues in section 4.

Fukushima's system above treats all the varieties of the compounds *uniformly*—there is a single set of principles (1) (with independent corollaries). According to such an account, different degrees of productivity among various varieties seen above are not recognized as such.

3. Observing the Productivity Numerically

In this section we first observe numerically how the productivity pie is divided among lexical V-V compounds. The results of two investigations about a large amount of data—including both basic and compound verbs—are reported. It turns out that the five examples of argument matching introduced above (or the four types of transitivity-based classification: (i) TV-TV, (ii) IV-IV, (iii) TV-IV, and (iv) IV-TV types) can be consolidated and re-grouped into three basic patterns.

3.1. The current observations

First, the ratios of *basic* intransitive and (di)transitive verbs in Japanese are determined. This is accomplished by counting basic verbs found in *Bunrui Goi Hyo, Zoho Kaitei-ban* (Classificatory Vocabulary Charts, Expanded and Revised Edition) (2004) published by the National Institute for Japanese Language (NIJL). Though the verb chart here is by no means exhaustive, it offers a highly plausible estimate of such ratios due to the large size of the data set. The group of basic verbs consists of 4283 mono-morphemic native verbs. This grouping excludes (a) the combinations of some element and the light verb *su* (like *gakkari-su* 'get.disappointed' and *ryokoo-su* 'travel-do') and (b) derived complex verbs including compound verbs. The ratios are IV: 25% (N=1066) vs. TV: 75% (N=3217), respectively. From this result, we can see that the basic verbs in Japanese are predominantly (di)transitive.

Second, moving over to V-V compounds, I adopt Tagashira and Hoff's (1986) list (N=1157)—one of the most extensive (but not exhaustive) V-V compound lists available in the literature—as the basis for counting different types of lexical V-V compounds.From this list, I have excluded the following items: (a) syntactic V-V compounds, (b) lexical V-V compounds with semantically non-transparent/non-literal components (e.g. *-kom* 'inward-ing' and *tori-simar* 'take-tighten', i.e. 'crack.down.on'), and (c) lexical V-V compounds with semantically 'de-verbalized' V_2s (e.g. *kaki-nagur* 'write-batter', i.e. 'write in uncontrolled manner' (see Matsumoto 1996 and 1998). This shrinks the original list to N=732.

The preliminary count of lexical V-V compounds types (still based on the concept of transitivity) turns out to be: TV-TV: 77%, IV-IV: 14%, IV-TV: 6%, and TV-IV: 3%. This result not only confirms the general tendency observed by Seki (1977) and Toratani (2002) but also accentuates the overwhelming strength of transitive verbs.

However, this transitivity-based outcome leaves some questions unanswered. For one thing, though it is true that the TV vs. IV ratios are roughly three to one (see above), that alone does not explain the pattern of productivity in question. The IV-TV or TV-IV pattern, in particular, should be more productive/frequent than the IV-IV type, since the former is constructed, at least in part, employing quantitatively dominant TVs.

Moreover, a transitivity-based approach is rather simplistic and ignores potential contributions of relevant semantic properties of verbs including semantic roles and other related theoretical concepts (see Kageyama 1993, Himeno 1999, Matsumoto 1996 and 1998, Yumoto 2005, inter alia). For example, Kageyama suggests that the concept of unaccusativity plays a significant role in this connection. Though the adequacy of such a concept is challenged by Matsumoto (1998), unaccusativity does direct our attention to the semantic complexity of IVs, and the necessity for treating them with care. Or consider the traditional θ-roles. They are quite inadequate both empirically and conceptually for the purpose of linguistic investigation (see Ladusaw and Dowty 1988, Dowty 1991, Davis 2001, inter alia). However, they too remind us of the fact that, for example, transitive verbs not only demand two syntactic arguments per se but also assign different semantic roles to these arguments.

For these reasons, it seems that our understanding of the compound phenomenon would be more profound and revealing if we take into account semantic properties of verbs. My suggestion here is that Dowty's theory of thematic proto-roles offers precisely what we need for the following reasons. First, being independently motivated, they avoid all the empirical and theoretical problems encountered by the traditional θ-roles. Second, unlike the traditional θ-roles, they are flexible, and they do not classify verbal arguments discreetly and rigidly. Third, the semantic side of the concept of unaccusativity is an automatic consequence in the thematic proto-role framework—'unaccusative' verbs are just those intransitive verbs calling for a subject argument with proto-patient entailments. It turns out that the productivity problem from the perspective of transitivity pointed out above is resolved by employing a proto-role approach.[5]

3.2. The current observations revamped

To better understand the productivity issues, first, I have further examined the semantic properties (roles) of the basic verbs utilizing again, the vocabulary charts by NIJL. This time the new observations are made relying on the concept of thematic proto-roles. Also we keep in mind that the head verb of a V-V compound plays a significant role in argument matching—(1b) above is head-driven.

Additional observation 1: In the ARG-STs of the basic TVs and IVs, the percentage of NP_{ga}s that are proto-agents is 92%, while only 8% of such NP_{ga}s (all of which are from IVs) are proto-patients. An argument with proto-agent entailments is labeled as 'proto-agent' here for convenience sake (likewise for 'proto-patient'). This means that if we find a basic Japanese verb, its subject is very likely to be a proto-agent.

Additional observation 2: In ARG-STs of the basic IVs, the percentage of NP_{ga}s that are proto-agents is 69%, while 31% of such NP_{ga}s are proto-patients. So, considering only IVs, we find the ratio of proto-patient NP_{ga}s dramatically increases. Though it is generally rare to find proto-patient subjects (see above), the likelihood of finding them is higher with respect to IVs. This suggests that IVs are semantically more complex than TVs.

Moving on to lexical V-V compounds (from Tagashira and Hoff's list), some additional observations are available from the proto-role perspective.

Additional observation 3: Among the lexical V-V compounds of the TV-TV

[5] In this connection, other theories of the lexicon (those supposing 'lexical conceptual structure' (Jackendoff 1983/1997) or 'qualia' (Pustejovsky 1995, Jackendoff 1997)) could be coupled with a transitivity-based approach to give rise to a 'neo-transitivity approach'. Though both empirical and conceptual consequences of these and current approaches are distinct, I do not attempt to explicate them in this short paper.

type, 100% of their NP_{ga} arguments are proto-agents, e.g. *tataki-war* 'hit-break'. There simply is no proto-patient NP_{ga} found in this category.[6]

Additional observation 4: The lexical V-V compounds of the IV-IV type are divided into two kinds. The NP_{ga}s in the heads' ARG-STs are either (a) proto-agents (80%), e.g. *odori-tukare* 'dance-get.tired', or (b) proto-patients (20%), e.g. *umare-kawar* 'be.born-change'. The ratios here are not so drastically different from the ones reported in *additional observation 2* above.

Additional observation 5: For the IV-TV type compounds, 100% of NP_{ga} arguments are proto-agents in the ARG-STs of the head verbs, e.g. *naki-haras* 'cry-make. swollen'. No NP_{ga} is a proto-patient in this environment. This finding is reminiscent of *additional observation 3* above.

Additional observation 6: The head verbs of the TV-IV type compounds allow either (a) proto-agent NP_{ga}s (60%), e.g. *ture-sar* 'take-go.away' or (b) proto-patient NP_{ga}s (40%), *suri-kire* 'rub-cut'.

Tagashira and Hoff's compound verb list is by no means exhaustive and conclusions drawn based on it would not be definitive. However, when viewed from the current perspective, there is an interesting commonality emerging among the four transitivity-based compound types. When the head verbs of compounds are TVs (*additional observations 3* and *5*), the heads' NP_{ga} (i.e. subject) arguments are exclusively proto-agents. This suggests that the TV-TV and IV-TV types can be consolidated into a single pattern. Reflecting this observation, we can re-categorize the compound types into the following three basic patterns which are tentative and descriptive labels for exposition. A pattern here is qualitatively non-uniform (or 'contaminated') due to the fact that it represents a mixture of, for example, the quantitative and semantic complexities that are independent of each other. See section 4 for a more succinct statement regarding the productivity types.

- *Pattern 1*: TV-TV, IV-TV (83%)
- *Pattern 2*: IV-IV (14%)
- *Pattern 3*: TV-IV (3%).

[6] A reviewer points out that *fuki-das* 'blow-emit' (Imaizumi and Gunji 2002) is an 'unaccusative transitive verb', which shows the following alternation pattern: (i) *Kakoo-ga yoogan-o huki-das-u* 'The opening of a volcano spews out lava' as opposed to (ii) *Kakoo-kara yoogan-ga huki-das-u* 'Lava gushes out of the opening of a volcano'. One thing to note here is that according to the approach assumed in this paper, (i) is more basic than (ii). That is due to the fact that when compound formation takes place in the lexicon, these two verbs are *basic* lexical items—namely, regular transitive verbs. We note that *Kakoo-ga yoogan-o huk-u* and *Kakoo-ga yoogan-o das-u* are just fine, while **Kakoo-kara yoogan-ga huk-u* and **Kakoo-kara yoogan-ga das-u* are not well-formed. That means that regular compound formation gives rise to (i) first. And then, a post-compounding process (whatever that may be) converts *huki-das* into a secondary 'unaccusative transitive verb' seen in (ii). I am not going to adumbrate such a process here. Case-marker alternation does take place for V-V compounds. See Fukushima (2005) for lexical case-marker (re-)adjustment exemplified by compound verbs like *katari-akasu* 'speak-spend.a.night'.

Pattern 1 is characterized by the overwhelming quantitative dominance of TVs. When a TV is employed to form a novel V-V compound, the non-head will very likely be a verb with a NP_{ga} argument that is a proto-agent (see *additional observations 1* and *2*). This situation would make argument matching very straightforward. In most cases, the ARG-ST of the head will be retained as is.

In contrast to *pattern 1*, *pattern 2* is more complex. First, it is far less prevalent due to the quantitative inferiority of IVs. Simply, they are not as numerous as TVs. Second, the semantic complexity of IVs—with their NP_{ga} arguments being either a proto-agent or proto-patient—renders argument matching rather complicated. There may be many instances where matching fails. Meanwhile, if a match is found it would be straightforward—between two proto-agents or two proto-patients. In this way the ARG-ST of the head will surface unaltered.

This leaves *pattern 3*, which is even more complex than *pattern 2*. First, *pattern 3* inherits the quantitative disadvantage and the semantic complexities of IVs described above. Second, the correspondence between the input ARG-STs (of the non-head TV and the head IV) and the output ARG-ST (of the entire compound) can be complicated in the following ways. (a) The head's ARG-ST could be modified as in *ture-sar* 'take-go.away' with the NP_{o} argument of (non-head) *ture* 'take' added to the output ARG-ST. Or (b) dismissal of the non-head's argument may take place as in *ki-kuzure* 'wear-get.out.of.shape' with the NP_{ga} argument of (non-head) *ki* 'wear' is abandoned. All this results from complex ways according to which argument matching has to be carried out. Of course, as we see below, alteration of the head's ARG-ST potentially can take place for *patterns 1* and *2* as well. However, it is much more prevalent and drastic for *pattern 3*.

3.3. Summary of the current observations

We have seen that numerically, there indeed are differences in productivity among lexical V-V compounds. It is, however, one thing to simply *detect* and *describe* the patterns of productivity, and it is another to single out the *factors* that are responsible for giving rise to such patterns. And the latter is the subject matter to which we turn in the next section.

4. Capturing the Productivity

Based on the observations above, a descriptive generalization capturing the patterns of productivity is offered in this section employing the perspectives of optimality theory (OT).[7]

An OT grammar supposes Gen as the universal mechanism that generates candidate linguistic forms (sequences of speech sounds, words, argument structures, syntactic structures, etc.). The candidates are tested against a language-particular constraint hierarchy H by Eval, the universal evaluational function. The output of

[7] However, this does not mean the current author accepts/endorses the entire framework of OT unconditionally. The aspects of OT that are suitable for the description of gradience are adapted here.

this process yields the most harmonic candidate (the winner). The constraint hierarchy H consists of faithfulness and markedness constraints. (Additionally, Con, a putatively universal set of constraints, is supposed. See McCarthy (2002) for a concise introduction to OT.) Specifically, a simplified OT-style approach outlined below follows the conception of 'markedness' envisaged by Vogel (2006). Vogel, in offering an account of case conflicts within free relative clauses in German, suggests that the constraints of OT, except for faithfulness constraints, are markedness constraints. Such constraints can be employed to determine relative markedness of a given structure in comparison to others, and are suitable for predicting gradience.

4.1. An OT approach

The main intuition here is: *the more complex the process of argument matching employed for a given V-V compound becomes, the less productive the compound will be*.

Six constraints (9a–g) are proposed by incorporating the effects of (1) in section 2 above and other general constraints concerning case properties, legitimacy of ARG-STs, the correspondence between input and output ARG-STs, semantic congruity, etc.[8] (One more constraint—Anti-alternative Lexicalization (AAL)—is added below.) They are markedness constraints that collectively determine the least marked candidate, namely, the winner.

(9) a. Semantic Incongruity (SI)
b. Proper Proto-role Correspondence (PPRC)
c. Post-match Argument Retention (PMAR)
d. Case Congruity (CC)
e. Argument Dismissal (AD)
f. Head ARG-ST Modification (HAM)
g. Ranking (preliminary): SI >> PPRC >> PMAR >> CC >> AD >> HAM

The first constraint (9a) penalizes semantic incongruity and anomaly. This is a general constraint and its effect is not limited to compound formation. One example of such semantic anomaly is (7b, c) above.

(9b) is a specialized V-V compound constraint. It discourages a structure if there is an overlooked match between arguments with the same proto-role entailments or an erroneous match between arguments with different proto-role entailments.

Structures are favored by (9c) where arguments (of either the head or non-head) are retained *after* a match has already been found between arguments. This is special to V-V compound formation.

(9d) is a general constraint that regulates the adequacy of case properties dis-

[8] The motivation for most of the constraints here is simply inherited from the framework of Fukushima (2005). The ranking of the constraints, however, is another matter. Not unlike in the general OT practice, the ranking seems to be rather post hoc. For now, we can consider the ranking to be an embodiment of descriptive generalization concerning the productivity phenomenon. The search for independent motivation is left for further research.

played by ARG-STs. For example, an ARG-ST like <NP_o, NP_o> or <NP_{ga}, NP_{ga}, NP_{ga}> is not allowed. (An ARG-ST indicates not only the number of syntactic arguments but also the case properties of each argument.)

Another general constraint (9e) prevents dismissal of arguments from ARG-STs—a verb like *kowas* 'destroy' cannot be used as an intransitive verb *as is*.

Finally, (9f) checks if the head's ARG-ST has been altered or not. Possible alterations come in different forms. Inheriting an argument from the non-head verb is one instance of such alternation. Or dismissing an argument from the head's ARG-ST is another possibility. This is a specialized V-V compound constraint.

Obviously, some of the constraints display overlapping coverage. For example, both (9e) and (9f) would apply to an ARG-ST configuration of the head verb where an original argument is missing. However, as seen below, the coverage of the respective constraints is not identical. I retain the most general formulation for each. The (preliminary) ranking of the constraints is in (9g).

Let me demonstrate how the constraints are applied by Eval to the examples introduced above. Our first example is (3) *naki-sakeb* 'cry-scream'. Representative candidate ARG-STs here and below are generated by Gen. The ARG-STs of the head and those of the output are indicated with an underline to make comparisons easy.

(10) *naki-sakeb* 'cry-scream' (*pattern 2*)

	SI	PPRC	PMAR	CC	AD	HAM
a. √ *nak*: $\underline{<NP_{ga\boxed{1}}>}$ *sakeb*: $\underline{<NP_{ga\boxed{1}}>}$ } $\underline{<NP_{ga\boxed{1}}>}$						
b. *nak*: $\underline{<NP_{ga\boxed{1}}>}$ *sakeb*: $\underline{<NP_{ga\boxed{2}}>}$ } $\underline{<NP_{ga\boxed{1}}, NP_{ga\boxed{2}}>}$		*		*		*
c. *nak*: $\underline{<NP_{ga\boxed{1}}>}$ *sakeb*: $\underline{<NP_{ga\boxed{2}}>}$ } (i) $\underline{<NP_{ga\boxed{1}}>}$ or (ii) $\underline{<NP_{ga\boxed{2}}>}$		*			*	*
d. *nak*: $\underline{<NP_{ga\boxed{1}}>}$ *sakeb*: $\underline{<NP_{ga\boxed{2}}>}$ } $\underline{< >}$	*	*		*	**	*

(10b) violates PPRC, CC, and HAM. The two NP_{ga} arguments with proto-agent entailments are not matched. There are two NP_{ga}s in the ARG-ST—a choice allowed only for stative predicates, which *naki-sakeb* is not. The ARG-ST of each head is altered. PPRC, AD, and HAM are violated by (10c-i/ii). The two NP_{ga} are not matched. In both cases one argument is dismissed from one of the heads and its ARG-ST is altered. (10d) flouts SI, PPRC, CC, AD, and HAM. In addition to other violations, the output ARG-ST would ultimately render *naki-sakeb* semantically vacuous. (An empty ARG-ST is not included below.) The most favored ARG-ST is (10a) with no violation and is the winner in this case. Other non-dvandva IV-IV type compounds are handled similarly.

Next we consider (4) *tataki-war* 'hit-break'. (11a) is the winner in this case. In (11b) we see mirror image argument matching (also violating Anti-alternative Lexicalization (AAL) below) and it ends up violating SI and PPRC. The 'agent' of hitting will be the broken object and the 'agent' of breaking is the one who is hit—semantic incongruity. The two NP_{ga}s (proto-agents) on one hand and the two NP_{o}s (proto-patients) on the other are not matched. Rather proto-agents are matched with proto-patients, respectively. The candidate in (11c) is generated by dismissing the NP_{ga} argument of the non-head verb and means something along the lines of 'Due to unidentified person's hitting some object, the separate "agent" of breaking ends up destroying the object.' Though the situation is not semantically and/or pragmatically unimaginable, the ARG-ST is penalized by PPRC and AD. The two NP_{ga}s are supposed to match but do not. And one NP_{ga} is dismissed. (11d) ends up dismissing the NP_{o} argument of the non-head verb and ends up violating PPRC, PMAR, and AD. Other TV-TV and IV-TV types are accounted for in more or less equivalent fashion.

(11) *tataki-war* 'hit-break' (*pattern 1*)

			SI	PPRC	PMAR	CC	AD	HAM
a. √	*tatak*: $<NP_{ga\boxed{1}},NP_{o\boxed{2}}>$ *war*: $\underline{<NP_{ga\boxed{1}},NP_{o\boxed{2}}>}$	$\underline{<NP_{ga\boxed{1}},NP_{o\boxed{2}}>}$						
b.	*tatak*: $<NP_{ga\boxed{2}},NP_{o\boxed{1}}>$ *war*: $\underline{<NP_{ga\boxed{1}},NP_{o\boxed{2}}>}$	$\underline{<NP_{ga\boxed{1}},NP_{o\boxed{2}}>}$	*	*				
c.	*tatak*: $<NP_{ga\boxed{1}},NP_{o\boxed{2}}>$ *war*: $\underline{<NP_{ga\boxed{3}},NP_{o\boxed{2}}>}$	$\underline{<NP_{ga\boxed{3}},NP_{o\boxed{2}}>}$		*			*	
d.	*tatak*: $<NP_{ga\boxed{1}},NP_{o\boxed{3}}>$ *war*: $\underline{<NP_{ga\boxed{1}},NP_{o\boxed{2}}>}$	$\underline{<NP_{ga\boxed{1}},NP_{o\boxed{2}}>}$		*	*		*	
e.	*tatak*: $<NP_{ga\boxed{1}},NP_{o\boxed{2}}>$ *war*: $\underline{<NP_{ga\boxed{1}},NP_{o\boxed{2}}>}$	(i) $\underline{<NP_{ga\boxed{1}}>}$ or (ii) $\underline{<NP_{ga\boxed{2}}>}$		*	*		*	*

The next example *ture-sar* 'take-go.away' (5) is a rather rare instance where a NP_{o} argument of the non-head verb is inherited into the compound's ARG-ST. In this way, even the winner (12a)—the best candidate—ends up violating one constraint, namely HAM. This is in good contrast to the examples in (10) and (11), where no violation is observed for the winning candidates. This aspect is important when we consider the nature of gradience in productivity below. (12b) violates PPRC and AD. A match between the two NP_{ga}s (proto-agents) is ignored and then the head's NP_{ga} and NP_{o} (proto-patients) are matched violating PPRC twice. And then, the NP_{ga} argument of the non-head verb is dismissed. PPRC is disregarded twice by (12c). The two NP_{ga}s should match but they don't. As above, the head's NP_{ga} and NP_{o} should not match but they do. HAM is not satisfied either—the ARG-ST of the head is altered. The NP_{o} argument is dismissed despite the fact that there is a match between the two NP_{ga}s, a violation of PMAR. This is against AD as well.

(12) *ture-sar* 'take-go.away' (*pattern 3*)

		SI	PPRC	PMAR	CC	AD	HAM
a. √	*ture*: $\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle$ *sar*: $\underline{\langle NP_{ga\boxed{1}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle}$						*
b.	*ture*: $\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle$ *sar*: $\underline{\langle NP_{ga\boxed{2}} \rangle}$ } $\underline{\langle NP_{ga\boxed{2}} \rangle}$		**			*	
c.	*ture*: $\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle$ *sar*: $\underline{\langle NP_{ga\boxed{2}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle}$		**				*
d.	*ture*: $\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle$ *sar*: $\underline{\langle NP_{ga\boxed{1}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}} \rangle}$			*		*	

Ki-kuzure 'wear-go.out.of.shape' in (6) is a very rare example. Again even the best candidate of all, (13a), violates AD that is higher-ranked than HAM violated by *ture-sar* above. (13b) could possibly mean that the 'agent' of wearing wears some specific garment and he/she (not the garment) gets out of shape. That state of affairs does not seem to be semantically impossible. This example, however, is in violation of PPRC and HAM. The same basic meaning of (13b) is shared by (13c) (with the object of wearing is unspecified this time, i.e. dismissed and existentially quantified), which ends up violating PPRC, PMAR, AD, and HAM. Finally, in (13d) we see a complete dismissal of the non-head arguments, giving rise to the violations of PPRC and AD.

(13) *ki-kuzure* 'wear-go.out.of.shape' (*pattern 3*)

		SI	PPRC	PMAR	CC	AD	HAM
a. √	*ki*: $\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle$ *kuzure*: $\underline{\langle NP_{ga\boxed{2}} \rangle}$ } $\underline{\langle NP_{ga\boxed{2}} \rangle}$					*	
b.	*ki*: $\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle$ *kuzure*: $\underline{\langle NP_{ga\boxed{1}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle}$		*				*
c.	*ki*: $\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle$ *kuzure*: $\underline{\langle NP_{ga\boxed{1}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}} \rangle}$		*	*		*	*
d.	*ki*: $\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}} \rangle$ *kuzure*: $\underline{\langle NP_{ga\boxed{3}} \rangle}$ } $\underline{\langle NP_{ga\boxed{3}} \rangle}$		*			**	

What we have seen above is how a *single* lexical V-V compound is evaluated with regard to its candidate ARG-STs. From the results, we know which ARG-ST is the most optimal for one particular compound. Does the way of evaluating candidate ARG-STs for a single compound have anything significant to offer when faced with the general patterns (*patters 1–3* above) of productivity? My answer is yes. Vogel (2006) states that "[t]he possibility of *a comparison of all winners* in the OT grammar of a particular language with respect to their markedness is an important feature that distinguishes OT from ordinary models of generative grammar" (p. 257, emphases mine). Moreover, "[m]arkedness can be seen as the correlate of gradience within [an] OT grammar" (p. 253).

Then, in addition to comparing different candidate ARG-STs of a *single* V-V compound, we can also compare the winning ARG-STs of *different* V-V compounds to show the gradient markedness of the results of argument matching, so long as the set of markedness constraints used to evaluate them are held constant. Due to the abstraction from particular forms of compound verbs, such comparisons can be considered not merely comparisons between ARG-STs of different compounds per se but rather between *different mechanisms* of argument matching. The result of the comparison between winning ARG-STs of different compounds is shown in (14).

(14) Comparison of winners from (10) through (13)

	SI	PPRC	PMAR	CC	AD	HAM	*arg-matching types*
a. *pattern 1*							regular (TV-TV or IV-TV)
b. *pattern 2*							regular (IV-IV)
c. *pattern 3*						*	altered ARG-ST (TV-IV)
d. *pattern 3*					*		dismissed arg (TV-IV)

In (14) we abstract away from particular compound verbs (or their candidate ARG-STs) and examine how marked the different *ways* of arriving at a given ARG-ST might be. In this connection, Sells (2001) proposes to compare different *inventories* (or groups) of Cebuano word forms expressing voice, instead of comparing individual word forms. Vogel's and Sells' approaches are, so to speak, 'higher order' applications of OT to which the current project adds another dimension.

As indicated above, the ratios of the *patterns 1–3* are 83%, 14%, and 3%. How do we make sense of these ratios? The following is my story based on the vocabulary survey in section 3 and the markedness constraints (9) in this section.

Frist, we recall that (a) transitive verbs are far dominant quantitatively (75% of all the basic verbs) and (b) a NP_{ga} argument is most likely be a proto-agent (*additional observations 1–2*). Since (b) brings about violation-free argument matching, we expect *pattern 1* to be both quantitatively wide spread and qualitatively unmarked compared to others—*pattern 1* commands the highest productivity.

Second, intransitive verbs are (a) quantitatively inferior as opposed to transitive verbs and (b) semantically complex regarding their NP_{ga} arguments (either proto-agents or proto-patients; see *additional observation 2*). These two factors collectively reduce combinatory possibilities compared to those of transitive verbs. However, when matching two NP_{ga}s that are either exclusively proto-agents or exclusively proto-patients, intransitive verbs give rise to straightforward argument matching. We expect *pattern 2* to be no more marked qualitatively (in terms of the markedness constraints) than transitive verbs but simply disadvantaged quantitatively.

Third, *pattern 3* is more handicapped due to the fact that (a) the head is an intransitive verb with all the complexities described above, and (b) the non-head verb is a transitive verb. Since argument matching is head-driven, if the head's NP_{ga} argument is not a proto-agent, then the non-head's NP_{ga} argument has to be dismissed. On the other hand, if the head's NP_{ga} argument is a proto-agent (yielding a match), then the non-head verb's non-subject argument(s) has/have to be inherited into the compound's ARG-ST, altering the head's argument taking property. Either way, the process of argument matching becomes complex violating the constraints above. *Pattern 3* is qualitatively more marked than *patterns 1–2*.

4.2. Rare compounds and the refined compound types

Thus far I have constructed an OT-style account for the *general* productivity patterns of the V-V compounds. However, there are cases that spill over from the general patterns. That is because complex argument matching, though rare, takes place for *pattern 1/2* as well. One such instance is *ne-sizumar* 'sleep-become.quiet' (15), which belongs to *pattern 2*. Since the head's NP_{ga} argument is (normally) a proto-patient and the non-head's is a proto-agent (with sentience), there is no match. This latter argument is dismissed and existentially quantified.[9]

(15) a. Ie-ga ne-sizumat-ta
house-NOM sleep-become.quiet-PAST
'The house became quiet as (unspecified) sleepers fell asleep.'
b. *ne* (non-head): ARG-ST<$NP_{ga\boxed{2}}$>
c. *sizumar* (head—the schema (2b)): ARG-ST<$NP_{ga\boxed{1}}$>
d. *ne-sizumar*: ARG-ST<NP_{ga}>
e. $\lambda y \exists x.(\mathbf{ne}'(x)\ \&\ \mathbf{sizumar}'(y))$

The results of applying the constraints (9) are given in (16). Here we are back to comparing candidate ARG-STs of a single compound verb. The best candidate is (16a) with one violation (AD). Since *sizumar* 'become.quiet' is not a stative predicate, the case arrangement like (16b), violating CC, is not possible. The existence of an extra argument indicates that the head's argument structure has been altered going against HAM. (16c) is similar to (16a) but the crucial difference between the two is that it is the head's argument that is dismissed in (16c). It then violates one extra constraint, namely HAM, in addition to violating AD as in (16a). A series of constraints, SI, CC, AD, and HAM, are violated by the empty ARG-ST in (16d). All that renders this *pattern 2* compound as marked as (14d), which is *pattern 3*. Not surprisingly, *pattern 2* compounds like this are very rare.

9 Alternatively, the head verb *sizumar* can be construed as indicating an intentional action. Then, the two NP_{ga} arguments (with proto-agent entailments) are matched, avoiding argument dismissal. But that renders the compound semantically distinct from what it is in (15a).

(16) *ne-sizumar* 'sleep-become.quiet'

	SI	PPRC	PMAR	CC	AD	HAM
a. √ *ne*: $\langle NP_{ga\boxed{1}} \rangle$ *sizumar*: $\underline{\langle NP_{ga\boxed{2}} \rangle}$ } $\underline{\langle NP_{ga\boxed{2}} \rangle}$					*	
b. *ne*: $\langle NP_{ga\boxed{1}} \rangle$ *sizumar*: $\underline{\langle NP_{ga\boxed{2}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}}, NP_{ga\boxed{2}} \rangle}$				*		*
c. *ne*: $\langle NP_{ga\boxed{1}} \rangle$ *sizumar*: $\underline{\langle NP_{ga\boxed{2}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}} \rangle}$					*	*
d. *ne*: $\langle NP_{ga\boxed{1}} \rangle$ *sizumar*: $\underline{\langle NP_{ga\boxed{2}} \rangle}$ } $\underline{\langle \; \rangle}$	*			*	**	*

Another highly marked example that does not belong to *pattern 3* is (7a). For this example, an additional constraint is proposed: Anti-alternative Lexicalization (AAL). This penalizes argument matching employing the concept of alternative lexicalization. Given AAL, the final ranking of the constraints is: SI >> AAL >> PPRC >> PMAR >> CC >> AD >> HAM. In (17) comparisons are carried out for different ARG-ST candidates and (17a) turns to be the best one. (Because PMAR and CC are not violated by any of the candidates, they are not included in (17).) Though (17a) and (17b) are on a par regarding AAL, the latter violates HAM in addition. The remaining possibilities in (17c–e), being semantically incoherent, all violate SI, the highest ranked constraint.

(17) *yuzuri-uke* 'give-receive'

	SI	AAL	PPRC	AD	HAM
a. √ *yuzur*: $\langle NP_{ga\boxed{3}}, NP_{o\boxed{2}}, NP_{ni\boxed{1}} \rangle$ *uke*: $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}}, NP_{kara\boxed{3}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}}, NP_{kara\boxed{3}} \rangle}$		*			
b. *yuzur*: $\langle NP_{ga\boxed{3}}, NP_{o\boxed{2}}, NP_{ni\boxed{1}} \rangle$ *uke*: $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}}, NP_{kara\boxed{3}} \rangle}$ } $\underline{\langle NP_{ga\boxed{3}}, NP_{o\boxed{2}}, NP_{ni\boxed{1}} \rangle}$		*			*
c. *yuzur*: $\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}}, NP_{ni\boxed{3}} \rangle$ *uke*: $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}}, NP_{kara\boxed{3}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}}, NP_{kara\boxed{3}} \rangle}$	*				
d. *yuzur*: $\langle NP_{ga\boxed{4}}, NP_{o\boxed{2}}, NP_{ni\boxed{1}} \rangle$ *uke*: $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}}, NP_{kara\boxed{3}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}}, NP_{kara\boxed{3}} \rangle}$	*		*	*	
e. *yuzur*: $\langle NP_{ga\boxed{3}}, NP_{o\boxed{1}}, NP_{ni\boxed{2}} \rangle$ *uke*: $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}}, NP_{kara\boxed{3}} \rangle}$ } $\underline{\langle NP_{ga\boxed{1}}, NP_{o\boxed{2}}, NP_{kara\boxed{3}} \rangle}$	*		*		

Reflecting what has been said so far, (18) is the comparison of markedness regarding how argument matching is accomplished. We note that all these are the winning candidates from the respective evaluation processes (10–13) and (16–17). Again we are comparing the mechanisms generating ARG-STs here.

(18)

	SI	AAL	PPRC	PMAR	CC	AD	HAM	*arg-matching types*
a. *pattern* 1								regular (TV-TV·IV-TV)
b. *pattern 2*								regular (IV-IV)
c. *pattern 3*							*	altered ARG-ST (TV-IV)
d. *pattern 3*						*		dismissed arg (TV-IV)
e. *pattern 2* (15a)						*		dismissed arg (IV-IV)
f. *pattern 1* (7a)		*						mirror image (DTV-DTV)

From (18) we see that it is not merely the transitivity-based patterns that are significant for determining how marked a given argument matching process is. For example, (15a), which is *pattern 2*, can be as marked as *pattern 3*. Also (7a)/(17a), belonging to *pattern 1*, should be less marked than *pattern 3*, though it is more marked than the latter. One thing to note is that it is extremely rare to find examples like (15a) and (7a)/(17a). Usually, for *patterns 1–2* the probability of encountering simpler argument matching is higher than for *pattern 3*. This expectation does not materialize for (15a) and (7a)/(17a), making the examples highly unexpected and more marked than instances of *pattern 3*.

In (18) the notion of transitivity (or grammatical relation) does not play a central role. Nevertheless, I kept the distinction between *pattern 1* and *pattern 2* just to underscore the uniqueness of intransitive verbs vis-à-vis transitive verbs—their quantitative inferiority and the proto-role complexities for their NP_{ga} arguments. In fact, these two patterns are more or less equivalent in terms of markedness as long as argument matching remains straightforward. Also the fact that there are sub-varieties of *pattern 3* indicates that it is a mixed category with regard to markedness. In addition, as mentioned above, the semantic varieties of V-V compounds do not make a difference, either—different degrees of productivity are observed within a single semantic variety (like 'cause' and 'manner'). The type-wise productivity of the compounds, then, should be characterized from the perspective that designates the complexities of argument matching (the final column of (18)) as the primary factor. After we factor out the quantitative discrepancy between intransitive and transitive verbs and the (unresolved) issues concerning lexicalization (i.e. accidental gaps), the degree of transparency in the process of argument matching will emerge as the most significant dimension.

In view of all these considerations, (18) can be reformulated as (19).

(19)

arg-matching types	SI	AAL	PPRC	PMAR	CC	AD	HAM
a. regular argument-matching							
b. altered argument structure							*
d. dismissed argument						*	
e. mirror image		*					

According to (19), (a) regularity (straightforwardness) in argument-matching, (b) alternation of the head's ARG-STS (i.e. slight complexity in argument-matching), (c) dismissal of arguments (i.e. increased complexity in argument-matching), and (d) mirror image matching (i.e. the maximal complexity in argument-matching, a.k.a. alternative lexicalization) are the central factors giving rise to productivity variation among lexical V-V compounds.

5. Concluding Remarks

The new approach outlined above is characterized in the following way. First, it started out investigating the ratios of the basic intransitive verbs as opposed to the transitive ones, which are 25% vs. 75%. This is important for demonstrating the quantitative advantage of transitive verbs. If there are far more transitive verbs, then it is reasonable that *pattern 1* is the most prevalent quantitatively.

Second, it abandons the traditional transitivity-based perspective. Such an account lacks the capability to capture semantic complexities of various sorts that play a significant role. Specifically, the concept of transitivity is dissected by considering the distribution of thematic proto-roles assumed by the arguments of the basic verbs. Transitive verbs are simple in that their NP_{ga} arguments are exclusively proto-agents. In contrast, intransitive verbs turned out to be semantically complex with their NP_{ga} arguments being either proto-agents or proto-patients. This proto-role complexity for IV-IV combinations and quantitative disadvantage are the factors rendering *pattern 2* less productive than *pattern 1*.

Third, the complexity of intransitive verbs proves to be a further obstacle for *pattern 3*. For any *pattern 3* compound, it would be necessary to alter the head verb's ARG-ST, dismiss an argument of the non-head verb, or both. These choices violate the constraints in (9e, f) above, rendering *pattern 3* generally more marked compared to *patterns 1–2*.

Fourth, by removing the factors in (18) unrelated to formation of new ARG-STS, we have finally arrived at the four main factors determining the productivity of the compounds, which classify the compounds into four qualitatively distinct argument matching types. While these four types may not be exhaustive, together they constitute a new perspective through which the productivity issues can be investigated further.

The current project has also addressed the issue regarding how gradience should be viewed in an OT grammar. The outcome here points to the necessity for employing markedness constraints not only for comparing different candidates of a single expression but also for comparing winning candidates from different evaluations.

Even if the approach described above could be on the right track, what has been said is not the entire story. For one thing, the constraints in (9) and, especially its ranking need to be motivated independently. Some of the constraints and their ranking are intuitively reasonable. For example, the workings and ranking of SI (9a) seem to be well-motivated. Why, in contrast, should it be the case that AD (9e) is more marked and ranked higher than HAM (9f)?

For another thing, why should the list of constraints be limited to those in (9)? Wouldn't it be possible or even desirable to incorporate, for example, other lexical semantic constraints noted by researchers like Matsumoto (1996, 1998) (preventing, for example, **tukare-suwar* 'get.tired-sit.down') and Himeno (1999) (preventing, for example, **nobori-agar* 'ascend-ascend' or **nobori-sagar* 'ascend-descend')?

Though there are some loose ends left to be tied up, the current project has opened up a new possibility for investigating the productivity issues regarding lexical V-V compounds in Japanese by going beyond the limit of the traditional transitivity-based approach.

References

Bauer, Laurie (2001) *Morphological productivity*. Cambridge: Cambridge University Press.

Davis, Anthony (2001) *Linking by types in the hierarchical lexicon*. Stanford: CSLI.

Dowty, David (1991) Thematic proto-roles and argument selection. *Language* 67: 547–619.

Fukushima, Kazuhiko (2005) Lexical V-V compounds in Japanese: Lexicon vs. syntax. *Language* 81: 568–612.

Imaizumi, Shinako and Gunji Takao (2002) Goiteki fukugo-ni okeru fukugo jisho [Complex state of affairs manifested through lexical compounding]. In: Takane Ito (ed.) *Bunpo riron: Rekishikon to togo* [Grammatical theory: Lexicon and syntax], 33–59. Tokyo: University of Tokyo Press.

Himeno, Masako (1999) *Fukugodoshi-no kozo-to imiyoho* [Structure and semantic usage of compound verbs]. Tokyo: Hitsuzi Syobo.

Jackendoff, Ray (1983) *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff, Ray (1997) *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Kageyama, Taro (1993) *Bunpo-to gokeisei* [Grammar and word-formation]. Tokyo: Hitsuzi Syobo.

Ladusaw, William and David Dowty (1988) Towards a non-grammatical account of thematic roles. In: Wendy Wilkins (ed.) *Syntax and semantics 21: Thematic relations*, 62–73. San Diego: Academic Press.

Lyons, John (1977) *Semantics*. Cambridge: Cambridge University Press.

Matsumoto, Yo (1996) *Complex predicates in Japanese: A syntactic and semantic study of the notion 'word'*. Stanford: CSLI.

Matsumoto, Yo (1998) Nihongo-no goiteki fukugodoshi-niokeru doshi-no kumiawase [Verb combinations displayed by lexical V-V compounds in Japanese]. *Gengo Kenkyu* 114: 37–83.

McCarthy, John (2002) *A thematic guide to optimality theory*. Cambridge: Cambridge University Press.

National Institute for Japanese Language (NIJL) (2004) *Bunrui goi hyo, zoho kaitei-ban* [Classificatory vocabulary charts, expanded and revised edition]. Tokyo: Dainihon Tosho.

Pollard, Carl and Ivan Sag (1987) *Information-based syntax and semantics Volume one: fundamentals*. Stanford: CSLI.

Pollard, Carl and Ivan Sag (1994) *Head-driven phrase structure grammar*. Chicago: University of Chicago Press and Stanford: CSLI.

Pustejovsky, James (1995) *The generative lexicon*. Cambridge, MA: MIT Press.

Seki, Kazuo (1977) *Kokugo fukugodoshi-no kenkyu* [Studies on Japanese compound verbs]. Tokyo: Kazamashobo.

Sells, Peter (2001) Form and function in the typology of grammatical voice systems. In: Geraldine Legendre, Jane Grimshaw, and Sten Vikner (eds.) *Optimality-theoretic syntax*, 355–391. Cambridge, Mass: MIT Press.

Tagashira, Yoshiko and Jean Hoff (1986) *Handbook of Japanese compound verbs*. Tokyo: Hokuseido.

Toratani, Kiyoko (2002) *The morphosyntactic structure and logical structures of compound verbs in Japanese*. Unpublished doctoral dissertation, State University of New York at Buffalo.

Vogel, Ralf. (2006) Degraded acceptability and markedness in syntax, and the Stochastic interpretation of Optimality Theory. In: Gisbert Fanselow, Caroline Féry, Ralf Vogel, and Matthias Schlesewsky (eds.) *Gradience in grammar*, 246–269. Oxford: Oxford University Press.

Yumoto, Yoko (2005) *Fukugodoshi/haseidoshi-no imi-to togo: Mojuru keitairon-kara mita nichieigo-no doshi keisei* [The semantics and syntax of compound verbs/derived verbs: Verb-formation in Japanese and English viewed from a modular morphological perspective]. Tokyo: Hitsuzi Syobo.

Author's contact information:
Foreign Languages
Kansai Gaidai University
16-1 Nakamiya Higashino-cho
Hirakata, Osaka 573-1001, Japan
e-mail: kaz@kansaigaidai.ac.jp

[Received 23 December 2007;
Accepted 30 June 2008]

【要 旨】

日本語の語彙的複合動詞におけるタイプ別生産性について
——プロト意味役割を用いた説明——

福島 一彦
関西外国語大学

日本語の語彙的複合動詞におけるタイプ別生産性をプロト意味役割の観点から体系的に調査する。ここでの「タイプ」とは，結合される動詞の項を同定する際の異なる同定方法を指す。各種の語彙的複合動詞の生産性の違いは，他動詞の数的優位性のみならず，動詞の統語的項の同定における複雑性に起因すると主張する。そのような複雑性と生産性の相関関係を最適化理論（OT）の有標性制約を利用することで説明する方向性を模索する。このような制約は，ただ単に最適な個別の複合語の形を決めるだけでなく，最適な個々の複合動詞を作り出す各種の過程そのものの比較をも可能にする。本稿の特徴は，語彙的複合動詞の生産性が文法関係の視点のみから単に数量的に記述されてきたことから離脱し，なぜそのような数量的生産性の相違が見られるのか説明を企てるところにある。