Productivity of compounds: an application of construction morphology

Yoshihiko Asao

Kyoto University

September 26, 2008
ICCG-5 @ Austin, TX
Productivity of compounds cannot be adequately described with existing measures.

The idea of construction morphology is useful to describe the productivity of compounds.

These ideas will be illustrated by investigating Japanese verbal compounds.
Question:

How to describe productivity of compounds?
Productivity of compounds has often been described in informal ways.

Are quantitative measures available?

As for derivational affixes, quantitative measures for productivity have been proposed.
Baayen’s productivity measure for derivational affixes (Baayen & Lieber 1991; Baayen 1992)

$$P = \frac{n_1}{N}$$

where

- $n_1$: the number of hapax legomena (words that occur only once)
- $N$: the token frequency
\( \mathcal{P} \) can be naively applied for compounds

<table>
<thead>
<tr>
<th></th>
<th>( \mathcal{P} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominal compounds</td>
<td>0.225</td>
</tr>
<tr>
<td>-tje</td>
<td>0.253</td>
</tr>
<tr>
<td>-ing</td>
<td>0.038</td>
</tr>
<tr>
<td>-heid</td>
<td>0.114</td>
</tr>
<tr>
<td>-er</td>
<td>0.076</td>
</tr>
</tbody>
</table>

(Baayen 1993: 183)

But what does this figure exactly mean?
The high value of $P$ for nominal compounds can mean:

- any pairs of nouns are able to form compounds (when semantic or other conditions are met)

But it is also possible that

- Only a few nouns can occur as second constituents of compounds, and they are very productive

This point is illustrated by Japanese verbal compounds
### Japanese verbal compounds

<table>
<thead>
<tr>
<th>Compound Verbs</th>
<th>Deverbal Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasiri-tudukeru (run-continue) ‘keep running’</td>
<td>sara-arai (dish-wash) ‘dishwashing’</td>
</tr>
<tr>
<td>osi-taosu (push-fell) ‘push down’</td>
<td>te-dukuri (hand-make) ‘handmade’</td>
</tr>
<tr>
<td></td>
<td>mahô-tukai (magic-use) ‘magician’</td>
</tr>
</tbody>
</table>
By adopting Pinker and Prince’s (1991) Dual Mechanism Theory, Ito & Sugioka (2002) claim that both compound verbs and deverbal compounds can be divided into two subtypes: rule-based and analogy-based.
Japanese verbal compounds

productive (rule-based) compound verbs

- hazimeru ‘begin’  0.069
- tudukeru ‘continue’  0.052
- eru ‘be able’  0.040

An example:

tabe-hazimeru
eat-begin
‘begin to eat’

Another example:

guguri-hazimeru
google-begin
‘begin to google’
Japanese verbal compounds

**argument compounds: rule-based**
- gomi-sute (garbage-throw.away) ‘garbage disposal’
- mado-sime (window-close) ‘window closing’

**adjunct compounds: analogy-based**
- te-gaki (hand-write) ‘handwritten’
- hi-yake (sun-burn) ‘sunburn’
But some adjunct compounds are as productive as productive compound verbs, which are considered to be rule-based:

- -umare ‘be born’ 0.073
- -gurasi ‘live’ 0.058

An example:

kare-wa TÔkyô-umare da.
he-TOP Tokyo-be.born be
‘He was born in Tokyo.’

Another example:

kare-wa Tekisasu-umare da.
he-TOP Texas-be.born be
‘He was born in Texas.’
Are these productive adjunct compounds rule-based?

No, according to the previous studies:

..productivity of these adjunct compounds are not due to the rule operating on abstract categories but rather they are formed by analogy based on a specific head, ..

(Sugioka 1996: 237)

How about compound verbs like -hazimeru (begin), which are considered to be rule-based but nonetheless have specific heads?
Japanese verbal compounds

To summarize,

- productive compound verbs like -hazimeru ‘begin’ are considered to be rule-based because their first constituents are open-ended, while

- adjunct deverbal compounds like -umare ‘be born’ are considered to be analogy-based because their second constituents are lexically specified

In fact, both classes behave in the same way
Japanese verbal compounds

This confusion arises because the rule/lexicon dichotomy does not work — they are partially lexical and partially schematic.
Linguistic knowledge can be described by a network of constructions.

A construction is a pair of form and meaning, e.g. 

\[
[[S] [V] [O_1] [O_2] / S \text{ cau}se\ O_1 \text{ to receive } O_2 \text{ by } V\text{ing}]
\]

(cf. Goldberg 1995)

The idea of construction grammar is also useful in morphology (Booij 2005, 2007)

e.g. 

\[
[[V]-er / \text{one who } V\text{s}]
\]
Japanese verbal compounds are instances of partially lexically specified constructions, or constructional idioms

- [[V]-hazimeru / \texttt{BEGIN TO V}]
- [[N]-umare / \texttt{BORN IN N}]

Argument deverbal compounds, such as mado-sime (window-close, ‘window closing’), are instances of the fully schematic construction

- [[N]-[V] / \texttt{VING N}]
The productivity of compounds can be replaced by the strength or degree of entrenchment of each construction.

How to calculate the degree of entrenchment?
A simple simulation model

- Compounds are given to the model one by one, based on actual frequency data
- A construction is established when commonality between compounds is found by the model
- A construction gains strength when it sanctions a compound
- Constructions must be maximally specific; the most specific construction is chosen whenever more than one constructions are available
Simulation

[X]-begin

eat – begin
drink – begin
Productivity of compounds

Simulation

[X]-[X]

eat – begin

drink – begin
Simulation

Procedure

- 10,000 compound verbs are randomly extracted from a newspaper corpus
- Compounds are given to the model
- The degree of entrenchment is calculated for each established construction
- The same procedure was then applied to deverbal compounds
Results

- \([\text{[N]}-\text{[V]}]\) has much higher degree of entrenchment (348) than \([\text{[V]}-\text{[V]}]\) do (120)
- The constructions \([\text{[V]}-\text{hazimeru (begin)}]\) (37) and \([\text{[N]}-\text{umare (be.born)}]\) (22) can be found among the most entrenched partially lexical constructions
- However, some results does not accord well with native speakers’ intuition and Baayen’s \(p\); for example, \([\text{[V]}-\text{komu (get.in)}]\) has higher degree of entrenchment than \([\text{[V]}-\text{hazimeru (begin)}]\), although it is known to be unproductive. This is possibly because the data is too small
Productivity of compounds cannot be fully described by a one-dimensional scale.

Construction Morphology is useful to describe productivity of compounds.

Simulation studies can model how constructions are acquired and used.
My slides will be available at
http://ling.bun.kyoto-u.ac.jp/~asaokitan/

Search my name on the web
Thank you!


References II


