CBA 2010: The workshop on Corpus-Based Approaches to Paraphrasing and Nominalization
Universitat de Barcelona, 1-2 December 2010

Some Creative Aspects of Nominalization: An Analysis of Hapax Legomena in English

Junya MORITA
Kinjo Gakuin University
morita@kinjo-u.ac.jp

1. Introduction

Aim: To elucidate certain creative aspects of nominalization by analyzing the hapax nominals extracted from a large-scale corpus.

Organization: the definition of hapax legomena and method of their research → internal and external structures of hapax nominals → pragmatic functions → relationship of the structures and functions → contextually-motivated hapax nominals

2. Hapax Legomena and the Method of Their Research

2.1. Significance of hapax legomena

Figure 1.   lexicon
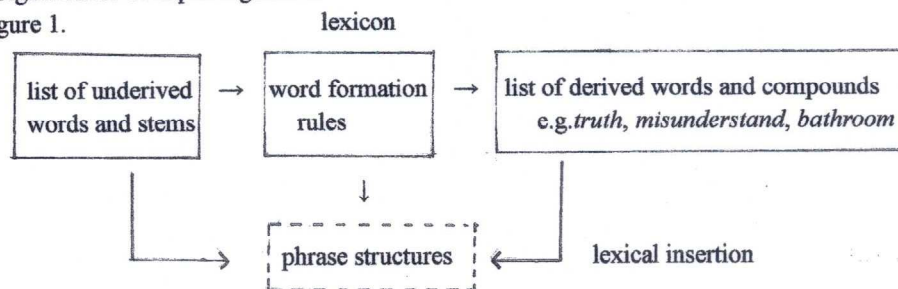
- Regular complex words can be removed from the lexicon, which may be generated on demand by word formation devices.

  e.g. *de-registration*: V + suffix → $[V_{suf}]_N$ *joke-production*: N + $[V_{suf}]_N$ → $[N \ V_{suf}]_N$

- "The list of words which a speaker has at his command at a given moment is not closed and the speaker always has the capacity to make up new words, in particular novel complex nominals (cf. Aronoff (1976: 19)).

- What items are not lexically listed? What outputs of word formation rules are directly inserted into the N-position of a syntactic structure?

→   low frequency words

  reason: A number of experimental evidence shows that highly frequent complex words are typically retrieved from the mental lexicon faster and more accurately than low frequent ones. →   While highly frequent forms are stored and easily accessible, infrequent forms are generally composed by some rule and accordingly their extraction without error would require more time. (Hay (2003: 77-81))

- a most useful criterion of low frequency = the status of hapax legomena

  hapax legomena: a word which occurs only once in a corpus, and is hence coined once for a particular occasion by a single speaker

This extremely low frequency word is crucially used in measuring "productivity"—the probability of encountering new formations (Baayen and Renouf (1996)).

◎Examination of hapax nominals is indispensable for investigating the true nature of creative nominalization.

2.2. Method of research and the result

Method

- By making use of the "wild card" function of the research engine: http://view.byu.edu/ reg3.asp?c =aybfiyfml, the frequency of complex words ending with nominal suffixes has been checked.
- Nominalization is here defined as the process of forming a nominal on the basis of a verb phrase (and so a gerund is left out from our discussion, which correlates with a clause that can encode aspectual distinctions). Our focus is on the nominal which denotes a process or activity.
- The nominal form of [self-V_{suf}]$_N$ (e.g. *self-notification*), the nominal whose base verb is of zero frequency (cf. *disidentify* (token frequency 0) and *disidentification* (hapax)), or the nominal with a minor spelling (e.g. *despatchment*) is not the subject of our investigation.

Result

- 248 types of hapax nominals are identified in the British National Corpus (BNC).
- Of the 248 hapaxes discerned, 170 have no entry in *The Oxford English Dictionary* (2nd ed.). → Hapaxes are principally coined online in working memory.
- The hapaxes are classified in terms of syntactic and discoursal patterns.

3. The Internal and External Structures of Nominalizations and Their Functions

3.1. Internal and external structures

Internal structures

Table 1. *Five types of internal structures and the hapax rate of each type*

Type	Structure	Example	Rate	
I a.	(Det) + derived N	*(the) destruction*	28.6%	32.3%
b.	(Det) + derived N + N	*(the) destruction process*	3.6%	
II a.	(Det) + derived N + of + NP	*(the) destruction of the city*	12.9%	16.5%
b.	possessive + derived N	*(the) city's destruction*	3.6%	
III a.	(Det) + [N + V_{suf}]$_N$	*(the) city-destruction*	35.9%	49.6%
b.	(Det) + [N + V_{suf}]$_N$ + N	*(the) city-destruction process*	13.7%	
IV	possessive + derived N + of + NP	*Caesar's destruction of the city*	1.2%	
V	possessive + [N + V_{suf}]$_N$	*Caesar's city-destruction*	0.4%	

- Type III is the most productive, followed by Type I. The two types make up 80% of the whole hapaxes recorded, and accordingly they are the major devices for online formation of nominals.
- Note a common feature of Type I and III: it is composed of a single word, a derived word (*destruction*) or compound (*destruction process*), optionally accompanied by a determiner.
- Type II with one syntactically realized argument is not so productive, and Type IV and V hapax nominals with an additional argument are unproductive.

External structures

Table 2. *Five types of external structures and the hapax rate of each type*

Type	Syntactic Function	Example	Rate
A	subject	(1)	13.3%
B	complement of preposition	(2a)	54.4%
C	complement of verb	(2b)	26.6%
D	complement of conjunction	(8)	1.6%
E	others	(6c) (head word)	4.1%

· Hapax nominals can dominantly serve as complements of prepositions or verbs when they are used in the formation of larger structures.

3.2. Functional properties

Three communicative functions of hapax nominals: ①discourse cohesion, ②focus, and ③brevity.

①discourse cohesion

(1) It becomes the subject of innumerable short stories and songs, of films that reduce their audiences to tears. And *this glamourization* obliterates the social reality. (BNC HH3: 2652)

        "rheme" ↑_____↑ "theme": pro-form or substitute

            thematic linkage   (Quirk et al. (1985: 1430-31))

· A nominal serves as thematic substitute to contribute to discourse cohesion. → Type I and III internal structures are optimal for substitutes. Type A external structure is optimal for a thematic expression. Thus, these structures are closely related to the function of discourse cohesion.

②focus

(2) a. A child's name should only be removed from the register when it is agreed unanimously at the review that an inter-agency protection plan is no longer necessary to protect the child. The possibility of *de-registration* should be considered at each review. (BNC J76: 388)

   b. Fourthly, no central guidance has emerged on how to coordinate at the local level, care management, the care programme approach, and hospital discharge procedures, thus inviting *triplication of planning effort*. (BNC FT4: 1177)

   c. A further charge is that, being introduced in an era of expenditure restraint, performance measurement has focused largely on measuring, and minimizing, inputs (costs and manpower) … The emphasis is upon *input cost minimization* … (BNC G19: 205)

(2a): the single derived noun *de-registration*, with its direct object *a child's name* unexpressed →
    Type I nominalization therefore serves to emphasize the activity which the noun denotes and de-emphasize the object of the designated activity.

(2b): When the object of nominal is the focus of attention, Type II nominal is chosen following the principle of end-focus.

(2c): Type III (object-incorporating compound) has an important part to play in suppressing the prominency of the object, namely in backgrounding the object (Rice and Prideaux (1991)).

③Type I and III serve as a device of brevity; to construct concise and sensible nominals. A special conception can be produced by compressing a propositional content into a word.

(3) a. No one agrees, for example, on what is "hazardous" waste. By their respective standards, Britain generates only 4m tons of it, America 250m. Nor is there agreement on how to handle it. Britain uses "*co-disposal*," the burial of solid and hazardous wastes together. (BNC HSF: 445)

   b. Without the pinpoint contact of the tips of the claws, the animals may find themselves

slipping and crashing to the ground. The expression of confusion observed on the faces of such cats as they pick themselves up is in itself sufficient to turn any cat-lover against the idea of *claw-removal*. (BNC BMG: 673)

To conclude this section: a major factor influencing the choice among nominal forms has to do with the communicative functions; an optimal nominal form is created in order to package the information in the most appropriate way.

4. Contextuals

- A hapax, an online word coinage, is mainly created depending on context (Clark and Clark (1979)).
- Context-sensitive hapax nominals play a core part of hapax nominalization in performing the previously-discussed communicative functions, in particular discourse cohesion.

4.1. Classification of contextuals

three groups: (a) anaphoric contextuals, (b) cataphoric contextuals, and (c) contextuals requiring the speaker's and hearer's mutual knowledge

(a) anaphoric contextuals: (i) the ones which are directly derived from the preceding VP structures as in (4), or the ones which are constructed depending on the prior correspondent phrases as in (5).

(4) But unconsciously, the object, as an indispensable element of a painting was also discredited; it is *this "discreditation" of the object* which has dominated art for the last half century—it all goes back to Monet! (BNC E9W: 25)

- this: a referring item or anaphoric determiner, which requires the referent or antecedent of the anaphor (Schmid (2000: 27-28))
- the noun *discreditation* appears in quotation marks → a temporarily created true contextual

(5) a. All in all, an ever-growing institutional centralization was matched by a burgeoning theological *pluralization*. (BNC CRK: 47)

   b. Disclosing things about yourself often encourages other people to open up too. But in inappropriate situations and when talking with someone who is quite guarded, *over-disclosure* may be interpreted as intrusive. (BNC CEF: 1400)

(b) cataphoric contextuals: (i) the process is given before its explanation as in (6b), and (ii) "headline" use in (6c).

(6) a. … those plants … should experience enhanced growth with increasing carbon dioxide in the atmosphere (this is termed the *carbon fertilization* … (BNC GU5: 663)

   b. These ideas have, anyway, suffered a *gallicization*, which places the French at the centre of the cosmos. (BNC FAJ: 2170)

   c. *Repavement*: All pavements/sidewalls in Los Angeles are in dire need of resurfacing. (BNC CAL: 970)

(c) contextuals requiring the speaker's and hearer's mutual knowledge: An addresser can express with a nominal the event which (s)he assumes is known to the addressee.

(7) It's taking the form of inquiries about *your US naturalization*. (BNC FYV: 1872)

Result 1

More than half of the attested hapax nominals are contextuals: 135 (54.4%).

→ Given that hapax nominals are online new coinages, nonce nominal formation is principally made context-dependently.

Result 2

The number of each type is: Type a (110 (81.5%)), Type b (17 (12.6%)), and Type c (8 (5.9%)).

→ anaphoricity of nonce nominal formation: it plays a crucial role of affording a concise form of

the propositional unit introduced into the prior context.

4.2. Relationship of contextuals to internal and external structures
internal structures and contextuals

Table 3. *Three main types of internal structures and their context-dependency*

Type	Structure	Example	Context-dependency (D)
I a.	(Det) + derived N	*(the) destruction*	0.662 } 0.663
b.	(Det) + derived N + N	*(the) destruction process*	0.667
II a.	(Det) + derived N + of + NP	*(the) destruction of the city*	0.500 } 0.561
b.	possessive + derived N	*(the) city's destruction*	0.778
III a.	(Det) + [N + V$_{suf}$]$_N$	*(the) city-destruction*	0.472 } 0.472
b.	(Det) + [N + V$_{suf}$]$_N$ + N	*(the) city-destruction process*	0.471

Context-dependency (D)=N_C/N_H, where N_C is the number of contextuals and N_H is the number of hapaxes.

◎Type I and Type IIb (passive nominals) are highly context-dependent.
→ This reflects the function of <u>discourse cohesion</u> which they carry out.
· Type I (cf. example (1)): Since the complement of the Type I nominal has to be retrieved for its interpretation, it naturally relies for information on the context, so that the linkage between the nominal and the preceding sentence is naturally established.
· Type IIb (passive nominals, cf. example (8)): All the passive nominals attested turned out to contain possessive pronouns (e.g. *its reidentification*), which shows that the pronouns naturally link to their antecedents for the identification of the nominal complements.

External structures and contextuals

Table 4. *Four main types of external structures and their context-dependency*

Type	Syntactic Function	Example	Context-dependency (D)
A	subject	(1)	0.636
B	complement of preposition	(2a)	0.489
C	complement of verb	(2b)	0.530
D	complement of conjunction	(8)	1.000

◎Type A and Type D are highly context-dependent.
→ This reflects the function of <u>discourse cohesion</u> which they carry out.
· Type A (cf. example (1)): Subjects tend to look for anaphoric linkage to the prior materials, since subjects are unmarked topics (Lambrecht (1994: 132)) and topic expressions designate the topic referents anaphorically (Lambrecht (1994: 187)).
· Type D: The conjunctions involved are all those which introduce temporal adverbial phrases, as in (8). A temporal adverbial clause or phrase normally expresses what is known to the addressee, hence providing the temporal background for the main clause (Lambrecht (1994: 125)).
(8) ... which was situated near the church where <u>she eventually had her cell</u>: ... But there is no evidence that Julian had ever been a nun *before her immurement* ... (BNC CD4: 467)
To conclude section 4.2: Contextuals reflect how closely the structural properties of nominals are bound up with a function of nominals — discourse cohesion.

4.3. Contextuals and a morphological constraint

"No Phrase" constraint: A phrasal category cannot become a word-internal element (cf. Roeper and Siegel (1978: 202)). *[large bank]er *[every animal] eating (dinosaur)

BUT, phrase-incorporating words are coined often provisionally in a discourse.

(9) a. Even soil with a "fairly high clay content" became acidified at depth. Dry deposition of sulphur dioxide may, says a DoE review, also "cause rapid *surface soil acidification* (over months rather than years)."(BNC AM4: 1392)

    b. Because it contains material about the Ardakkeans that might stimulate your sense of humour. ... Even an advanced Intelloid can not be properly programmed with an inclusive appreciation of *human laughter-stimulation* ... (BNC G3G: 702)

(9a): · To achieve the function of discourse cohesion, a nominal compound (*soil acidification*), but not nominal phrase or clause, should be employed as an anaphoric linker of two sentences.

    · To have the discourse informative, the contrast between the present process (*to be acidified at the surface*) and the previously mentioned process (*to be acidified at depth*) has to be brought out, leading to more specific description of the first element of the compound (*surface soil*).

(9b): Humor includes a lot more than laughing. *laughter-stimulation* → *human laughter-stimulation*

· In sum, the manifest of an attribute of word-internal noun against the morphological constraint is motivated by the functions of discourse cohesion and informativeness of a discourse.

5. Conclusion

· An in-depth analysis of the hapax nominals identified in the BNC

→ Some creative aspects of the online nominalization depicted in Figure 1 have been disclosed: its optimal inner- and outer-structures, context-dependent communicative functions, and the close interrelationship of the forms and functions.

References

Aronoff, Mark (1976) *Word Formation in Generative Grammar*, MIT Press, Cambridge, MA.

Aronoff, Mark (1980) "Contextuals," *Language* 56, 744-758.

Baayen, R. Harald and Antoinette Renouf (1996) "Chronicling *The Times*: Productive Lexical Innovations in an English Newspaper," *Language* 72, 69-96.

Clark, Eve and Herbert Clark (1979) "When Nouns Surface as Verbs," *Language* 55, 767-811.

Hay, Jennifer (2003) *Causes and Consequences of Word Structure*, Routledge, New York.

Lambrecht, Knud (1994) *Information Structure and Sentence Form*, Cambridge University Press, Cambridge.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik (1985) *A Comprehensive Grammar of the English Language*, Longman, London.

Rice, Sally and Gary Prideaux (1991) "Event-Packing: The Case of Object Incorporation in English," *BLS* 17, 283-298.

Roeper, Thomas and Muffy E. A. Siegel (1978) "A Lexical Transformation for Verbal Compounds," *Linguistic Inquiry* 9, 199-260.

Schmid, Hans-Jörg (2000) *English Abstract Nouns as Conceptual Shells*, Mouton de Gruyter, Berlin.

Corpus and Dictionary

The British National Corpus

The Oxford English Dictionary on CD-ROM (Version 2.0), 1999, Oxford University Press.