



HAL
open science

Semantics of the Sino-Japanese derived noun ' verbal noun + sha (“person”)

Raoul Blin

► **To cite this version:**

Raoul Blin. Semantics of the Sino-Japanese derived noun ' verbal noun + sha (“person”) '. 2017.
hal-01571625

HAL Id: hal-01571625

<https://hal.science/hal-01571625>

Preprint submitted on 4 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Semantics of the Sino-Japanese derived noun <verbal noun + *sha* (“person”)>

Raoul BLIN
 CNRS-CRLAO
 blin@ehess.fr

Abstract: We provide a set of rules to predict to which argument of the embedded verbal noun the Japanese derived noun <verbal noun - *sha* (“person”)> can be co-referent with. The rules are based on the *system of argument structures* of the verbal nouns but we do not use the distinction external/internal argument.

Keywords: Japanese, morphology, syntax, nominal derivation, verbal-nouns, system of argument structures

1. Introduction

In Japanese, at first glance, most occurrences of the derived nouns of the form

< verbal noun - 者 (*sha*, ‘person’) >

refer to the agent (or proto-agent in Dowty, 1991 ; *dōsashu* in Kageyama, 2002) of the event expressed by the embedded verbal noun. For example, <*sanka*_{VN}+*sha*> (*sanka*: ‘participation; to participate’) refers to the agent of the event expressed by the embedded verbal noun *sanka* (e.g. 1). Thus, the few studies (partly) consacred to this derived form mainly focused on this interpretation (Miyajima, 1997; Kageyama, 2002; Ono, 2016).

- (1) *gēmu no < sanka*_{VN} - *sha > wa minna kaetta.*
 game GEN <particip._{VN} - person> TOP all leave
 ‘All the participants to the game have left. ’

When observing a large corpus, agent interpretation is certainly the most frequent type of interpretation. But we cannot infer any general rule from this observation. Indeed, as we shall see in this paper, vn-*sha* refers to one of the human arguments of its embedded verbal noun. Most of the verbal nouns suffixed by -*sha* have only one human argument, and this argument is the agent (see section 4.1). Accordingly, the portion of vn-*sha* forms subject to agent interpretation is the highest portion. If we put aside these ‘obligatory’ cases of agent interpretation, interpreting vn-*sha* is not as straightforward as it might appear. First, some derived nouns can refer to the patient of

the embedded verbal noun (e.g. 2), or to other roles, like the recipient or the beneficiary (e.g. 3).

(2) 処刑, *shokei*: 'condemnation'

argument structure: { agent , patient }

shokei-sha: litt. 'condemnation - person', 'convicted' ; refers to the patient

(3) 輸血, *yuketsu*: '(blood) transfusion'

argument structure: { agent , (object) , beneficiary/recipient *ni* }

yuketsu-sha: litt. 'transfusion-person' ; can refer to the beneficiary 'recipient' .

Such non-agentive interpretation was indicated by (at least) by Sugimura (1986) who gave only one example: *taihosha* 'arrested person'. But can we consider this case to be an exception, just like the author. A systematic analysis must still be conducted to build the complete list of complements (argument or circumstantial). Some derived nouns have more than one possible interpretation. Such is the case for *rachi-sha* ('kidnap+person'), derived from the verbal noun *rachi* ('kidnapping'). Without context, the Japanese informants interpreted the derived noun as the agent, i.e. the 'kidnapper'. But given the following context and distribution (e.g. 4), they all interpreted the derived noun as being the patient. All of them considered this sentence to be natural.

(4) Context: *nihon to kitachôsen ga saikin okonatta kaidan de*,¹

'In the talks which have been recently conducted
by Japan and North Korea

nihonjin rachi -sha mondai ga gidai ka sare,...

Japanese kidnap-person problem NOM discussed

the issue of Japanese abductees will be discussed.'

We will qualify as 'ambiguous' any derived noun open to various interpretations depending on the context. The ambiguity raises at least two questions: (i) Does a verbal noun which has more than one human argument systematically produce an ambiguous derived noun in *-sha*? (ii) Does the ambiguous derived noun always refer to only one of the arguments (as in the case above) or can it also refer to the union of all the human arguments?

To date, the derived form <verbal noun - *sha*> in Japanese and the equivalent forms in Western languages have mainly been studied in the framework of referential semantics, by focusing on

1 <http://japan.dailyink.com/japanese/read.php?cataId=nk00400&num=16594> (18/07/2013) ; original

sentence is 日本と北朝鮮が最近行った会談で、日本人拉致者問題が議題化され ; Also:

<http://kokkai.ndl.go.jp/SENTAKU/sangiin/180/0081/18004160081003c.html> 有田芳生君 '過去清算と拉致者問題を同時に解決する'

which thematic roles of the embedded verbal noun the derived noun refers to (for example Fradin (2005) for French, Booij (1986) for English and Dutch). We consider that this approach is unsuitable, as the very notion of thematic roles is not clearly defined (Danlos 2000), and there is a lack of complete lists (Dowty 1991). In order to bypass this problem of definition, we will describe the meaning of *vn-sha*, within the framework of what we could call an *inferential semantics*. This consists firstly of syntactically describing the argument structure(s) of each verbal noun. By using the particle *wich* marks each complement (argument or circumstantial) in an unambiguous way², the descriptions are free of uncertainty. Secondly, we determine the logical (inferential) relationships between the derived form and the arguments of its embedded verbal noun. For example, rather than saying that *rati-sha* designates the agent, the patient or some dative role, we will say that in context, *rati-sha* is coreferent with either the argument designated by the *ga*-argument of the verb in the context or the *wo*-argument, regardless of the “role”. To that end, we use tests based on substitutions of words and inferences between sentences. We will describe those tests in section 3. This approach is all the more interesting given that, on the one hand, there are many exhaustive syntactic descriptions of the argument structures of verbal nouns in Japanese. On the other hand, it is easier to perform tests based on substitutions and inferences. Indeed, it is easier to ask a (non-linguist) informant whether a given sentence can be deduced from another sentence in a given context, than to ask if a given word expresses the ‘role’ of ‘(proto-)agent’ or ‘(proto-)patient’. Lieber (2004) provided a theoretic system based on an ontology framework. However, as with systems of semantic roles, the entire ontology framework is not given, nor is its (logical) consistency proven.

Some linguists have tried to provide a general explanation of the interpretation of suffixes like *-sha*. In English, Levin & Rappaport (1988) suggested that *-er* refers to the external argument, independently of the semantic roles. Roy & Soare (2014) conducted a similar analysis for French. This hypothesis is interesting as it is based on syntactic properties rather than semantic criteria. Unfortunately, this analysis does not seem appropriate for Japanese. The first reason is largely exemplified in this article: *vn-sha* can be co-referent with the arguments marked by the particles *to*, *ni* or *wo*, which are usually considered to be internal arguments. Ono (2016) showed that this hypothesis does not hold, even for agent nominals. The second is a highly sensitive issue: the notion of internal/external argument is correlated to the existence of a verbal phrase within the meaning of generative grammar. However, the existence of such a constituent is not widely recognised for Japanese, even in generative grammar (Inoue 1976). In a short discussion focusing on agent interpretation, Kageyama (1999) suggests analysing the derived form within the framework of the generative lexicon (Pustejovsky 1991) but does not address any other arguments or explain how to proceed. Few other theoretic frameworks have been suggested, mainly focusing

² There are only a very low number of verbs and verbal nouns which admit two arguments with a same particle.

on how to discriminate between agent/experiencer/instrument interpretations in English. By adopting a specific theoretical framework, we run the risk of establishing rules which will be difficult to export within other theoretical contexts. Accordingly, in this paper we strive to present a surfastic description that could be widely used, regardless of the theoretical point of view and final use. In other words, we inscribe ourselves in a taxonomic rather than a theoretical approach (see for discussion Danjou-Flaux & Fichez-Vallez, 1985). But, even if we give up providing a single reason to observed phenomena, we do not just provide an unstructured collection of facts. We aim for a 'most efficient set of rules' (a minimal set of rules for a maximum coverage of data). To be compared to other descriptions, we furnish quantified evaluations of this set.

The first purpose of this study is to offer an exhaustive list of the possible interpretations of the derived nouns <verbal noun - *sha*> by examining a large number of cases. Apart from a very small number of lexicalised cases, we assume that the set of possible interpretations of a derived noun <vn-*sha*> is produced following specific rules. The second purpose of the study is to identify those rules. This entails determining which data to input in the derivation rules to produce the set of possible interpretations. We will not try to determine which possible interpretation should be applied to which context. That is a different area of research. In this paper, we will provide rules for which the input data are the syntactico- and logico-semantic properties of the argument structure of the embedded verbal noun.

As is frequently the case in linguistics, our research enters a grey area where some derived forms or interpretations are neither widely accepted nor rejected, or for which the information is contradictory. As we would like our results to be used also in applied linguistics, such as the machine interpretation of texts, we cannot ignore such cases that necessarily occur when processing large corpora. In this paper, we will also strive to describe such cases. To this end, in addition to the traditional qualitative description, we will also deal with quantitative data to help us make decisions while deciding whether or not to accept difficult cases and how to interpret them. By doing so, we follow authors like (Yatabe 1999) who provide quantitative evaluation of acceptability. But we hope to do one more step ahead by explicitly explain the causal relation between the quantitative data and our theoretical decision.

This paper is divided as follows: In the second section, we will set out the main morphological properties of the derived noun. The third section is dedicated to the presentation of the methodology. The fourth section will then comprise the analysis.

To simplify the presentation, we will use the following abbreviations: 'vn' signifies 'verbal noun', whereas 'vn-*sha*' designates the derived noun < verbal noun - *sha* >. Given two statements P1 and P2 and a context, 'P1 \Rightarrow P2' means that P2 can be inferred from P1 in this context. 'x-argument' means 'the argument marked by the case particle x'. 'x-interpretation' means that the derived form is co-referent with the x-argument. We use the term 'co-reference' in a naive way. Indeed, we understand two phrases in a given context to be co-referent if they refer to *exactly* the same thing

(entity, set of entities, notions, etc.). We write structures of n arguments as follows: $\{ \text{arg}_1 p_1, \dots, \text{arg}_n p_n \}$. arg and p are respectively the arguments and their case particle. It is understood that all the argument structures are given for the active form of the verbal noun. The order of the arguments is not relevant here. Instead of 'arg', we use 'H' to indicate that an argument is exclusively human. For example, one of the argument structures will be written $\{ H1 \text{ ga}, H2 \text{ wo} \}$. In order to help the reader who is not familiar with Japanese, in some glosses, we indicate the semantic role which is usually attributed. However, the information has no theoretical value.

2. Field of study

In order to clearly define the scope of the rules we will provide in this study, we will describe in detail the derived forms on which we will focus, together with their components.

2.1. *Morpho-syntactic properties and semantic category of <vn+sha>*

Let's first define the derived nouns <nv+sha> we study. (1) They are of the form <verbal-noun - suffix *sha* (者)>. (2) They belong to the category of common nouns. (3) For each derived noun, at least one interpretation is human. Among the numerous derived forms, we will restrict our focus to (4) the ones belonging to the Sino-Japanese stratum. Let's detail each point of the definition:

1) The *vn-sha* is usually described as a derivation obtained by concatenating the Sino-Japanese suffix *sha* to an autonomous verbal noun. We will detail the properties of the two constituents in the following sections (2.3 and 2.4).

2) The *vn-sha* belongs exclusively to common nouns. They must then be distinguished from equivalent derived nouns in French. In French, some of them can be either nouns or adjectives (Dubois & Dubois-Charlier 1999, etc.)

(5) *chahut-eur*

Adjectival use: Julien est un enfant *chahuteur* ('Julien is a rowdy child')

Nominal use: Julien est un *chahuteur* ('Julien is a [person/child who is] rowdy')

3) With the exception of a very small number of cases we consider to be idiomatic, *vn-shas* refer to humans. This means that *vn-shas* are hyponyms of *hito* ('person'). As any [+human] common noun, it can be counted with the classifier *nin / ri*:

(6) *go nin no sanka -sha shika inakatta*

5 nin GEN particip.-person only were

'There were only 5 participants.'

vn-sha can also refer more generally to moral persons. See for example *sanka-sha* (litt. 'participation - person'; 'participant'), which can refer to humans, countries or companies. As the notion of 'moral person' is difficult to define, we will focus in this paper on the 'human' interpretation. *vn-sha* differs from its Western equivalent derived nouns, many of which can refer also to instruments and tools:

- (7) English: *mix-er* ('tool to mix', 'mixer')
 French: *broy-eur* ('tool to grind', 'grinder')
 Dutch: *zend-er* ('tool to transmit', 'transmitter')

4) Although the derivation applies to the verbal nouns of all the lexical strata of Japanese (native Japanese, Sino-Japanese, *gairaigo* 'foreign words'), we will restrict this study to Sino-Japanese verbal nouns. Indeed, unlike for native Japanese verbal nouns, the *vn-shas* derived from Sino-Japanese verbal nouns are numerous and involve all the varieties of argument structures. In addition, contrary to foreign verbal nouns, the argument structures of the Sino-Japanese verbal nouns are exhaustively described (Case Frame Database, Kawahara & Kurohashi (2010); Verb Thesaurus, Takeuchi, Inui, Takeuchi, & Fujita (2008) Goitaikei, Ikehara et al. (1997), etc.). Our long-term objective is to extend the system of interpretation rules to all the verbal nouns of all strata. We will then take care to establish the rules independently of the lexical stratum.

To our knowledge, there is no exhaustive, large-scale study about *vn-sha*, and *a fortiori* about the inferential relations between *vn-sha* and the arguments of the embedded verbal noun. In a generalist monolingual dictionary (e.g. Daijirin, Matsumura & Sanseido Editor, 2006; Daijisen, Editor group of "Daijisen" & Matsumura, 1998; etc.) few definitions paraphrase the agent interpretation by using a gloss of the form: VN-*sha* = VN *suru hito/mono* ('a person who does something').

- (8) *kaitaku - sha = kaitaku suru hito (Daijirin)*
 'clear - person = person who develops [something new].'

This paraphrase can be used to detect the agent interpretation but cannot be used for other interpretations.

Sugimura (1986) has a more pragmatic approach and indicates a tendency to refer to what he calls *shutaisha*. This term is not standard in Japanese linguistics or in general linguistics. Moreover, the term is very ambiguous. Following the general definition in monolingual dictionaries, it designates the 'agent' of an event. So it is close to the usual analysis in linguistics, but not sufficient, given that other interpretations are possible. In philosophy, it refers to the very subjective 'main person' of an event. Unfortunately, Sugimura does not provide details or associate linguistic

structures to illustrate this notion. As such, we cannot use this analysis as is.

2.2. Productivity

To our knowledge, there is no quantitative study about the productivity of the *vn-sha*. Generally, it is accepted to be a productive derived noun (see Sugimura (1986)). There is much information in support of this view. Our corpus (corefjp 0.002.140619, 50 million sentences³) contains about 1503 different *vn-shas*, derived from Sino-Japanese *vns*. According to the lexicon jaLexgram-0.10 (Blin 2015), of the 8,202 verbal-nouns for which the lexical stratum is given, 6807 (83%) belong to the Sino-Japanese stratum. Accordingly, at least one-fifth of the (Sino-Japanese) *vns* can be derived with *-sha*, which is a high ratio. Furthermore, this number far exceeds the number of lexicalised derived nouns, around 78 in the generalist Daijirin dictionary (Matsumura 2006), which contains many old constructions, and five entries in the mecab-naist-dic (0.4.3-20080812)⁴. Moreover, the high productivity is confirmed by the apparition of new semantically compositional *vn-shas* derived from a verbal noun imported from foreign languages: *kôchingu-sha* ('person who coaches' or 'person who is coached'), or based on mixed compound verbal nouns : *fûru.katsudô - sha* (litt. 'full [foreign word] activity [Sino-Japanese word] person', 'a fully active person').

The productivity varies depending on the lexical stratum. Nouns derived from native *vns* are very few in our corpus. This suggests that productivity is low. We consider them as idioms. The *vn-shas* derived from foreign *vns* are also few. This can be explained by the relatively low number of foreign *vns*. However, their number will increase in the future, as will the number of derived forms. *vn-shas* derived from Sino-Japanese *vns* are far more numerous. Although no more new *vns* are imported from Chinese, native speakers of Japanese still create neologisms from which *vn-shas* are derived.

2.3. Verbal nouns

Japanese verbal nouns (*dômeishi*) are syntactically defined as follows⁵. (1) They can be used as nouns. (2) When the support⁶ verb *suru* is post-posed, they constitute a verbal phrase. The *vns* category is large (more than 13,000 in the jaLexgram (Blin 2015); more than 17,000 in mecab-naist-dic, including redundant entries) and open. As seen above, a majority belongs to the Sino-Japanese lexical stratum.

In this study, we will focus on the Sino-Japanese *vns* (Kobayashi 2004) of two sinograms⁷. The

3 http://rkappa.fr/sagace/tutoriel/sagace4-2/adj-i_frequences.php

4 We only counted the *vn-shas* built with *vns* which are categorised as verbal-nouns in the mecab-naist-dic. mecab-naist-dic can be download for free at <http://osdn.jp/projects/naist-jdic/downloads/32511/mecab-naist-jdic-0.4.3-20080812.tar.gz>

5 Among many others, see for example (Tsuji-mura 2013) for a simple description, and (Ohara 2000) for a formal and exhaustive description.

6 We intentionally avoid the term 'light verb' to sidestep the debate about the 'light' or 'heavy' nature of *suru* (Kobayashi 2004). The distinction between them is not relevant to our study.

7 Sino-Japanese words are obtained by concatenating morpho-phonological and semantical units that are

list of vns we used to extract derived vn-*shas* (see section 3.1) contains vns from one to three sinograms but mainly two. In the corpus, no verbal noun with more than two sinograms is post-posed by *sha*. On the other hand, some authors (for example Miyake, 2010) consider vns with one sinogram as verbs instead of vns. In this study, we prefer to avoid uncertain data. We have thus excluded the Sino-Japanese verbal nouns/verbs of one sinogram. In the corpus, of the one-sinogram vns that appear in a vn-*sha*, only *shi* ('death') can be used as a noun.

2.4. Morphological properties of *sha*

The suffix⁸ *sha* ('person', mostly written with the sinogram 者) is a dependant morpheme. It cannot be used as a noun, and it cannot be determined by the demonstrative *sono* ('this'): **sono sha*. It then differs from morphemes like *ryô* ('quantity') that are morpho-syntactically more ambiguous. Indeed, *ryô* can be used as a suffix (e.g. *osui-ryô* 'quantity of contaminated water') or as a noun (*sono ryô*, 'this quantity'). Like other Sino-Japanese suffixes such as *shitsu* ('room') and *kan* ('building'), *sha* can be used as an auxiliary of numeral (i.e. *josuushi* see for example an exhaustive list in Iida, 2003). It is a 'morphological' auxiliary in the system of Blin (2006). Table 1 provides a comparison between *sha*, *shitsu* and *ryô*

	<i>sha</i> 者 'person'	<i>shitsu</i> 室 'room'	<i>ryô</i> 量 'quantity'
Suffix	<i>kenkyû-sha</i> litt. 'research + person' 'researcher'	<i>kenkyû-shitsu</i> litt. 'research - room' 'laboratory'	<i>osui-ryô</i> litt. 'contaminated water - quantity' 'quantity of contaminated water'
Noun:	* <i>sono sha</i> * 'this <i>sha</i> '	?? <i>sono shitsu</i> 'this room' ⁹	<i>sono ryô</i> 'this quantity ; the quantity of'
Numeral auxiliary	<i>môsikomi ga 5sha wo koeta...</i> ¹⁰ registra. NOM 5 <i>sha</i> O exce. '... registrations exceeded 5 persons'	<i>kenkyû-shitsu 5 shitsu wo seibi</i> res.-room 5 shitsu O prepa. 'to prepare five laboratories'	?? <i>osui -ryô 3 ryô</i> cont. water-quantity 3 quantity ?? '3 quantities of contaminated water'

Table 1: Comparison of -*sha*, -*shitsu*, *ryô*

Masuoka & Takubo (1992) sub-classify *sha* as a 'nominal suffix' (*meishisei setsubiji*) by contrast to the 'adjectival suffix' (*keiyôshi setsubiji*). Accordingly, *sha* belongs to a very large and heterogeneous category, including register modifier suffixes (*san*, *sama* etc.) which have no semantic value. We do not believe this very heterogeneous sub-classification to be useful in our analysis.

sha can suffix non-verbal nouns, like the common noun (cn) *gengogaku*: *gengogaku-sha* (litt.

specific to the Sino-Japanese stratum, and according to rules that are almost specific to this stratum. In Japanese, because these units are mostly transcribed with one sinogram, by convenience, the length is usually expressed in number of sinograms, even in linguistics.

8 For a recent overview about apparented suffixes, see for example (Sugioka & Ito 2016)

9 About this case, see a discussion in Blin (2006).

10 <http://www.city.ota.gunma.jp/005gyosei/0100-001tosi-plan/files/propo-5syasennte.pdf>

'linguistics - person', 'linguist'). Like *vn-sha*, these derived nouns designate humans. It is important here to note that the noun *gengogaku* does not have any argument structure. This means that *sha* does not necessarily deal with the arguments of the word it suffixes. We would like to attribute a unique representation to *sha*, although it occurs in different derivations (<*vn-sha*> and <*cn-sha*>). To do this, we attribute to *sha* the common semantic properties of the two derived nouns, indeed the +human feature. If we reduce the meaning of *sha* to this simple interpretation, then what is the source of the other semantic specificities of the two derived nouns ('[person] who practices' in <*cn-sha*>; '[person] who is the argument of the embedded vn' in <*nv-sha*>)? We exclude that they come from the radical (resp. the *cn* and the *vn*), as we assume that such a radical cannot contain all the interpretations of all their possible derivations. Consequently, we have to accept that these semantic specificities are carried out by the derivation rules, not by the suffix itself. This hypothesis means that we do not consider that these derived nouns are semantically compositional in the narrow sense of the term: in our analysis, their meanings are not entirely carried out by their components. It is also against the lexical hypothesis which consists in attributing all the morpho-syntactic and semantic information to the representation of the lexical entries, like in Combinatory Categorical Grammar (see for example Bekki 2010).

The present study strictly focuses on *sha* and does not take into account the morphs that could be considered as variants. The first one of these possible variants is the suffix *sa*; e.g. *kujû-sa* (久住者 'A person who has been practising his religion in a temple for a long time'). The derivation in *-sa* is not productive and cases are rare. *sha* is vocalized in a very few number of words (three cases in the general dictionary *Daijirin*, Matsumura & Sanseido Editor, 2006): *sankan-ja* (算勘者, 'person who calculates'), *jikyô-ja* (持経者, 'person who carries sūtra»), *shyugen-ja* (修験者, 'ermite'). Some *vns* can be suffixed with both the vocalized form and the non-vocalized form of *-sha*, such that both derivations are synonymous. It is the case of *annai* ('to guide, to inform'): *annai-sha* (litt. 'person who informs', 'a guide') and *annai-ja* (same meaning, lexicalised in the *Daijirin* dictionary). We assume that the derivation with the vocalized form is not productive and not predictable. We consider the vocalized form as well as the previous forms as idiomatic forms. The rules we set forth in this paper do not apply to them. Based on its distributive and semantic properties, *sha* is also close to the native Japanese *mono*, usually transcribed in *hiragana*, and sometimes with the sinogram 者. *mono* suffixes the radical of the native Japanese verbs. For example, the noun *abare mono* (litt. 'to be turbulent + *mono*', 'turbulent person') is derived from the verb *abare-ru* ('to be turbulent'). We do not exclude that *sha* and *mono* are allomorphs. For methodological reasons, we prefer to distinguish between them and focus on *sha*. We hope this study can contribute to the debate on the similarity of these two suffixes.

As we mentioned above, *sha* has some semantic similarities with the French suffixes *-eur*, *-é*, *-aire*, the English suffix *-er*, etc. But there is no one-to-one correspondence. The semantic spectrum of *sha* in *vn-sha* seems larger than that of the French and English suffixes. For example,

the semantic values of *sha* in *vn-sha* are covered by several suffixes in French. *sha* is close to *-eur* and *-ant* (e.g. *invit+ant* ‘invitee’) for non-instrumental agent interpretation. For patient value, it is closer to *-é* (e.g. *invit+é* ‘invited’). When designating the beneficiary/recipient, it is close to *-air* (e.g. *destinat+aire* ‘recipient’). Since the correspondence between Japanese and European suffixes is not one-to-one, we will focus our observations on Japanese data. We are not trying to correlate Japanese data to European languages data; such a correlation must be made later.

3. Method

In this section, we present the method of investigation. We first build an exhaustive list of attested *vn-shas*. Secondly, for each meaning of each verbal noun embedded in these *vn-shas*, we define the argument structures. We then analyse with which of the arguments the derived *vn-sha* can be co-referent.

3.1. Building the list of *vn-shas*

We listed the *vn-shas* occurring in a rich, large-scale and heterogeneous corpus comprising 10 million sentences. The research has been limited to the occurrences of the *vn-shas* in the scheme <verbal noun + 者 + particle>. The occurrences of *vn-sha* not followed by a particle have been left out, but we do not expect that this to have any significant impact on the list. We have used the list of vns provided in the dictionary *mecab-naist-jdic* (0.4.3-20080812). It contains around 8,000 Sino-Japanese vns, including semantic compositional entries that have not been taken into account (see section 2.3). After a manual post-analysis process, we obtained a list of around 1,503 *vn-shas*. The total number of occurrences is around 285,000.

3.2. Determining the argument structure of the verbal nouns

For each meaning of each verbal noun that occurs in the *vn-shas* of the corpus, we (manually) listed all the possible argument structures which contain at least one human argument. The reason for not including just any argument structure is based on the following rule we put forward (about valuation vns, see discussion in section 4.13). This rule is induced from our observations on corpus.

RULE 1 : Apart *valuation verbal nouns*, there is no verbal noun with non-human argument that can produce a human *vn-sha*.

Because of a lack of place, we do not demonstrate the following rule, which is also induced from observations:

RULE 2 : Vn-*sha* can be only co-referent with the argument of the verbal noun, not the circumstantial complement.

Unlike existing databases of argument structures (Case Frame Database, Verb Thesaurus, Goitakei, etc.), we do not only attribute a list of argument structures to each verbal noun. We also determine the logical relationship between the argument structures of said list. For each meaning of each verbal noun, the set of the list of argument structures and the logical relationship between these structures will be called a *system of argument structures* related to this meaning and verbal noun. Before showing examples, let us present our terminology.

Given the meaning of a verbal noun and the associated argument structure, we call 'implementation' of this structure any sentence where the verbal noun occurs within this meaning and within this argument structure. The following example (9) represents the implementation of the argument structure { H1 *ga* , H2 *wo* } of the verbal noun *tôroku* ('to enrol').

- (9) *tanaka sensei*_{H1} *ga* *gakusei*_{H2} *wo* *jugyô ni* *tôroku* *shita*.
 Takana prof. NOM student O course in *enrolled*
 'Professor Tanaka enrolled the students to the course.'

For the sake of simplicity, given a specific verbal noun, we will say that from an argument structure AS1 one can infer an argument structure AS2 if, in any context, from any implementation of AS1 one can infer an implementation of AS2. We will then write $AS1 \Rightarrow AS2$. We would like to highlight the fact that this formula is just an abbreviation. It must not be interpreted as a formula of formal logic. We assume that in such a formula, the occurrences of a same symbol designate the same argument. For example, $\{ H1 \textit{ ga}, H2 \textit{ wo} \}_{AS1} \Rightarrow \{ H2 \textit{ ga} \}_{AS2}$ means that from the argument structure { H1 *ga* , H2 *wo* } one can infer the argument structure { H2 *ga* } where the *wo*-argument of AS1 is identical to the *ga*-argument of AS2.

Let us now describe for example the *system* of argument structures of the verbal noun *tôroku* meaning '*to inscribe; to enrol*'. This verbal noun has at least two argument structures. The first one has only a *ga*-argument, whereas the second one has a *ga*-argument and a *wo*-argument. All the arguments are humans. The list of argument structures of *tôroku* with the meaning 'to enrol' is:

- { H1 *ga* } (= AS1)
 { H1 *ga* , H2 *wo* } (= AS2)

In any context, from any implementation of AS2 we can infer AS1 where the *ga*-argument is the *wo*-argument of AS2. For example, from (9) we can infer (10):

- (10) *gakusei* *ga* *tôroku* *shita*.
 student NOM enrolled

'The student enrolled.'

To account for this phenomenon, we will add to the list of argument structures an inference rule to show the logical relation between those two structures:

$$\{ H_1 \textit{ga} , H_2 \textit{wo} \}_{AS2} \Rightarrow \{ H_2 \textit{ga} \}_{AS1}$$

As we will see in the following sections, this systemic description will be useful in the analyses.

3.3. Determining the referential value of *vn-sha*

To determine with which argument of its embedded *vn* the derived *vn-sha* is co-referent, we used some tests of inferences. We submitted the tests to Japanese native speakers.

3.3.1. Test with copulae sentence

Given a *vn* and its argument structure $\{ \textit{arg}_1 p_1 , \dots , \textit{arg}_n p_n \}$ for any *i* from 1 to *n*, we consider that *vn-sha* is co-referent with the argument *i* if from any implementation IMP of the argument structure at least one of the following inferences is true:

$$\begin{aligned} \langle \textit{arg}_i p_i \dots \textit{arg}_n p_n \textit{nv light_verb.} \rangle_{IMP} &\Rightarrow \langle \textit{arg}_i \rangle \textit{ga} \langle \textit{vn-sha} \rangle \textit{copula.} \\ &\quad \text{'The } \langle \textit{arg}_i \rangle \textit{ is the } \langle \textit{vn-sha} \rangle \textit{'}} \\ &\Rightarrow \langle \textit{vn-sha} \rangle \textit{ga} \langle \textit{arg}_i \rangle \textit{copula.} \\ &\quad \text{'The } \langle \textit{vn-sha} \rangle \textit{ is the } \langle \textit{arg}_i \rangle \textit{'}} \end{aligned}$$

For example, let's take the *vn shidô* ('to direct (a student)'). We want to know with which argument of *shidô* the derived form *shidô-sha* ('director') is co-referent. Only one argument structure is associated with the meaning 'to direct (a student)': $\{ H_1 \textit{ga} , H_2 \textit{wo} \}$. It contains two human arguments: H1 and H2. The first is marked by the case particle *ga* (corresponding to the grammatical subject). The second is marked by the particle *wo* (corresponding to the grammatical object, the person who is directed). As shown in the following example, the test succeeds only for argument H1 (*ga*). We then conclude that *shidô-sha* can be co-referent only with the *ga*-argument of the *vn shidô*.

(11) *tanaka sensei ga tarô wo shidô suru.*

Takana prof. NOM Tarô O direct

'Professor Tanaka is the director of Tarô.'

⇒ (12) *tanaka sensei ga shidô-sha da.*

Takana prof. NOM direct-person do

(litt.) 'Professor Tanaka is the direction-*sha*'

'Professor Tanaka is the director.'

- ⇒ (13) *tarô ga shidô -sha da.*
 tarô NOM direction-person COP
 (litt.) 'Tarô is the director.'

3.3.2. Test with substitution:

Some copulae-sentences produced in this way may be unnatural (but not ungrammatical or logically inconsistent). In this case, we will suggest one more test. Let's consider a verbal noun VN, AS={ arg₁ p₁ , ... , arg_n p_n } one of its argument structures, a context *C* in which the implementation IMP of AS is true, IMP1 a statement which includes VN-*sha* and which is true in *C*, and IMP2 a statement obtained by substituting one of the arguments of VN in IMP to VN-*sha*. For example, let's consider the sentence (12) as IMP, assumed to be true in a context *C*. VN is *shidô*. IMP1 will be illustrated by (15). We obtain IMP2 (resp. IMP2') by just copying IMP1 in which we substituted to VN-*sha* (i.e. *shidô-sha*) the *ga*-argument (resp. *wo*-argument) which appeared in IMP:

- (14) IMP1: [*tarô no*] <*shidô-sha*> *ga naita.*
 [*Tarô GEN*] *direct -person NOM cried*
 'The director [of Tarô] cried.'
- IMP2: <*tanaka sensei*> *ga naita.* = IMP1 [*tanaka sensei* ← *shidô-sha*]
 Tanaka prof. NOM cried
 'Professor Tanaka cried'
- IMP2': <*tarô*> *ga naita.* = IMP1 [*tarô* ← *shidô-sha*]
 Tarô NOM cried
 'Tarô cried.'

If IMP2 is logically equivalent to IMP1, we can conclude that in the context *C*, VN-*sha* is co-referent with the argument which has been substituted for it. For example, in *C*, IMP2 is logically equivalent to IMP1. We can then conclude that *Tanaka sensei* (i.e. the *ga*-argument in IMP) and *shidô-sha* are co-referent. On the contrary, however, IMP1 and IMP2' are not equivalent. We can then conclude that *shidô-sha* and Tarô (i.e. the *wo*-argument) are not co-referent.

3.4. Interpreting uncertain data

For some derived nouns, informants were very uneasy in deciding whether or not to accept the interpretations. Our policy is as follows: we accepted the interpretation if at least one occurrence was found in a reliable text (see for example section 4.8). We consider to be reliable any official or journalistic text, as we assume such texts are reviewed before publishing. If no occurrence of the

interpretation existed in the corpus, we asked the informant to conceive a short story on his own in which the debated interpretation was supposed to be acceptable. If he could not produce such a story, we considered that the interpretation was not acceptable. If he could, we asked one more informant to explain with which argument the occurrence of the *vn-sha* is co-referring. If the answer differed from the interpretation of the first informant, we rejected the debated interpretation. For some derived nouns, this procedure might not be enough to decide whether or not an interpretation is acceptable. For that reason, when quantifying the observations in section 4, we produced two results. We produced a first *reliable* result by considering uncertain interpretations as impossible. The second, *permissive* result was calculated by accepting the uncertain interpretations. To determine the rules, we took into account both the results and the difference between them. Permissive results are indicated between brackets [].

For some ambiguous derived *vn-shas*, informants had a remarkable reaction. They first chose one of the possible interpretations and rejected or felt uncertain about the other interpretations. But when such other possible interpretations were found in a reliable corpus, they finally acknowledged them. This phenomenon is frequent enough to be taken into account. We assume that this is not due to a lack in competence (all of the informants teach Japanese at university level). We will call the first choice the ‘default interpretation’ and the second one the ‘derogative interpretation’. In the observations, we attempt to determine whether or not a *vn-sha* has a default and derogative interpretation. When there are such interpretations, we try to determine which interpretation is the default or the derogative one.

4. Observations

We group the derived *vn-shas* according to their interpretation and to the system of argument structures of the embedded *vn*. In this study, a category is then defined based on the properties of the system of argument structures, coupled with the interpretation rule of the derived form. For each group, we describe the specificities of both these aspects. ‘Exceptions in a category’ are *vns* which own the system of argument structures of the category, but the corresponding interpretation rule does not fit. We assume that rare exceptions do not call into question the existence of the category. Accordingly, we will not discuss such exceptions.

There are around 9 simple case markers in Japanese, and several complex markers (the list depends on the author). After a first general observation, we concluded that the possible co-referent arguments are the ones marked by the case particles *ga*, *wo*, *ni* and *to*. We then focused on these arguments only. In addition, we also discuss *kara*-arguments for specific reasons that we will explain. As no logical links could be established between the categories, we present those rules in an arbitrary order. We suggest some generalisation hypotheses in section 5.

4.1. Verbal nouns with only one human argument

951 vn-*shas* in the corpus (63% of the total number of vn-*shas* of the corpus) are derived from a vn which has only one human argument. The category includes cases derived from a vn with only one argument (e.g. 15) and cases derived from a vn with more than one argument; but only one human argument (e.g. 16):

- (15) *shibô* { H *ga* }
shibô-sha, 死亡者, litt. 'death - person', 'dead person'

- (16) *yunyû* { H *ga* , X *wo* }
yunyû-sha, 輸入者, litt. 'import - person', 'importer'

In any case, the only human argument is the *ga*-argument. For that reason and due to the high number of vns with only one human argument, the majority of the vn-*shas* found in the corpus are *de facto* co-referent with the *ga*-argument. We cannot exclude that the quantitative superiority of this interpretation might influence the choice of the interpretation of ambiguous vn-*shas*. Indeed, when the listener hesitates between two or more interpretations for an ambiguous vn-*sha*, s/he may prefer the interpretation which is the most common one, namely the *ga*-argument.

4.2. Symmetrical vns with two argument: type *kyôyû* ('to share')

Symmetric verbs (*nitta*, 1974; *Teramura*, 2011) like *kyôyû* ('to share') type vns have at least two argument structures:

- AS1= { H1 *ga* , H2 *to* (, X *wo*) }
 AS2= { HH *ga* (, X *wo*) } (where HH designates a set of at least two humans)

For example:

- (17) 1- AS1= { H1 *ga* , H2 *to* , (X *wo*) }
tanaka san_{H1} ga umeda san_{H2} to ie_X wo kyôyû shiteiru.
 Tanaka Mr. NOM Umeda Mr. p.with house O share
 'Mr.Tanaka share a house with Mr.Umeda'
- 2- AS2 = { HH *ga* (, X *wo*) }
 < *tanaka san to umeda san* >_{HH} *ga ie_X wo kyôyû shiteiru.*
 Tanaka Mr. p.and Umeda Mr. NOM house O share
 'Mr. Tanaka and Mr.Umeda share a house.'

Both argument structures are equivalent, as demonstrated by the example (17). The arguments commute without changing the logical meaning of the statement (Levin, 1993; Borillo, 1971) as follows :

$$\{ H1 \textit{ ga} , H2 \textit{ to} , (X \textit{ wo}) \} \Leftrightarrow \{ H2 \textit{ ga} , H1 \textit{ to} , (X \textit{ wo}) \} \Leftrightarrow \{ H1 \textit{ to} H2 \textit{ ga} , (X \textit{ wo}) \}$$

Within the framework of the system of argument structures, we can define symmetrical *kyôyû* type vns category as follows:

KYÔYÛ

Argument structures

AS1 = { H1 *ga*, (X *wo*), H2 *to* }

AS2 = { H1 *to*, H2 *ga*, (X *wo*) }

Properties

{ H1 *ga*, (X *wo*), H2 *to* }_{AS1} \Leftrightarrow { H1 *to*, H2 *ga*, (X *wo*) }_{AS2}

Observed cases: In the corpus, the category contains 43 (2.86% of the total number of verbal nouns which are derived with *sha* in the corpus) verbal nouns: *kaishoku* ('eating together'), *dôshitsu* ('to be together in the same room'), *kôtai* ('alternate'), etc.

vn-sha is coreferent with the *ga*-argument of AS2. Consequently, in a context in which (17-2) is true, *kyôyû-sha* is co-referent with the two arguments (H1 and H2) of AS1 at the same time. Indeed, from (17-2) we can infer both copulae-sentences CS and CS', and CN":

CS *tanaka san wa , kono ie no kyôyû -sha no hitori dearu.*

Tanaka Mr. TOP, this house GEN be co-owner - person GEN one copula

'Mr. Tanaka is one of the co-owners of this house.'

CS' *umeda san wa , kono ie no kyôyû -sha no hitori dearu.*

Umeda Mr. TOP, this house GEN be co-owner-person GEN one copula

'Mr. Umeda is one of the co-owner of this house.'

In the interpretation rule of *vn-sha*, it is enough to say that *vn-sha* is co-referent with *ga*-argument of AS2. Because of the logical relationship between argument structures, from this statement we can automatically infer that *vn-sha* can be co-referent with both *ga* and *to*-arguments of AS1, *at the same time*.

RULE 3: *vn-shas* derived from a symmetrical verbal noun are co-referent with the *ga*-argument of AS2.

This rule apparently contradicts the interpretation of sentence (18) below in the context (17-1). Indeed, according to rule 3, (18) means that Tanaka met with Mr Umeda *and* Mr Tanaka. This is

clearly wrong because verb *au* ('meet') can not apply when the *ni*-argument is the *ga*-argument himself. In other words, "(usually, no one can meet with him/herself". Such contradiction is not limited to *vn-shas* (see 19 for example). Nevertheless, it is not necessary to modify rule 3. Indeed, we assume that a very general pragmatical rule avoid wrong interpretations. This rule says that 'in a given context, a predicate applies to arguments it can be applied'¹³. For example, in (18), *atta* applies to a subset of people that Tanaka can meet, within the set of *vn-shas*. In the context, the lonely possible subset is the singleton {Umeda}. The same pragmatical rule apply to (19).

(18) Context: 17-1

tanaka san wa kyôyû-sha ni atta

Tanaka Mr TOP co-owner-person DAT met

'Mr. Tanaka met with [a/the] co-owner(s) [except Tanaka himself].'

(19) Context: tanaka is young

tanaka san wa wakai hito ni atta.

tanaka Mr TOP young person DAT met

'Mr. Umeda met [a] young person(s) [except himself, despite he is young].'

4.3. Exclusion of *ni*-argument : *sôdan* ('to conseil') type

Vns like *sôdan* ('consult') are partially similar to symmetrical vns, but the arguments are not always commutative. We think that *sôdan* type nvs correspond to meso-symmetrical verbs (Nitta, 1974). *Sôdan* has the same two argument structures as symmetrical vns.

(20) (AS1)

<uchi no buchô>_{H1} ga <sochi no buchô>_{H2} to sôdan shite, isshoni kimeta.

our GEN manager NOM your GEN man. p.to have a cons., together decid.

'Our manager had a consultation with you manager, and made the decision together.'

(21) (AS2)

<uchi no buchô to sochi no buchô>_{HH} ga sôdan shite, isshoni kimeta .

our GEN man. p.and your GEN man. NOM had a cons., together decid.

'Our manager and your manager had a consultation and made a decision together.'

Both sentences (20) to (21) can be true at the same time. It is not excluded that *sochi no buchô* also asked for counsel. Since they 'made a decision' (*kimeta*) 'together' (*isshoni*), we must

¹³ See other example of application of this pragmatical rule in Renaud (1996) , Blin (2017).

acknowledge the reciprocity in the ‘discussion’ (*sôdan*). Accordingly, in this example, *ga*- and *to*-arguments are commutative, and AS1 and AS2 are equivalent. But this logical relation is not systematic. For example, in (22), reciprocity is impossible: general knowledge and stereotypes determine that a doctor does not consult a wounded patient. It is the wounded patient who consults the doctor:

- (22) *ude no itami ni tsuite, kanja_{H1} ga isha_{H2} to sôdan suru.*
 elbow GEN pain p.about patient NOM doctor p.to consult
 ‘About a pain in the elbow, the patient consults the doctor.’

We can conclude that for *sôdan* type vns, the commutativity is not produced by the meaning of the vn itself. It is produced (or repressed) through the context.

The *sôdan* type vn has one more argument structure: { H *ga*, H *ni* }. The meaning is almost the same as for AS1, except that commutativity is forbidden for any context.

- (23) *uti no buchô ga socchi no buchô ni sôdan shite , kimeta..*
 our man. NOM your manager with have a consultation, decided
 ‘Our manager consulted your manager, and made the decision.’

In this case, we clearly understand that the one who asked for counsel was *uchi no buchô*, not *socchi no buchô*. By adding ‘*isshoni*’ which establishes reciprocity, the sentence would sound slightly unusual. We can summarise the argument structure system as follows: the rule of non-commutativity is such that commutativity is not automatically produced from AS1, nor is it automatically repressed. It then depends on the context.

SÔDAN

Argument structures

AS1= { H1 *ga*, H2 *to* }

AS2= { HH *ga* }

AS3= {H1 *ga*, H2 *ni* }

Rule of non-commutativity for AS3

In any context, if { H1 *ga* , H2 *ni* }_{AS3} is true, then { H2 *ga*, H1 *ni* } is false.

Observed cases: The category contains 24 (1.59%) verbal nouns: *hankô* (‘to revolt’), *hantai* (‘to oppose to somebody’), *kyôryoku* (‘to cooperate’), etc.

We can observe that, if the argument structure AS1 of a given *sôdan*-type verbal noun is implemented in a true sentence, the derived vn-*sha* can refer to the *ga*-argument. Thus, *sôdan-sha* is co-referent with *uchi no buchou* in (22) and with *isha* in sentence (23). If reciprocity is

established through the context, then *vn-sha* refers to both the *ga*-argument and the *to*-argument. In principle, when reciprocity is repressed for whatever reason (context or occurrence of particle *ni* instead of *to*), *vn-sha* refers only to the *ga*-argument. We can summarise the interpretation of *vn-shas* derived from *sôdan* type vns as follows:

RULE 4: When a *vn* is of the *sôdan* type, the *vn-sha* is co-referent with at least the *ga*-argument of the structures AS2 and AS3.

Nevertheless, when informally consulted, the informants *sometimes* accepted that *vn-sha* could be also co-referent with the *ni*-argument. We assume that a broad usage of the particle *ni* exists such that *ni* can be confused with *to*.

4.4. *Kettei* type verbal nouns ('to attribute')

The verbal noun *kettei* has many meanings. In this section, we will look only at *kettei* within the meaning of 'somebody attributes to somebody else a status/function/role'. *Kettei* type vns have an argument structure with human *ga*- and *wo*-arguments, and a *ni*-argument which designates a social function, a role or a status:

- (24) *i'inkai kaichô_{H1} ga tanaka kyôjyu_{H2} wo gakuchô_X ni kettei shita.*
 comity director_{H1} NOM Takana Pr_{H2} O director_X p.*ni* decided
 'The director of the comity decided to put Professor Tanaka as director.'

As the notions of social function, role or status, could be debatable, and to avoid confusion, we will restrict our observations to cases where the syntagm *N-ni* is commutative with the syntagm *< N no { chi'i, yaku, yakuwari}>* (*chi'i* 'position', *yaku* 'role', *yakuwari* 'role') without modifying the meaning of the sentence. For example, (24) and (25) are semantically equivalent:

- (25) *i'inkai kaichô_{H1} ga tanaka kyôjyu_{H2} wo gakuchô_X no chi'i ni kettei shita*
 comity director_{H1} NOM Takana Pr_{H2} O directeur_X GEN post p.*ni* decided
 'The director of the comity nominated Professor Tanaka to the post of director.'

Accordingly, we then define the category of *kettei* type vns as follows:

KETTEI

Argument structure

AS1= { H1 *ga*, H2 *wo*, X *ni* }

Logical constraints

{ H1 *ga*, H2 *wo*, X *ni* }_{AS1} ⇔ { H1 *ga*, H2 *wo*, X *no* Y *ni* } where Y is one of the

following words or any of their hyponyms: *chi'i*, *yaku*, *yakuwari*.

Observed cases: the category contains 21 (1.4%) verbal nouns: *kettei* ('to attribute'), *ninkan* ('to appoint to an office'), *ninmei* ('to appoint'), *sentei* ('to choose'),...

Those vns produce derived forms which, depending on the context, can be co-referent with the *ga*-argument (100%) or the *wo*-argument (83% [72%]), but never with both arguments simultaneously:

- (26) Context: *iryô ga tarinai toki ni, isha wa kanja no naka de, tasukeru hito wo kettei shinai to ikenai.* 'When the medicine is not sufficient, the doctor has to decide who will be saved among the patients.'
- kettei -sha wa, sekinin ga omoi desu.*
 decide-person TOP, responsibility NOM heavy
 'The person who decides has a big responsibility.'

From (24) we infer:

- (27) *kettei -sha wa Tanaka kyôju da.*
 decide-person TOP Takana Pr_{H2} COP
 'The person who has been decided is the Professor Tanaka.'

vn-sha is never co-referent with the *ni*-argument, even if the *ni*-argument is a noun that usually refers to a human. It can be explained by the fact that in this position, any noun N has the non-human meaning < N no {*chi'i*, ...} >.

RULE 5: *Vn-sha* derived from a *kettei* type verbal noun can be co-referent with the *ga*- and the *wo*-argument, but not with both arguments simultaneously.

4.5. Verbal nouns with a 'detrimental' *wo*-argument

Verbal nouns like *rachi* ('kidnap') and *shobatsu* ('punish') refer to events that have a negative effect on the human referred to by the *wo*-argument refers. Indeed, we can safely assume that most people would consider being 'kidnapped' or 'punished' as negative. We will qualify these vns as a 'detrimental *wo*-argument'. Some words cannot be considered to constitute a detrimental *wo*-argument in any context. *Jomei*, for example, which means 'exclude the noun of somebody' (referred to by the *wo*-argument) can refer to an administrative procedure (neutral event) or to

being side-lined (negative impact). We did not include such verbal nouns in this category. We were unable to build a satisfying definition using only syntactic criteria. We suggest two definitions.

wo-detrimental verbal nouns

Argument structure

AS= { H1 *ga*, H2 *wo* }

Intensional definition:

From a implementation 'H1 *ga* H2 *wo* VN *shita*' it can be inferred *in any context* that:

H2 *ni wa, son ni naru*
H2 p.*ni* TOP, loss p.*ni* become

'To H2, it is a loss.'

Extensional definition:

The VN is hyponymous or synonymous to:

ryôjôku (陵辱, 'to rape, to humiliate')

yokuryû (抑留者, 'to detain')

yûhei (幽閉, 'to confine, to incarcerate')

...

ryakushu (略取, 'to capture')

satsuriku (殺戮, 'to massacre')

Observed cases: The category contains 56 (3.72%) vns.

We recommend using the extensional definition as the intensional one is subjective and difficult to make operational. The second definition can be simplified if there are logical inferences between the vns of the list. If such is the case, the list can be limited to the most general vns. Hyponyms can be deleted.

For many of those vns, the informants did not agree about the co-referent argument. Co-reference with the *ga*-argument is always possible. Co-reference with the *wo*-argument is possible in many (but not all) cases. The choice depends on the context (see *rachi*, 'to kidnap' in (4) and the related discussion). Nevertheless, we can observe that the more often the passive form is used, the more ambiguous the *vn-sha* become. To obtain this result, we measured in a large corpus (10 million sentences) the frequencies of the active and passive form of each vn. We divided the whole set of vns into two subsets depending on whether or not the frequency of the passive form exceeds 45%. This percentage was determined based on manual tests. We first counted in each category the number of derived *vn-sha* that can be co-referent with both the *ga*- or *wo*-interpretation, leaving no doubt (i.e. informants were sure). We obtained a second result by including the vns of which the informants were unsure. The first result, regarding which informants were certain, showed a statistically significant correlation between the frequency of the passive form of the vn and the ambiguity of the derived *vn-sha*. The second, permissive result, which included uncertainty, demonstrated no statistically significant result.

frequency of the passive form / occurrence of the embedded vn	number of vn-shas	Interpretation
>45%	21	ambiguous
	2	non ambiguous
<45%	12	ambiguous
	20	non ambiguous

Table 2: Ambiguity of *vn-sha* depending on the frequency of the passive form of the embedded vn in the corpus (for example: for 21+2 vns, more than 45% of the occurrences are in the passive form; the derived noun of 21 of these vns are ambiguous). Reliable results.

(Table 2) In the '<45%' category, there are 12 ambiguous forms which can be co-referent with the *wo* argument. This number increases to 20 for the uncertain results. This figure is quite significant, and we cannot exclude the *wo*-interpretation. Accordingly, we suggest that the *wo*-argument is a derogative value in this category:

RULE 6: For any *wo*-detrimental verbal noun, if the frequency of the passive form exceeds 45% for all occurrences of the verbal noun, then the derived form can be co-referent with the *ga* or *wo*-argument. Both interpretations are not possible at the same time. For other *wo*-detrimental *vn-shas*, the interpretation is the *ga*-argument by default. The *wo*-argument is a derogative value.

This correlation is unique. We did the same test for all the *vn-shas* derived from a verbal noun with more than one human argument, but no such correlation occurred. At this stage of the observations, it is not possible to establish a reliable causal relation between the ambiguity and the frequency of the passive form (see discussion in section 5, however).

4.6. Verbal-nouns with 'beneficial' *wo*-argument

Contrary to the *wo*-detrimental vns, some vns designate an event with a positive effect on the *wo*-argument, such as *kaihô* ('liberate'). We will refer to such vns as '*wo*-beneficial'. The difficulties to define this category are similar to the difficulties to define the *wo*-detrimental vns category. We use an extensional definition:

***wo*-beneficial verbal nouns**

Argument structure

AS= { H1 *ga*, H2 *wo* }

Extensional definition:

A verbal noun with an argument structure of the type { H1 *ga*, H2 *wo* } will be called *wo*-beneficial if it is hyponymous or synonymous to:

annai ('to guide')
kyûshutsu ('to help')

...

Observed cases: The category contains 18 (1.2%) vns.

77% [88%] of the *vn-shas* in the category can be co-referent with the *wo*-argument, whereas 100% can be co-referent with the *ga*-argument. As there is a slight quantitative difference between the *wo*- and *ga*-arguments, we can assume that the *ga*-argument is the default interpretation, with the *wo*-argument being the derogative one.

RULE 7: Derived noun from 'wo-beneficial' vns can be co-referent with the *ga*-argument (default interpretation) and the *wo*-argument (derogative interpretation).

4.7. Verbal nouns of type *jita*

We classify vns as *jita* verbal nouns, such as *tôroku* ('enrol'), when they have two argument structures: a transitive (29) and an intransitive structure (30). The intransitive structure can be inferred from the transitive one:

(28) *sensei ga gakusei wo tôroku shita.*
 teacher NOM student O enrolled
 'The teacher enrolled the student.'

⇒ (29) *gakusei ga tôroku shita.*
 student NOM enrol
 'The student enrolled.'

It should be noted that the acceptability of the argument structures and the inferences are debated by informants for some vns like *idô* ('displacement'). The transitive argument structure with a human *wo*-argument is contested by some informants who prefer to use the factitive form. For example, instead of the sentence (31), they prefer the sentence (32) with the factitive form.

(30) *sensei ga gakusei wo idô shita.*
 professor NOM student O move
 'The professor moved the students.'

(31) *sensei ga seito wo idô sasetu.*
 professor NOM student O move factitive

'The professor made the student move.'

For these informants, it is difficult to discuss the acceptability of the inference rule between the transitive form (which they do not accept) and the intransitive form. Nevertheless, we have decided to accept the transitive form as it has been confirmed for many vns, and since the occurrences are included in the corpus and registered in some lexical databases (like the transitive form of *idô* ('move') in the Goitaikei). When accepting the transitive form, informants also accept the inference rule. We then have to take this rule into account, although the use of such a transitive form is (still) rare or limited to some specific jargons. We thus define JITA verbal nouns as follow:

JITA

Argument structure:

AS1= { H1 *ga* , H2 *wo* }

AS2= { H1 *ga* }

Properties

{ H1 *ga* , H2 *wo* }_{AS1} ⇒ { H2 *ga* }_{AS2}

Observed cases: the category contains 8 (0.53%) vns: *kôtai* ('change'), *shujutsu* ('to operate'), *tôroku* ('to registrate'), ...

The derived *vn-shas* can refer to both *ga*- and *wo*-arguments. Let's consider a context where assessment (28) and then (29), are true. In such a case, both of the following copulative sentences are acceptable, but *not at the same time*:

(32) <*tôroku - sha*> wa *sensei da*.
 enrol - person TOP teacher COP
 'The person who enrolled is the teacher.'

(33) <*tôroku - sha*> wa *gakusei da*.
 enrol - person TOP student COP
 'The person who enrolled is the student.'

It is not clear whether there is a default or derogative interpretation. Informants preferred the *wo*-argument of the transitive argument structure). But, for *shujutsu-sha* ('to operate-person'), they preferred the *ga*-interpretation, *i.e.* the 'person who operates' ('surgeon'), despite there being a specific word to designate this person (*shittô-sha*). Accordingly, we can conclude that the derived form is ambiguous. We will not specify here which interpretation is preferred in an underspecified context.

RULE 8: Derived nouns from *jita* vns can be co-referent with the *ga*-argument and the *wo*-

argument.

4.8. Limited tolerance for the *ni*-argument when referring to the recipient: *sôshin* ('to send') type verbal nouns

We classify as *sôshin* ('send') type the vns which designate the event in which somebody (referred to by the *ga*-argument) sends/gives/etc. something (*wo*-argument) to somebody (*ni*-argument). The *ni*-argument refers to the recipient, in the broadest sense:

- (34) *tanaka sensei ga tarô ni mêru wo sôshin shita*
 Tanaka prof. NOM Tarô DAT mail O sent
 "Professor Tanaka sent a mail to Tarô.'

The argument structure { H *ga*, H *wo*, H *ni* } is not specific to this type of vns and it is not sufficient to characterize *sôshin* type vns. As there are no other syntactic criteria to identify *sôshin* type vns, we also use semantic (logical) criteria: it is possible to obtain the following inference between a sentence built with a *sôshin* type vn and the copulae-sentence including the verb *uketotta* ('receive') or any hyponym of it:

- H1 *ga* X *wo* H2 *ni* VN *shita* ⇒ H2 *ga* X *wo* *uketotta hito dearu*.
 H2 NOM X O received person COP
 "H2 is the person who received X.'

Thus, the *sôshin* type verbal nouns are defined as follows:

SÔSHIN

Argument structure

AS= { H1 *ga* , X *wo*, H2 *ni* }

Inference rules

{ H1 *ga* , X *wo* , H2 *ni* } ⇒ "H2 *ga* X *wo* *uketotta hito da*.'

Observed cases: The category contains 106 (7.04%) verbal nouns: *kenkin* ('to financially contribute'), *bunpai* ('to distribute'), *zôyo* ('to donate'),...

For any *sôshin* type vn, the derived vn-*sha* can be co-referent with the *ga*-argument. The *ni*-argument is more controversial. In many cases, informants were uncomfortable to clearly exclude the co-reference with the *ni*-argument, but were unable to suggest a context wherein this interpretation could be natural. We found very few occurrences of such an interpretation in reliable official documents or commercial reports. Despite their surprise, the informants accepted these cases and the interpretation:

- (35) ...*rikkôho yotei -sha setsumei -kai shusseki-sha oyobi*
 candidate prevision - person explanation-meeting attendee and
*shiryô haihu -sha ichiran*¹⁴
 document distribute-person summary
 "summary of attendees to the briefing of candidates to the election, and of people
who received documentation'
- (36) (*hyô-9 zaitaku kôrei-sha sâbisu no jisseki to suiikei*)¹⁵
 "Table-9 estimation and real service for elderly at home'
bentô haitatsu -sha -sû =2,582
 lunch box distribution-person-number = 2,582
 "Number of persons to whom lunch box has been delivered = 2,582.'

In the two examples above, *vn-sha* is co-referent with the *ni*-argument. After seeing those examples, informants were more likely to accept many similar cases which they had initially rejected. As a result, while taking into account reliable and permissive interpretations, *sôshin* type *vn-shas* which can be co-referent with the *ni*-argument increased to 50%. We can assume that information from informants is too confusing. We then conclude as follows: since the *ni*-interpretation is confirmed in reliable corpora and for different *vn-shas*, we can assume that this interpretation is possible and not erroneous. However, as such a reliable interpretation applies to fewer than half of the total number of derived forms of this category, we must distinguish between this interpretation and the *ga*-interpretation, which is by far the most frequent one. We can then suggest the following rule:

RULE 9: When the verbal noun is of *sôshin* type, the *ga*-argument is the default interpretation, with the *ni*-argument being the derogative one.

4.9. *Chûmon* type

Chûmon type *vns* have human *ga*- and *ni*-arguments, and a *to*- or *wo*-argument. The *to*-argument is a citation:

14 original sentence: さいたま市議会議員補欠選挙（見沼区）の立候補予定者 - 説明会の資料配布者について - 記者発表資料 - 平成25年4月18日（木）報道機関

15 元気はつらつ高齢者計画の概要(案) (平成21年度～23年度) 紫波町 生活部 長寿健康課
http://www2143uc.sakura.ne.jp/cms/files/01139/koureisya_kossi.pdf

- (37) AS1= { H1 *ga*, H2 *ni*, X *to* }
bôî san ni "kôhî motte koi ' to, gaki ga chûmon shita.
 waiter DAT "coffee bring come' p.to kid NOM commanded.
 " 'Bring me a coffee' the kid commanded to the waiter.'

A *wo*-argument instead of the *to*-argument is possible, but not with a citation:

- (38) AS2= { H1 *ga*, H2 *ni*, X *wo* }
gaki_{H1} ga bôî san ni kôhî_X wo chûmon shita.
 kid NOM waiter DAT coffee O commanded.
 "The kid commanded a coffee to the waiter.'

In a lot of contexts, there exists a nominal phrase NP such that { H1 *ga*, H2 *ni*, X *to* }_{AS1} ⇒ { H1 *ga*, H2 *ni*, NP *wo* }_{AS2}. This is illustrated by the inference (37) ⇒ (38) where the NP is *kôhî*. Nevertheless, we can assume that the argument structure system is sufficient to characterise the *chûmon* category:

CHÛMON

Argument structure

AS1= { H1 *ga*, X *to*, H2 *ni* }

AS2= { H1 *ga*, X *wo*, H2 *ni* }

Observed cases: 50 (3.32%) verbal nouns: *shazai* ("apologize'), *jogen* ("to counsel.'), *senden* ("to advertise'), ...

Both reliable and permissive results show that only a very small portion (respectively 4% and 14%) of *vn-shas* can be co-referent with the *ni*-argument. Accordingly, we can assume that the *ni*-interpretation is impossible.

RULE 10: *vn-sha* derived from *chûmon* type verbal nouns are coreferent with the *ga*-argument only.

4.10. When the *ni*-argument designates the beneficiary

Among the *vns* observed, some of them accept a *complement* marked by *ni* or *no tameni*. This complement refers to the beneficiary of the event. Contrary to the *vns* discussed in the previous section, *ni* and *notameni* are semantically equivalent. Such is the case with *insatsu* ("print"):

- (39) *(boku_{H1} wa) sensei_{H2} ni shorui_x wo insatsu shita*
 me_{H1} TOP prof._{H2} DAT documents O printed
 "I printed a document for the professor.'
 ⇒ *(boku_{H1} wa) sensei_{H2} no tameni shorui wo insatsu shita*
 me_{H1} TOP) prof._{H2} p.for doc. O printed
 "I printed a document for the professor.'

Such a sentence is not common, but common enough to be taken into account. It is also understandable. The derived *vn-sha* cannot be co-referent with the *ni/notameni*-complement. We assume this is because the complement is a *circumstantial* one, not an argument. Indeed, there is not necessarily a beneficiary for these events.

4.11. *kara* argument, type *dokuritsu* ('to become independent')

Some vns have a *kara*-argument. Such is the case for *dokuritsu*.

- (40) *kodomo ga haha kara dokuritsu suru.*¹⁶
 child NOM mother p.from be independent
 "The child becomes independent of her mother..'

The category is defined as follow:

DOKURITSU

Argument structure

AS1= { H1 *ga* , H2 *kara* }

Observed cases: The category contains 4 (0.27%) verbal nouns: *koritsu* ('be isolated'), *dokuritsu* ('be independant'), ...

The derived *vn-sha* cannot be co-referent with the *to/kara*-argument. This leads to the following rule:

RULE 11: *vn-sha* derived from a *dokuritsu* type *vn* is coreferent with the *ga*-argument.

4.12. *jushin* ('to receive') type

As we have seen, *vn-shas* derived from *sôshin* ('send') type vns are ambiguous. This makes us wonder whether antonym vns like *jushin* ('receive') have similar properties. *Jushin* has a human *kara*-complement:

¹⁶ repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/160437%2F1/apk00100_071.pdf

(41) *tarô ga jirô kara mêru wo jushin shita.*

Tarô NOM Jirô p.kara mail O received

'Tarô has received a mail from Jirô.'

It is difficult to determine whether the *kara*-complement is just a circumstantial complement or if it possesses argument status. The syntactic properties are not sufficient to define the category. We added a logical constraint on the *ga*-argument, as follows:

JUSHIN

Argument structure

AS= { H1 *ga* , (X_{role} *wo*) , H2 *kara* }

Inference rules

{ H1 *ga* X *wo* H2 *kara* } ⇒ "H1 *ga* ukegawa da.'

H1 NOM recipient COP

"H1 is the recipient'

Observed cases: 10 (0.66%) verbal nouns: *jumei* ('to be commissioned'), *jueki* 'to beneficate', *jukyû* ('to receive a pension'), ...

The *vn-sha* cannot be co-referent with the *kara*-argument. The rule of interpretation for the derived *vn-sha* is then as follows:

RULE 12: A *vn-sha* derived from *dokuritsu* type verbal noun is coreferent with the *ga*-argument.

From the rules 11 and 12, we assume a more general rule:

RULE 13: VN-*sha* is never coreferent with *kara* complement.

4.13. Valuation verbal nouns: *zôka* ('to grow') type

When composing a verb with the support verb *suru*, *zôka* and *genshō* belong to a category of verbs we will call 'valuation verbs'. Valuation vns and verbs describe the evolution of the value of a measure (referred to by the *ga*-argument) with the passing of time. The *ni*-argument designates the final numerical value (*ichiman* 'ten thousand' in 42). The \emptyset -argument designates the quantitative difference between original value and final value (e.g. 43) (see also examples 45 and 46 in Blin 2013b):

(42) *kono 10 nen no aida ni,*

these 10 years GEN during ,
homuresu no kazu ga 10000 ni zôka shita
 homeless GEN number NOM 10000 to grow
 "In the past ten years, the number of homeless has grown to 10 thousand.'

- (43) *kono 10 nen no aida ni,*
 these 10 years GEN during ,
homuresu no kazu ga 10000 ∅ zôka shita
 homeless GEN number NOM 10000 grow
 "In the past ten years, the number of homeless has grown by 10 thousand.'

When the measure noun is made 'evident' through the context, it can be omitted (see example 44). In other word, to interpret this sentence, when the *ga*-argument is not a measure noun, it is necessary to "add' the most plausible measure noun according to the context. Therefore, we will add *kazu* in (41) which will be synonymous with (42):

- (44) ... , *homuresu ga 10000 ni zôka shita*
 ..., homeless NOM 10000 to grow up
 "... homeless has grown to 10 thousand.'

Valuation verbal noun can be defined as follows. This definition maybe should be extended by metonymy to human *ga*-argument but the metonymy is difficult to modelize. At this stage of the observations, we prefer to not consider the human *ga*-complement as an argument of the valuation verbal nouns.

Valuation verbal nouns

Argument structure

AS1= { measure_noun *ga* , numeral_group *ni* }

AS2= { measure_noun *ga* , numeral_group ∅ }

Observed cases: 3 (0.20%) verbal nouns: *zôka* ('to grow up ; increase'), *genshō* ('to decrease'). There are also *kyûzō* ('fast growth'), *kyûgeki* ('fast decrease') etc.

When the *ga*-argument designates a number of humans, then *vn-sha* designates the set of humans who have been added/removed. For example, if (44) is true, then the *vn-sha zôka-sha* refers to the 10,000 individuals who joined the initial set of homeless people. It can be expressed explicitly like as in sentence (45):

- (45) ... *zōka-sha no kazu ga 10,000 nin datta.*
 grow-person GEN number NOM 10,000 pers. COP
 "The number of new [homeless] is 10,000 people.'

5. Discussion and conclusion

In this paper, we have provided a set of rules to predict the *possible* interpretations of the derived noun *vn-sha*, depending on the properties of the system of argument structures of the embedded *vn*. The initial idea was to establish the rules by using only observable surfastic properties, such that these rules can be used by numerous researchers, whatever the theoretic framework. A second idea at this stage of the research was to provide a most efficient system of rules rather than a 'universal' explanation. Accordingly, we tried to describe the highest number of observed cases with the lowest number of rules possible. The system operates on a scale between two extremes: a system which contains as many rules as forms (to each form is applied a specific rule), and an ideal system where a single rule predicts the meaning of all the forms. The third and last idea was that exceptions do not automatically invalidate rules. Our system of rules is such that it can be evaluated using quantitative criteria: number of rules, number of observed cases to which these rules apply, and number of exceptions. This quantitative evaluation is an interesting way of objectively comparing the efficiency of different systems used to process the same data.

We finally obtained 13 categories of *vns* to which correspond 12 interpretation rules (a general rule covers the rules 11 and 12) for the derived *vn-sha*. Two categories (*wo*-detrimental and *wo*-beneficial) are partially based on extensional definitions. This system of rules covers 80% of the forms observed in the corpus. We must emphasise here that, at this stage of the study, we cannot say if any rule (and what kind of rule) applies to the remaining 20% of the forms.

In conclusion, we will now discuss the hypothesis of generalisation. A first general observation is that that *vn-sha* can only be co-referent with the human *ga*-, *wo*-, *to*- and *ni*-complements. Furthermore, only argumental complements are possible. The system of rules is such that it can be applied without making this distinction between arguments and adjuncts. It can then easily be used by linguists who prefer avoiding the distinction (general discussion in Manning 2003). All other complements are excluded. Except for the idiomatic form *gōshisha* ('enshrined person'), co-reference is always possible with the *ga*-argument, whatever its semantic value. We can see that for some *vns* (*kyōyū*-, *sōdan*-, *jita*-type), the co-reference with other arguments (*to* and *wo*) than *ga* is deductible from the logical properties of the system of arguments. For example, for symmetrical *vns*, co-reference with *to*-argument in argument structure AS1 is deductible from co-reference with *ga*-argument in argument structure AS2.

Of course, applying 12 rules or less to interpret a single morpheme is not satisfactory. The number should ideally be reduced. Let's speculate on a possible generalisation, while respecting

the constraint of processing only observable data. The *wo*-detrimental vns reveal interesting findings. We observed that *vn-shas* derived from *wo*-detrimental vns are ambiguous if the embedded vn mainly occurs in passive form. Suppose that passive form is a surfastic way of overshadowing one argument (indeed, the *ga*-argument) and then creating some defocusing effect (see for example the discussion in Shibatani 1990). On that basis, there may be a relationship between the point of focus and the choice of default *vn-sha* interpretation. Unfortunately, the correlation between the passive form and its (possible) focus effect is only true for one category: *wo*-detrimental vns. This is not enough to establish a general rule. Moreover, it could apply only to vns with the *wo*-argument (and not with the *ni*- or *to*-argument).

We then wondered if the choice of the default interpretation could have some relationship with a related notion of focus, the informativeness. Let's observe the detrimental vn *taiho* ('arrest'). By default, *taihosh*a refers to the *wo* argument. Some informants suggested that could be because the *ga*-argument is 'trivial and not interesting to point to'. We can see that the noun which represents the *ga*-argument is much more predictable than the *wo*-argument. It mainly designates people who can arrest other people, such as the police. Arrested people represent a greater variety. Based on this rough observation, we can suggest the following hypothesis: the ambiguous *vn-sha* is more frequently co-referent with the less predictable argument. In another words, to know the default interpretation, we have to know the predictability of the argument. A naive way to measure the predictability is to count the number of different nouns compatible with the position of *ga*-argument. To this end, we conducted a brief experiment: in a small but varied corpus (chiebukuro, 2009), we manually counted the different nouns which were explicitly given (no ellipsis) and which occupied the *ga*-position and the *wo*-position. For vn *taiho*, we observed that the *wo*-argument is much more explicit and varied¹⁷. Based on our speculative hypothesis, we can assert that the *wo*-argument is less predictable and is therefore the default value. This is consistent with our observations. Predictability can be measured for any argument, including *ni*-, *to*- and *wo*-complements. The hypothesis can then be applied to any complement. As the reader will no doubt understand, (in)validating the hypothesis of a relationship between the predictability and interpretation of *vn-sha* would be a major process and cannot be carried out here. It would be the next step of the research.

In this study, we have tried to innovate by quantifying the scope of the system of rules. We believe that we can calculate the possible meaning of 1,305 derived nouns among 1,503 represented in the corpus. For the remaining 198 cases, we have, on the one hand, the system of arguments of the embedded vn and, on the other hand, the list of possible interpretations of the derived <*vn-sha*> from this vn. But we could not find any productive rule to infer former data from latter data.

In a future phase of the study, we shall resolve these cases. While we admit that many of them

fit into the categories established in the current study, they are disqualified by certain external factors. For example, a given verbal noun can have two human arguments, but for one of them, there exists a useful term. In such a case, although the *vn-sha* can be co-referent with this argument, anyone would prefer the specific term and the *vn-sha* will appear impossible. While analysing the data in the current study, we have pointed out such particular phenomena, which could explain why some of the remaining cases seemed incompatible with the existing category. In the future study, it will be necessary to list these phenomena.

6. Bibliography

- Bekki, D. (2010). *Nihongo bunpo no keishikinron: katsuyotaikei, togokoza, imigosei [A formal theory of Japanese grammar: the conjugation system, syntactic structures, and semantic composition]*. Kuroshio.
- Blin, R. (2006). Formalisation des contraintes d'association des groupes numériques avec spécifiques et noms communs en japonais. *Cahiers de linguistique - Asie orientale*, 35(2), 241–275. <https://doi.org/10.3406/clao.2006.1758>
- Blin, R. (2013). sokuryou meisi no imi - tougo teki tokutyou ni kan suru kousatu [Measure Nouns in Japanese : Their Semantic and Syntactic Characteristics]. *Lexicon Forum*, (6), 173-202.
- Blin, R. (2015). Metadonnées du lexique-grammaire du japonais jalexGram-0.010 [metadata of Japanese lexicon-grammar jalexGram-0.010]. Consulté à l'adresse <http://goo.gl/3BvzGU>
- Blin, R. (2017). Désambiguïisation des noms propres à l'aide des connaissances sur autrui. *Linguisticae Investigationes*, 42(2).
- Booij, G. E. (1986). Form and meaning in morphology: the case of Dutch « agent nouns ». *Linguistics*, 24, 503-517. <https://doi.org/https://doi.org/10.1515/ling.1986.24.3.503>
- Borillo, A. (1971). Remarques sur les verbes symétriques français. *Langue française*, 11(1), 17-31.
- Danjou-Flaux, N., & Fichez-Vallez, E. (1985). Linguistique taxonomique et grammaire générative : le traitement de l'exception. *Langue française*, 66(1), 99–116. <https://doi.org/10.3406/lfr.1985.6348>
- Danlos, L. (2000). Discours causal et rôles thématiques. In *Des structures linguistiques à leur interprétation; mélanges en l'honneur d'André Borillo* (M.Plenat). Rodopi.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3), 547-619. <https://doi.org/10.2307/415037>
- Dubois, J., & Dubois-Charlier, F. (1999). *La Dérivation suffixale en français*. Paris: Nathan.
- Editor group of « Daijisen », & Matsumura, A. (1998). *Daijisen* (New edition). Shôgakkan.
- Fradin, B. (2005). On a semantically grounded difference between derivation and compounding. In *Morphology and its Demarcations* (John Benjamins, p. 161-182). Amsterdam, Philadelphia: Dressler W. U., D. Kastovsky, O. E. Pfeiffer & F. Rainer.
- Iida, A. (2003). *Kazoekata no jiten [Dictionary of counting in Japanese]*. Shôgakkan.
- Ikehara, S., Miyazaki, M., Shirai, S., Yokoo, A., Nakaiwa, H., Ogura, K., ... Hayashi, Y. (1997). *Goi-Taikei - A Japanese Lexicon* (Iwanami Shoten). Tokyo.
- Inoue, K. (1976). *Henkei bunpou to nihongo, T1, Tougo Kouzo wo chuushin ni [Transformational grammar and Japanese; T1; Syntactic structure]* (Taishukan).
- Kageyama, T. (1999). *Keitairon to imi [Morphology and Meaning]* (Kuroshio Shuppan).

- Kageyama, T. (2002). Dôshasyu meishi ni okeru goi to tôgo no kyôkai [Frontier between lexicon and syntax in agentive nouns]. *Kokugogaku*, 53(1), 44-55.
- Kawahara, D., & Kurohashi, S. (2010). Acquiring reliable predicate-argument structures from raw corpora for case frame compilation. In *Acquiring Reliable Predicate-argument Structures from Raw Corpora for Case Frame Compilation* (p. 1389-1393). Valetta, Malta.
- Kobayashi, H. (2004). *Gendai nihongo no kango doumeishi no kenkyû (Study of contemporary Japanese Sino-Japanese verbal nouns)* (Hitsuji Shobô). Tokyo, Japan.
- Levin, B. (1993). *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago Press.
- Levin, B., & Rappaport, M. (1988). Nonevent -er nominals: a probe into argument structure. *Linguistics*, 26, 1067-1083. <https://doi.org/https://doi.org/10.1515/ling.1988.26.6.1067>
- Lieber, R. (2004). *Morphology and lexical semantics*. Cambridge ; New York: Cambridge University Press.
- Manning, C. D. (2003). Probabilistic syntax. In *Probabilistic Linguistics* (The MIT Press, p. 289-341). Rens Bod, Jennifer Hay and Stefanie Jannedy.
- Masuoka, T., & Takubo, Y. (1992). *Kiso nihongo bunpoo [The basic grammar of Japanese]*. Tokyo: Kuroshio.
- Matsumura, A. (2006). *Daijirin*. Tokyo: Sanseido.
- Matsumura, A., & Sanseido Editor. (2006). *Daijirin* (3^e éd.). Sanseido.
- Miyajima, T. (1997). Hito meishi no imi to asupekuto-tensu [Aspect-tense and meaning of nouns of persons]. In *nihongo bunpou - taikai to hôhō*. Kawabata Yoshiaki, Nitta Yoshio.
- Miyake, T. (2010). « Ichiji kango suru » kei doushi wo megutte. In *Kango no gengo gaku* (Blin Raoul, Nakajima Akiko, Oshima Hiroko, p. 107-119). Tokyo: Kuroshio.
- Nitta, Y. (1974). taisyoudousi (Symmetrical Verb) と hantaisyoudousi (Meso-symmetrical Verb) と hitaisyoudousi (Anti-symmetrical Verb) – kakuseibun keisei kisoku no tameni. *kokugogaku kenkyuu*, (13), 56-68.
- Ohara, M. (2000). *An analysis of verbal nouns in Japanese*. University of Essex.
- Ono, N. (2016). Agent nominals. In *Handbook of Japanese Lexicon and Word Formation* (Mouton de Gruyter, p. 599-630). Kageyama Tarou.
- Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, 17(4), 409-441. <https://doi.org/10.1.1.14.5868>
- Renaud, F. (1996). The definite article: code and context. *Journal of Semantics*, 13(2), 139-180. <https://doi.org/10.1093/jos/13.2.139>
- Roy, I., & Soare, E. (2014). Les noms d'humains dérivés de participes : nominalisations en -ant et -é/i/u. *SHS Web of Conferences*, 8, 3197-3208. <https://doi.org/10.1051/shsconf/20140801352>
- Shibatani, M. (1990). *The languages of Japan* (Cambridge University Press).
- Sugimura, H. (1986). -sha, -ka. *Nihongogaku*, 5(3).
- Sugioka, Y., & Ito, T. (2016). Derivational affixation in the lexicon and syntax. In *Handbook of Japanese Lexicon and Word Formation* (Kageyama Taro, Kishimoto Hideki, p. 347-386). Mouton de Gruyter.
- Takeuchi, K., Inui, K., Takeuchi, N., & Fujita, A. (2008). Subclassification of verbal argument structure based on the inclusion of meaning. In *Annual meeting of NLP*. Tokyo: Tokyo University.
- Teramura, H. (2011). *Nihongo no shintakusu to imi [Syntax and Meaning of Japanese]* (Vol. I).

Tôkyô: Kuroshio shuppan.

Tsujimura, N. (2013). *An Introduction to Japanese Linguistics*. Wiley.

Yatabe, S. (1999). Particle ellipsis and focus projection in Japanese. *Language, information, text*, (6), 79-104.

7. Summary of results

Total of observed cases in the corpus	1506	
CLASSIFIED CASES	1307	86,79%
Verbal nouns with only one human argument	951	63,15%
Symmetric verbs with to argument: type kyôyû	43	2,86%
Exclusion of ni-argument : sôdan ('to conseil') type	24	1,59%
Kettei type verbal nouns ('to attribute')	21	1,39%
VNs with a 'detrimental' wo-argument	56	3,72%
VNs with 'beneficial' wo-argument	18	1,20%
VNs of type jita	8	0,53%
VNs of type sôshin (to send)	106	7,04%
VNs of type CHUUMON	50	3,32%
VNs of type DOKURITSU ('to become independant')	4	0,27%
VNs of type JUSHIN ('to receive')	10	0,66%
Valuation verbal nouns	3	0,20%
Idioms	13	0,86%

UNCLASSIFIED CASES	199	13,21%

% of human X-argument which can be co-referent with the derived form

	Certain	Uncertain cases (points)	Certain + Uncertain
ga-argument	99,7%	0,00	99,7%
wo-argument	63,9%	7,43	71,3%
ni-argument	23,6%	15,08	38,7%
to-argument (symetric vn)	94,1%	—	—