Zeitschrift für Wortbildung
Journal of Word Formation

Nominalization

Hrsg. von / Edited by

Gianina Iordăchioaia
Elena Soare

Petra M. Vogel • Elke Donalies • Ludwig M. Eichinger • Mechthild Habermann • Jörg Meibauer • Barbara Schlücker • Hans-Joachim Solms • Pavol Štekauer • Salvador Valera Hernández (eds.)

PETER LANG

Inhalt / Contents

Varia / Miscellaneous

Vorwort / Preface

Gianina Iordăchioaia & Elena Soare

Current trends in the study of nominalization

Nominalization has been at the forefront of linguistic research since the early days of generative grammar (Lees 1960, Vendler 1968, Lakoff 1970). The theoretical debate as to how a theory of grammar should be envisaged in order to capture the morphosyntactic and semantic complexity of nominalization, initiated by Chomsky's (1970) *Remarks on nominalization*, is just as lively today, after five decades during which both the empirical scope and the methodology of linguistic research have seen enormous progress. We are delighted to be able to mark this occasion through our collection, next to the anniversary volume *Nominalization: 50 Years on from Chomsky's* Remarks, edited by Artemis Alexiadou and Hagit Borer, soon to appear with Oxford University Press.

This collection represents a selection of the papers presented at the 8th JENom *Workshop on Nominalizations/Journées d'Etude sur les NOMinalisations*, organized at the University of Stuttgart in 2019 and aims to offer insights into the diversity of theoretical and methodological trends that the study of nominalization is currently following. It gathers work on several languages that ranges from addressing new empirical aspects of nominalization to contributing a better understanding of previously known formations, which are now addressed from a different theoretical and/or methodological perspective.

The empirical domain of this collection covers little-documented languages such as Ktunaxa (or Kutenai, an indigenous language isolate spoken in Canada and the USA) and Esahie (a Central Tano language spoken in Ghana), besides more widely-studied languages such as Irish, Italian, Japanese, French, and English. Different types of nominalizations are considered: from syntactically complex patterns (Ogawa, Niikuni & Wada and Gatchalian) to standard suffix-based nominalizations (Varvara and Wauquier, Hathout & Fabre), deverbal compounds (Knittel & Villoing), and zero-derived/conversion nouns (Iordăchioaia, Schweitzer, Svyryda & Buitrago Cabrera, Bloch-Trojnar, and Tribout), or a mixture of segmental and suprasegmental morphology (Broohm & Melloni).

From a theoretical and methodological perspective, the collection is just as broad: It offers synchronic, but also diachronic insights, syntax-oriented and lexicalist approaches, as well as corpus-based and experimental studies,

which contribute new empirical resources or statistical modeling of corpus data, as detailed in the brief summaries below.

Ogawa, Niikuni & Wada combine a diachronic and an experimental approach to a nominalization via juxtaposition strategy in Japanese. As they argue, coordination and juxtaposition of verbs or adjectives can form an NP if the conjuncts are antonymous to each other. This "Nominalization of Antonymous Combination" Construction (NACC), entirely new to the study of Japanese nominalizations, is identifiable by the alternation between Nominative and Genitive subjects and is only visible when considering diachronic data. The authors analyze this construction as a nominalization *via* an empty nominalizer, that is, similar to a zero-derivation, and show that it has been gradually developing in the history of Japanese. On the basis of a survey of the Corpus of Historical Japanese, they show how the various types of the NACC have developed diachronically, and an explanation is proposed along the lines of Ogawa's (2014) hypothesis of "syntactic constructionalization", which makes the right predictions about the differences in acceptability judgments among speakers of different ages with respect to the various types of NACC. These ratings are collected via an acceptability-rating experiment with 400 participants, which confirms the relevance of age in accepting the construction.

Broohm & Melloni investigate the role of tone in action nominalizations in Akan, Gã, Lɛtɛ, and Esahie (i.e. Kwa languages spoken in Ghana). Building on previous studies on word formation in Akan, Gã, and Lɛtɛ, in which tone is shown to accompany various morphological strategies such as suffixation, compounding, and reduplication, Broohm & Melloni argue that tone has a morphological function in some nominalization patterns (especially in Akan and Lɛtɛ), where it is the only indicator of the verb-to-noun transposition, in the absence of other morphology. In Esahie, in particular, they show that a (floating) high tone, in association with the suffix, plays a morphemic role in deverbal nominalization. They conclude that true conversion is not available in action nominals in Kwa languages, since in the absence of overt suffixes, some tonal change will be present to mark the nominalization. Unusual as tonal morphology may look from the perspective of European languages, the presence of suprasegmental morphology reminds us of stress shift in deverbal zero-derived nominals in English, as the authors also remark (cf. the contribution by Iordăchioaia et al. on English).

Gatchalian offers a theoretical study of nominalizations in Ktunaxa, an understudied language spoken in Canada and the USA. He uses data elicited from native speakers from Eastern British Columbia and proposes a typology of deverbal nominalizations in this language. Gatchalian argues that these nominalizations, formed by means of the left-peripheral

nominalizing particle *k*, fall into two classes, according to the level at which the particle attaches. vP-nominalizations (which in other theories may correspond to VoiceP) allow the realization of external arguments, while VP-nominalizations do not. When present, the external argument takes the shape of a possessor showing the same morphology as verbal subjects. The *k*-morpheme present in these nominalizations is analyzed as the head of a category-changing nP, which is able to take various levels of verbal structure as its complement. Besides an application of current theoretical tools to an understudied language, this article presents particularly interesting data, inasmuch as nominalizations are omnipresent in Ktunaxa, and show a complex structure even when they denote entities like *bed* or *pyjamas*.

The studies by Varvara and Wauquier, Hathout & Fabre contribute a computational linguistics perspective on the competition between different deverbal nominalizing suffixes in Italian and French. **Varvara** investigates the productivity domains of the suffixes *-mento* and *-zione* in Italian in order to understand what properties of the base verb play a role in the selection of one suffix over the other. She analyzes the occurrence of 678 nominalizations in a corpus from 1841 to 1947 and thus focuses on the *realized* (and not the *potential*) *productivity* of these suffixes (Baayen 2009). By means of a logistic regression model, she evaluates how the length in characters, the inflectional class, as well as the presence and number of other affixations influence the selection of one suffix over the other. In particular, she finds that *-zione* attaches to verbs of the first conjugation, while *-mento* favors those of the second and third conjugation. Moreover, *-mento* nominals are more likely to appear with morphologically complex verbs and especially those prefixed by *a-*, even though base verbs with greater length in characters tend to form nominalizations with *-zione*. Varvara's conclusion is that productivity constraints do not represent strict rules with binary outcomes but rather emerge as preferences with a graded effect.

Wauquier, Hathout & Fabre examine the degree of technicality in French nominalizations involving the suffixes *-age*, *-ion* and *-ment*, by combining Distributional Semantics and statistical modeling. Although all three suffixes appear to be approximately equivalent in forming action nominalizations, some semantic differences have been pointed out in the linguistic literature, and the authors are investigating their potential to be used as technical terms. After proposing a linguistic definition of technicality, Wauquier, Hathout & Fabre implement empirical, quantitative criteria based on corpora and lexical resources to determine to what extent these can adequately characterize the notion of technicality. Some of these criteria, such as the number of synonyms and definitions of a nominal, prove to be significant in discriminating *-age* as more technical than *-ion* and thus confirm the starting

hypothesis of the study. -*Ion* nominals are shown to be more heterogeneous than those with -*age*, while the results on -*ment* nominals are less clear. The authors highlight the need for additional criteria to evaluate technicality, including manual annotation, which would complement the aspects evaluated in their study.

Iordăchioaia, Schweitzer, Svyryda & Buitrago Cabrera address zero-derived nominals in English, which have received little attention in recent generative literature until Borer (2013). The core of the study is a newly created database of 1,000 English zero-derived nominals (e.g., *to walk > a walk*), which collects data on the semantic classes of their base verbs (e.g., change of state, psychological verbs, verbs of motion, communication, emission) and the different interpretations (event, result state, product, agent) that they may receive. The authors investigate the interpretation patterns of these zero-derived nouns in relation to the semantic type of their base verbs (i.e. result or manner, following Rappaport Hovav & Levin 1998), by also addressing some challenges to previous generalizations made in the literature. While some of these observations are confirmed by the database, zero-derived nominals are shown to also display some unexpected properties such as the relization of event readings with argument structure, which bring them closer to suffix-based nominals than previously assumed.

Bloch-Trojnar offers a syntactic analysis of Irish deverbal nominals from the perspective of Grimshaw's (1990) distinction between Complex Event, Simple Event and Result Nominalizations. On the basis of various empirical tests, she argues that Irish deverbal nominals do not encode complex events with aspectual properties; they may represent simple event nominals (also found in light verb constructions), which show event implications, and result nominals, which are devoid of any verbal properties and resemble lexical nouns. Bloch-Trojnar implements her analysis in Distributed Morphology and argues that simple event nominals incorporate a verbalizing vP, which is missing in result nominals. For the realization of internal arguments in simple event nominals, which lack aspectual structure, she argues in favor of a theory in which argument realization is independent of aspect – such as in Alexiadou's 2017 on Greek synthetic compounds) and against Borer's (2013) approach, in which argument structure is dependent on the projection of aspect.

Knittel & Villoing examine French verb-noun compounds that receive a Means interpretation (*couvre-pied* 'blanket', lit. cover-feet) and are derived from stative bases. These nominals are shown to be ambiguous between a Means and an Instrument reading, which leads the authors to discard previous treatments in terms of verbal homonymy, as put forward in Lexematic Morphology. Instead, Knittel & Villoing adopt the notion of a polysemous lexeme to account for this dual interpretation. They propose a formal account

based on Kratzer (2000) and Rothmayr (2009), by adopting Rothmayr's (2009) hypothesis of bi-eventive verbs. The verbal bases are asumed to have both an agentive (eventive) and a stative component, which accounts for the double Means/Instrument value of the verb-noun compounds under investigation. The distribution of the Instrument vs. Means/Instrument values is shown to rely on the state that the referent of the noun involved in the compound acquires after the occurrence of the event described by the verbal base. If this state is reversible, a double Means/Instrument reading obtains, while a permanent state entails a "pure" Instrument reading.

Tribout addresses what is called, especially in the French lexicalist literature, "nominalization by conversion" in French, also known as zero-derivation, i.e. pairs in which the output and the base of this word formation process have the same form such as in *marcher* 'to walk' – *marche* 'walk'. Tribout discusses the issue of directionality raised by these pairs, which challenge the traditional conception of derivational rules. After summarizing both derivational and non-derivational approaches to conversion, Tribout discusses and dismisses various criteria which have been used to determine directionality: diachronic ones, such as date of first attestation or etymology, as well as synchronic ones, such as semantic relations, noun gender or verb inflection. These criteria are evaluated on a corpus of 3,241 French morphologically complex noun-verb pairs, with a clear directionality. While all criteria may contribute to a decision in particular cases, none of them systematically applies to all noun-verb pairs. Tribout argues that the directionality of conversion in French is not determinable – which, she suggests, is also the case with some suffix-based derivations – and pleads for paradigmatic morphology as an appropriate framework to account for such formations.

The diversity of empirical, theoretical, and methodological contributions in this collection once again demonstrates that, even fifty years after Chomsky's *Remarks*, nominalization remains just as exciting a domain of research. It shows great potential for new hypotheses and ideas either by investigating lesser-studied languages or by applying new methodologies to well-known data. The empirical and theoretical picture that emerges is that of a variety of nominalizing constructions that span the lexical, morphological, and syntactic domains in a gradual continuum both within individual languages and crosslinguistically.

Given the variety of nominalizing constructions and approaches that it comprises, the present volume contributes new perspectives and advances in the field of word formation across languages. It may be of interest to scholars who work on syntax, morphology, lexical semantics, corpus linguistics, statistical modeling, and historical linguistics.

Acknowledgements:

References

Alexiadou, Artemis. 2017. On the complex relationship between deverbal compounds and argument supporting nominals. In Maria Bloch-Trojnar & Anna Malicka-Kleparska (eds.), *Aspect and valency in nominals*, 53–82. Berlin: de Gruyter.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick Jacobs & Peter Rosenbaum (eds.), *Readings in English Transformational Grammar*, 184–221. Waltham, MA: Ginn.

Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merjy Kytö (eds.), *Corpus linguistics. An international handbook*. Vol. 2, 899–919. Berlin: de Gruyter.

Borer, Hagit. 2013. *Taking form. Structuring Sense*. Volume III. Oxford: Oxford University Press.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press.

Kratzer, Angelika. 2000. Building statives. In Lisa Conathan, Jeff Good, Darya Kavitskaya, Alyssa Wulf, & Alan Yu (eds.), *Proceedings of the Twenty-Sixth Annual Meeting of the Berkeley Linguistics Society*. 385–399. Berkeley: Berkeley Linguistics Society.

Lees, Robert B. 1960. *The grammar of English nominalizations*. The Hague: Mouton.

Lakoff, George. 1970. *Irregularity in syntax*. New York, NY: Holt, Rinehart and Winston.

Ogawa, Yoshiki. 2014. Diachronic demorphologization and constructionalization of compounds from the perspective of distributed morphology and cartography. *Interdisciplinary Information Sciences* 20. 121–161.

Rappaport Hovav, Malka & Beth Levin. 1998. Building verb meanings. In Miriam Butt & Wilhelm Geuder (eds.), *The projection of arguments: Lexical and compositional factors*, 97–134. Stanford: CSLI Publications.

Rothmayr, Antonia. 2009. *The structure of stative verbs*. Amsterdam: Benjamins.

Vendler, Zeno. 1968. *Adjectives and nominalizations*. The Hague: Mouton.

Gianina Iordăchioaia
Universität Stuttgart
Institut für Linguistik: Anglistik
Keplerstraße 17
70174 Stuttgart
gianina@ifla.uni-stuttgart.de

Elena Soare
Université de Paris 8 - UFR Sciences du Langage
UMR 7023 du CNRS
2 rue de la Liberté
93256 Saint Denis
elena.soare@univ-paris8.fr

Beiträge/Papers

Yoshiki Ogawa, Keiyu Niikuni & Yuichi Wada

Empty nominalization over antonymous juxtaposition/coordination and the emergence of a new syntactic construction

Abstract: In Japanese, direct combination of verbs or adjectives by coordination (with *to* 'and') or juxtaposition (with its empty counterpart) can form a NP, if the conjuncts are antonymous to each other; the coordinator *to* 'and' can combine only NPs elsewhere. We claim that this is because there is a phonetically empty nominalizer that can nominalize each conjunct, and that the new nominal construction has been gradually developing in the history of Japanese. An acceptability-rating experiment targeting 400 participants shows that the younger speakers were likely to judge this construction more acceptable than the older ones, that this tendency is slightly weaker in the Nominative condition than in the Genitive condition, and that the coordination condition was significantly worse than the juxtaposition condition.

Keywords: antonymous juxtaposition, antonymous coordination, Nominative/ Genitive conversion, empty nominalizer, diachronic development of a new construction, intergenerational differences in acceptability

1. Introduction

In Japanese, a direct combination of verbs (V) or adjectives (A) by coordination (with an overt coordinator *to* 'and') or juxtaposition (without *to*) can form a NP if the conjuncts are antonymous to each other, as in (1) and (2):[1]

---

1 An anonymous reviewer has suggested that *to* in (1b) and (2b) may be more like the quotative marker *to* in Japanese, which is not restricted to an actual quotation (cf. Suzuki 2006), than the coordinator *to* in (3a), because the former is either absent or present in both antonyms, while in the latter, the second *to* is optional. However, we cannot share the reviewer's intuition, because the quotative marker in Japanese can never be doubled in the form of *A-to B-to*, unlike the *to* in (1b) and (2b). The same reviewer has also suggested that if the *to* in (1b) and (2b) is a quotative marker, the clause containing it is subordinate or embedded in some way, and hence a Genitive subject therein could be subjected to Hiraiwa's (2002) analysis, according to which a Genitive subject is licensed by a nominalized complementizer. However, there is no synchronic or diachronic evidence that the complementizer *to* in Japanese is a nominalized complementizer that could license a Genitive subject. For these reasons, we continue to assume that the *to* in (1b) and (2b) is a coordinator. See also section 3 and note 5 for relevant discussions.

(1) a. Kodomo-**no/(?)ga** iru$_V$ i-nai$_A$ de/niyotte, ... < VP & AP>
    child-NOM/GEN is is-not with/depending.on
    'Depending on whether you have a child or not, ...'

  b. Kodomo-**no/ga** iru$_V$ to i-nai$_A$ to de/niyotte, ... <VP & AP>
    child-NOM/GEN is CONJ is-not CONJ with/depending.on
    'Depending on whether you have a child or not, ...'

(2) a. Koe-**no/(?)ga** ookii$_A$ chiisai$_A$ de/niyotte, ... <AP & AP>
    voice-GEN/NOM large small with/depending.on
    'Depending on whether a voice is loud or quiet, ...'

  b. Koe-**no/ga** ookii$_A$ to chiisai$_A$ to de/niyotte, ... <AP & AP>
    voice-GEN/NOM large CONJ small CONJ with/depending.on
    'Depending on whether a voice is loud or low, ...'

The coordinator *to* 'and' can combine only NPs elsewhere, as in (3), and a different type of coordinator must be used to combine two or more VPs or APs, as in (4) and (5):

(3) a. Sakana **to** niku(-**to**)-o taberu. <NP & NP>
    fish CONJ meat-CONJ-ACC eat
    'to eat fish and meat'

  b. *chiisai **to** kawaii(-**to**) akachan <AP & AP>
    small CONJ pretty-CONJ baby (cf. (5a,b))
    'a small (and) pretty baby'

  c. *Sakana-o taberu **to** osake-o nomu(-**to**)-toki, … <VP & VP>
    fish-ACC eat CONJ sake-ACC drink(-CONJ)-when (cf. (4a,b))
    'when (you) eat fish and drink sake, …'

(4) a. Sakana-o tabe-**te** osake-o nomu. <VP & VP>
    fish-ACC eat-CONJ sake-ACC drink
    'to eat fish and (thereafter) drink sake'

  b. Sakana-o tabe-ta-**ri** osake-o non-da-**ri** suru.
    fish-ACC eat-PAST-CONJ sake-ACC drink-PAST-CONJ do
    'to repeat eating fish and drinking sake'

  c. Sakana-o taberu-**si** osake-mo nomu.
    fish-ACC eat-CONJ sake-also drink
    'to eat fish and also drink sake'

(5) a. akachan-wa chiisaku-**te** kawaii. <AP & AP>
    baby-TOP small-CONJ pretty
    'Babies are small and pretty.'

  b. akachan-wa chiisai-**si** kawaii.
    baby-TOP small-CONJ pretty
    'Babies are small and (also) pretty.'

In this article, we will argue that the coordinator *to* is available in (1b) and (2b) despite there apparently being no nominal conjuncts, because the coordination or juxtaposition of two VPs/APs that are semantically antonymous to each other involves syntactic nominalization of each conjunct or the entire conjunction via zero derivation. In allowing the alternation between Nominative and Genitive Cases for the subject, the construction in (1), which we henceforth refer to as the "Nominalization of Antonymous Combination" Construction (NACC), might be identified with the Nominative/Genitive Conversion (NGC), as in (6a,b):

(6) a. [Kyoo Taro-**ga/(?)no**   kuru **koto**]-wa dare-mo sira-nai. (*koto* = FN)
         Today Taro-NOM/GEN come fact-TOP any-MO know-NEG
         'Nobody knows (the fact) that Taro will come today.'
     b. [furuhon'ya-de   Taro-**ga/(?)no**   kat-ta]   hon
         used.book.store-at Taro-NOM/GEN buy-PAST book
         'the book that Taro bought at a used-book store'

In (6a,b) too, an adnominal finite clause has a subject whose Case morphology can alternate between Nominative and Genitive. Thus, if we only considered synchronic data, we might not be able to distinguish between the NACC and the NGC (cf. Hiraiwa 2002: 547). However, we will argue that the NACC in (1) and (2) should be distinguished from the NGC in (6) in diachronic terms: First, in the last 150 years or so, the NGC in (6) has seen a decrease in both type and token frequencies, while the NACC in (1) and (2) has seen an increase in both type and token frequencies; second, as far as the facts observed from the last century to the present are concerned, the NGC has been changing from a free alternation between the Nominative and Genitive subjects to the situation where the Nominative subject is increasingly preferred to the Genitive, while the NACC has been changing from a situation in which only the Genitive subject is allowed to one of free alternation between the Nominative and Genitive subjects. The two points will be shown by corpus studies and large-scale acceptability-rating experiments targeting hundreds of participants. The diachronic change in the NGCC shows that it has been undergoing what Ogawa (2014) calls "syntactic constructionalization," while the diachronic change in the NGC shows that it has been undergoing what Ogawa (2018) calls "clause shrinking." Given Bader and Häussler's (2010) experiments, there is a correlation between the frequency and acceptability of syntactic constructions. Hence, we predict that for the NGC with a decreasing frequency, younger speakers were likely to judge it less acceptable than older ones, whereas for the NGCC with an increasing frequency, the younger speakers are likely to judge it more acceptable than the older ones.

This article is organized as follows: Section 2 shows how the NACC and the NGC are similar to and different from each other. Section 3 shows that Scalise et al.'s (2009) semantic approach can explain the nominal property of the juxtaposition cases as in (1a) and (2a), but cannot be extended to the coordination cases as in (1b) and (2b). This section also shows how a syntactic analysis based on nominalization as zero derivation can accommodate the variants of the NACC in Japanese. Section 4 shows, on the basis of a survey of the Corpus of Historical Japanese (CHJ), how the various types of the NACC have developed diachronically, and explains its developmental stages in terms of Ogawa's (2014) hypothesis of "syntactic constructionalization." This section also proposes that the explanation of the developmental stages makes a prediction of differences in acceptability judgments of the various types of NACC by speakers of younger and older ages. Section 5 shows that the prediction made in section 4 is borne out by our own experiment targeting 400 native speakers of Japanese whose ages range from their 20s to their 60s, and presents a discussion of the statistical results. Section 6 is the conclusion.

2. Similarities and differences between the NACC and the NGC

Although there are numerous syntactic analyses of the NGC in Japanese (Harada 1971, Watanabe 1996, Hiraiwa 2002, Maki and Uchibori 2008, Miyagawa 2011, among many others), most of them are synchronic analyses and are basically divided into two subfields: a CP analysis and a TP analysis. The former posits that both the Genitive subject clause (GSC) and the Nominative subject clause (NSC) are a result of free alternation in a finite clause of a certain kind (Watanabe 1996, Hiraiwa 2002), while the latter argues that while the NSC is a both semantically and syntactically full-fledged finite clause, the GSC is semantically and/or syntactically defective in various senses. Thus, Miyagawa (2011) claims that the GSC has a defective tense and lacks a CP projection, with the result that feature inheritance from C to T is unavailable, which is why Nominative Case is not assigned to the subject; instead, the defective TP of the GSC is selected by another phase head D, from which its subject can receive Genitive Case. In this sense, the NGC is not a free alternation but a consequence of different syntactic manifestations and phase theory, a cross-linguistic component of UG. We suppose that although these previous analyses differ from each other in technical details, almost all of them (except for Watanabe 1996) share the assumption that the GSC is adjacent to a noun or nominal functional head. This assumption is reasonable because almost all instances of the GSC occur in adnominal clauses, such as relative clauses or nominal

complement clauses. Watanabe (1996) argues that the NGC in Japanese is also possible in a comparative clause, which apparently lacks a nominal superordinator, and hence the (ad)nominal property of the GSC is irrelevant to the Genitive subject licensing. However, Maki and Uchibori (2008) and Miyagawa (2011) argue convincingly that even in a comparative clause, the GSC can be optionally subordinated to an overt nominal element such as *teido* 'degree' or *no*, as in (7):

(7) Taroo-wa [Hanako-**ga/no**   yonda(-**teido/-no**) yori] takusan-no hon-o   yonda.
     Taroo-TOP Hanako-NOM/GEN   read-degree/NO   than   many-GEN book-ACC   read
     'Taroo read more books than Mary read.'   (Miyagawa 2011: 1270)

Hiraiwa (2002) claims that Genitive subject is licensed by a special verbal inflection that he refers to as the 'predicate adnominal (PA) form,' which is also called *rentai-kei* in Japanese linguistics, rather than the nominal element. This claim is allegedly supported by the well-formedness of the data such as (8a), where there is no overt nominal category that could license a GSC. Even in these cases, however, Miyagawa (2011) argues that an overt nominal element can immediately follow the P-A form, as in (8b), so that the Genitive Case in (8a) can also be licensed by the null counterpart of the overt nominals:

(8) a. John-wa   [ame-**ga/no**   yamu-made]   offisu-ni i-ta.
         John-TOP rain-NOM/GEN   stop.ADN-until   office-at be-PAST
         'John was at his office until the rain stopped.'
     b. John-wa   [ame-**ga/no**   yamu-**toki/zikan**-made] offisu-ni i-ta.
         John-TOP rain-NOM/GEN   stop.ADN-when/time-until office-at be-PAST

Hiraiwa (2002: 547) also argues that examples like (1b) are a case of the NGC licensed by the P-A form. Even in these cases, however, we can easily find a variant in which the bare V/A conjunct is immediately followed by an overt nominal complementizer *no*, as in (9a,b):

(9) a. Ken-**ga/no**   iru$_V$ **no**   **to**   i-nai$_A$ **no**   **to**   de/niyotte, ... (V&A)
         Ken-NOM/GEN   is COMP CONJ   is-not   COMP   CONJ   with/depending.on
         'Depending on whether Ken is here or not, ...'
     b. Koe-**ga/no**   ookii$_A$ **no**   **to**   chiisai$_A$ **no**   **to**   de/niyotte, ... (A&A)
         voice-GEN/NOM   large COMP   CONJ   small   COMP   CONJ   with/depending.on
         'Depending on whether the voice is loud or small, ...'

Then, along Miyagawa's (2011) lines, one might argue that the account given for (7) and (8a,b) would explain (1b) and (2b) as well, assuming that (1b) and (2b) also contain an empty nominal complementizer corresponding to the overt *no* in (7), (8b), and (9a,b).

However, there are three reasons against the identification of the NGC and the NACC. First, although the juxtapositional NACCs in (1a) and (2a) are semantically identical to, and hence should be syntactically related to, the coordinational ones in (1b) and (2b), the NGC analysis of (1b) and (2b) cannot itself relate the two variants straightforwardly, as the former does not allow an overt nominal complementizer as in (10) as a variant:

(10)　Ken-**no/(?)ga**　　iru$_V$　　(*no)　　i-nai$_A$　(*no)　　de/niyotte, ...
　　　Ken-NOM/GEN　　is　　　(COMP)　　is-not　COMP　　with/depending.on
　　　'Depending on whether Ken is here or not, ...'

Second, in the standard NGC, the Nominative Case on the subject is unmarked, while the Genitive Case is marked. Nambu's (2014) and Ogawa's (2018) independent corpus studies show that the GSC has become less frequent in the last 100 years or so for colloquial Japanese and written Japanese, respectively, and Niikuni et al.'s (2017) and Ogawa et al.'s (2017) investigations show that the GSC is becoming increasingly unacceptable among younger age groups if the predicate is eventive. Moreover, the CHJ shows that while the standard NGC was found frequently from the Heian period (about the 800s to 1200s AD) on, the NACC of the coordination type in (1b) and (2b) types was almost never found before the Meiji period (from the 1860s on), and when it first emerged as a new construction, the Genitive subject was more frequent than the Nominative (as will be seen in Table 2 below). Even in present-day Japanese, as shown in (1) and (2), the Nominative counterpart is degraded for some speakers compared to the Genitive ones, which are always acceptable. No such restriction is observed for the normal NGC. Third, the NACC is severely degraded with an overt Genitive subject on the second conjunct, which is intended to give a contrastive focus, as in (11a), and the degradation effect is eliminated if the overt nominal complementizer *no* is inserted as in (11b). However, no corresponding effect is found in any cases of the NGC:

(11)　a.　Otoko-ga/?*no　iru　　to　　onna-ga/?*no　iru　　to　　de-wa, ...
　　　　　male-NOM/GEN　is　　CONJ　female-NOM/GEN　is　　CONJ　depending.on-TOP
　　　　　'Depending on whether men are there or women are there, …'
　　　b.　Otoko-ga/no　　iru no to　　onna-ga/no　　iru no to　　de-wa, ...
　　　　　male-NOM/GEN　is　NO CONJ　female-NOM/GEN　is　NO CONJ　depending.on-TOP

For these reasons, we claim that the standard NGC and the NACC (whether juxtaposition or coordination) should be analyzed as different syntactic constructions, in that the Genitive Case in the NGC is licensed by an overt nominal head or a null nominal complementizer (which is a descendant of the

adnominal inflection), whereas the Genitive Case in the NACC is licensed by an empty nominalizer. What remains to be seen is whether the nominalization by an empty nominalizer is well-motivated or not. To answer to this question, we shall critically review Scalise et al.'s (2009) alternative semantic analysis of exocentric compounds in general in order to show that the combination of non-nominal antonyms leading to a nominal category as in (1) and (2) is not a semantic (universal) phenomenon but a syntactic phenomenon.

3. The nouniness of the juxtaposition/ coordination of antonymic non-nominals

In the context of exploring the nature of exocentricity of compounds, Scalise et al. (2009: 74–75) argue that compounds formed by antonymic predicates are universally retracted to nouns, whatever the original categories of their constituents. Some examples are given below:

(12) a.   saliscendi 'climb$_V$ + descend$_V$ = latch'   <Italian>
      b.   subibaja 'climb$_V$ + descend$_V$ = lift'   <Spanish>
      c.   Dàxiǎo 'big$_A$ + small$_A$ = size'   <Chinese>
      d.   Chángduǎn 'long$_A$ + short$_A$ = length'   <Chinese>

In (12a,b), two semantically antonymous motion verbs are combined by compounding and the output is a noun semantically related to the verbs. In (12c,d), two semantically antonymous dimension adjectives are combined by compounding and the output is a noun semantically related to the adjectives. They argue that "there is a semantic primitive which underlies these examples: the notion of path. The path, taken as an ordered series of values, underlies both the notion of trajectory and the notion of scale. In the first case, the path is an ordered series of locative points and in the second case, the path is constructed over a series of degree values."

As for the question of why these two types of antonymic compounds are exocentric, with the entire nominal category underived from the category of their constituents, they provide a semantic explanation: In V-V compounds of the type in (12a,b), the component verbs denote opposite transitions inside the same locative path defined vertically, as below:

(13)   ascend -----------------→|←----------------------descend
       – -- +
                         Vertical space scale   (Scalise et al.: 78)

As they express semantically opposite meanings, "if the two lexical items are combined it cannot be inside a category which is bound by a time

expression such as the verb. Instead, the two expressions combined have a coherent meaning if they are taken to refer to the interval of the path where the two lexical items can overlap." A similar argument also applies to A-A compounds of the type in (12c,d): "once the standard comparison is fixed as a point in the scale, it is impossible that the same object exhibits a value of a property which at the same time counts as short and long."

(14) Short ----------------------→|←--------------------long
      − -- +
                    Degrees in the length scale   (Scalise et al.: 77)
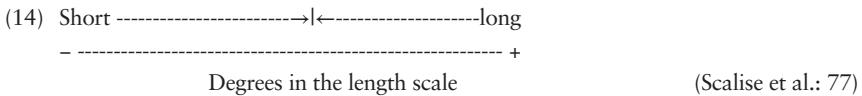
"The semantic denotation of a compound composed by two opposite adjectives is, therefore, incompatible with that of an adjective, but … it can be naturally used to name the whole scale."

   This semantic explanation of Scalise et al.'s (2009) is, however, not as motivated as it seems at first sight. It is true that one thing cannot satisfy the opposite value on a single scale at the same time. However, it is usually the case that every event and state extends over a certain spatiotemporal domain, rather than occupying a single spatiotemporal point. For example, although one thing cannot go up and down at the same time, it can repeat going up and down during a certain temporal interval. Although one thing cannot be short and long on a single scale, it can be short and long at different times (e.g., a tape measure can grow or shrink). Hence, in order to make a situation denoted in (13) and (14) contradictory, any semantic analysis of the exocentric compounds in (12) would have to prohibit the spatiotemporal extension of an eventuality denoted by the combined verbs or adjectives to any domain larger than a point. However, it is clear that such a semantic restriction is too strong to explain the empirical data. Thus, we can say that in Japanese, a combination of antonymous verbs does not refer to a vertical spatial scale itself but to an event of moving up and down repeatedly, as in (15a), or to a metaphorical extension of such an event, as in (15b):

(15) a. Agari-sagari-o   kurikaesi-te, kibun-ga   waruku-nat-ta.
          go.up-go.down-ACC repeat-CONJ mind-NOM bad-become-PAST
          'After going up and down repeatedly, I became sick.'
      b. Ano michi-wa agari-sagari-ga   hagesii.
          that road-TOP go.up-go.down-NOM drastic
          'That road goes up and down repeatedly and drastically.'

As *agari-sagari* in both (15a,b) are compounds denoting an event or its metaphorical extension, we need to say that the universal semantic principle can sometimes apply to a compound formed from antonymous verbs or adjectives and interpret it as a path but otherwise does not apply, thereby yielding a set of repeated events. But what is such a semantic universal?

Second, whatever explanation is made to induce the conclusion of semantic contradiction in (13) and (14) should not be extended to the possible noun phrases in (16a–c) and (17), which are composed of antonymous non-nominal words or phrases that are mediated by a coordinator and that behave as noun phrases as a whole:

(16)   a.   every now and then (= occasionally, once in a while, every time)
       b.   every here and there (= every place, everywhere)
       c.   With <u>every up and down</u>, you learn lessons that make you strong.

(17)   Agat-ta-ri   sagat-ta-ri-o   kurikaesi-ta.
       go.up-PAST-CONJ   go.down-PAST-CONJ-ACC   repeat-PAST
       'We repeated going up and going down.'

For example, in (16a), you can refer to every different point on a temporal path that is either identical to or different from the speech time without contradiction, from which the universal quantificational reading over time can be obtained. In (16b), you can refer to every different spatial point that is identical to or different from where you are without contradiction, from which you can obtain the universal quantificational reading over space. (16c) applies the same thing to every step of movement that is either upward or downward movement and obtains the non-contradictory meaning of "every movement." In (17), the coordination is mediated by *ri* 'and', a coordinator that combines two or more TPs to induce a reading of event repetition (cf. (4b)), and yet the entire coordinational structure is nominal, as shown by the Accusative Case- marker following it. If a contradictory situation can be avoided by the universal quantificational reading as in (16) or the repetitive readings as in (15) and (17), Scalise et al.'s semantic explanation of the nominal nature of (12a–d) cannot be a logical necessity. Hence, we will argue that the combination of two antonymous non-nominal categories, whether it is a word or a phrase, can be changed into a nominal category that refers to either a concrete object or an abstract event/state because the structure can involve some kind of syntactic nominalization.

More specifically, we claim that whenever two antonymous non-nominal words or phrases are juxtaposed or coordinated, a zero nominalizer as a functional category (represented as n) can be merged with either conjunct, so that (1a,b) have a syntactic structure as in (18a,b):[2]

---

2   An anonymous reviewer has asked how the second coordinator in *A-to B-to* in (1b) and (2b) appears in syntax. In this respect, the structures in (18a,b) contain innocuous simplification in that the syntactic position of the second coordinator *to* in (1b) and (2b) is not represented anywhere in the tree diagram. When more

(18)a.
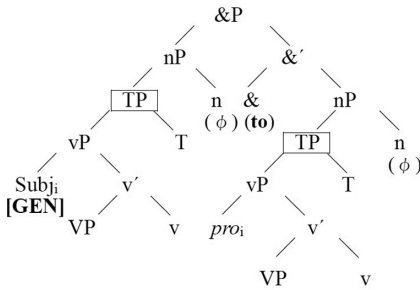
b.
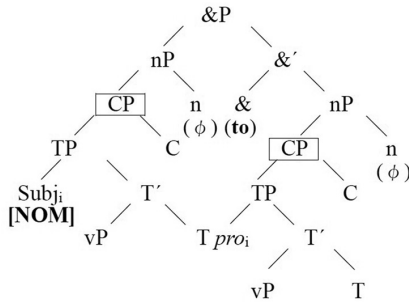
The left conjunct in (18a) has a Genitive subject, while that of (18b) has a Nominative subject, and in both cases the right conjunct has a coreferential subject *pro*. The two structures differ in terms of whether the nominalized category is TP or CP, but this distinction directly reflects Miyagawa's (2011) analysis of the NGC, according to which a Genitive subject is licensed by the phase head D (which in our analysis is replaced by n) that selects TP as

---

than one coordinator is represented in a coordinate structure, we can assume, following Johannessen (1998: 150), that the coordinator phrase (his CoP, our &P) as a functional projection can be stuck recursively. If we apply this assumption to (1b), where the linear order is *A-to B-to*, we can give it a more complex structure than (18a), as in (ia), where the first conjunct is merged with the lower coordinator (&2) whose Spec is filled with the second conjunct, the higher coordinator (&1) is merged with &P2, and the first conjunct is moved from the complement of &2 to the Spec of &1. Alternatively, we may assume that as in (ib), &P2 that occurs to the complement of &1 moves to the Spec of &1 to derive the same word order (Kayne 1994: 58).

(i) a. $[_{\&P1} [_{nP} [_{TP} Subj_j\text{-}GEN \ldots] n]_i [_{\&1'} \&1 \text{ (to) } [_{\&P2} [_{nP} [_{TP} pro_j \ldots] n] [_{\&2'} \&2 \text{ (to) } t_{nPi}]$

  b. $[_{\&P1} [_{\&P2} [_{nP} [_{TP} Subj_j\text{-}GEN \ldots] n] [_{\&2'} \&2 \text{ (to) } [_{nP} [_{TP} pro_j \ldots] n]] [_{\&1'} \&1 \text{ (to) } t_{\&P2}]$

However, we will continue to use the simpler structure in (18a) rather than (ia) or (ib) when discussing the issues in the text, as nothing in our argument hinges on the choice between the different coordinate structures.

its complement, while a Nominative subject is licensed by the phase head C that selects TP as its complement. (18a,b) enable us to explain why the coordinator *to* 'and,' which must always combine nominal elements as in (3), is used in (1b) and (2b) to combine two apparently non-nominal categories.

In (18a,b), we also assume that juxtaposition and coordination share the same structure, the only difference between them being whether the head of &P is covert or overt. There is an independent motivation for this assumption: First, in NP, two adjectives or two numerals that modify a noun can be either juxtaposed or coordinated, as in (19a,b); second, when more than two DPs/VPs are conjoined, a DP/VP that does not come last and another DP before it can be combined without an overt coordinator but with a comma, as in (19c,d):

(19) a. a long (and) happy life   <AP & AP>
      b. three thousand (and) eight hundred miles   <Numeral & Numeral>
      c. a man, (and) two women, and three dogs   <DP & DP>
      d. I see (, / and) hear and feel what's happened   <VP & VP>

For the same reasons, we assume that the overt coordinator is optional in (18a,b).

Note, however, that (18a) cannot be the only structure for a juxtaposition of non-nominals whose subject has a Genitive Case. This is because Genitive Case in Japanese can be assigned not only to a subject in a nominalized clause but also to a subject in a pure noun phrase, as in *John-no hon* 'John's book'. Hence, we assume that for juxtaposition, but not for coordination, a smaller and simpler structure like (20) is also available, at least for some speakers:

(20)

```
                      nP
                /            \
         Subjᵢ               n′
         [GEN]          /         \
                  √ A/√ V          n
                /        \
          √ A1/√ V1    √ A2/√ V2
```
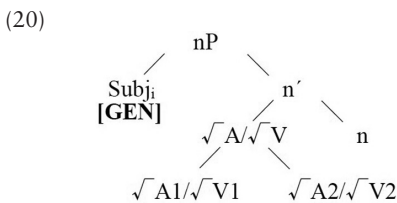
An anonymous reviewer has cast doubt on our proposed structure in (18a,b) and (20) on the basis of the following two facts: First, even when two verbs that are mutually antonymous are juxtaposed without an overt tense morpheme or an overt coordinator and appear to have a structure like (20), they can be modified by a manner-depicting adjectives as in (21), or the verb can take a direct object as in (22), so that they should have the larger structure like (18a):

(21) Bukka-no   kyuugeki-na   agari-sagari   niyotte, ...
     price-GEN   abrupt-AND   go.up-go.down   due.to
     'Because of (the repetition of) an abrupt increase and decrease of prices, ...'

(22) Otoshiyori-no   kaidan-no   nobori-ori-o   tetsudau.
     the.old-GEN   stairway-GEN   go.up-go.down-ACC   help
     'to help the elderly person's going up and down the stairway,'

It is important to note in this respect that the adnominal inflection on the adjective in (21) and the Genitive Case-marker on the direct object in (22) cannot be replaced by an adverbial inflection in (23a) or an Accusative Case-marker as in (24a) (in this respect, the NACC contrasts sharply with the NGC, where an adverb but not an adjective can co-occur), although the NACC with a Nominative subject are acceptable with them, as in (23b) and (24b):

(23) a. *Bukka-no   kyuugeki-ni   agari-sagari   de, ...
        price-GEN   abrupt-ADV   go.up-go-down   with
     b. Bukka-ga   kyuugeki-ni   agar-u   to   sagar-u   to de, ...
        price-NOM   abrupt-ADV   go.up-NONPAST   and go.down-NONPAST and with

(24) a. *Otoshiyori-no   kaidan-o   nobori-ori   de
        the.old-GEN   stairway-ACC   go.up-go.down-ACC   with
     b. Otoshiyori-ga   kaidan-o   nobor-u   to   ori-ru   to de, ...
        the.old-NOM   stairway-ACC   go.up-NONPAST   and go.down-NONPAST and with

Hence, we assume that when the NACC with a Genitive subject co-occurs with an adjectival modifier as in (21) or a Genitive-Case-marked object as in (22), it has a structure like (25a,b), where a nominal functional projection we tentatively represent as FP is located between nP and DP, rather than (18a), where vP and/or TP occurs below nP:

(25) a. [$_{DP}$ Subj-GEN [$_{FP}$ adjective/*adverb [$_{nP}$ √V1+√V2 n] F] D][3]
     b. [$_{DP}$ Subj-GEN [$_{FP}$ DP-GEN/*DP-ACC [$_{nP}$ √V1+√V2 n] F] D]

Second, the same anonymous reviewer has also cast doubt on our proposed structure in (18b) based on the fact that the realization of the coordinator *to* is not as freely optional as our theory would predict, pointing out the following minimal pairs of examples:

(26) a. Kodomo-ga   iru-inai-o   teema-ni   suru.
        child-NOM   be-be.not-ACC   theme-to   make
        'to place it on the agenda whether one has a child or not'
     b. *Kodomo-ga   iru-**to**   inai-**to**-o   teema-ni   suru.
        child-NOM   be-CONJ   be.not-CONJ-ACC   theme-to   make

---

3 See Ogawa (2001) for the distinction between nP (nominalizerP) and DP. Cf. also Cinque (2010) for the placement of each of the adjectives in NP in the Spec of various nominal functional categories.

(27) a. Shinchoo-ga takai hikui-o   hyooka-no   kijun-to   suru.
         height-NOM high short-ACC evaluation-GEN criterion-as make
         'to regard whether one's stature is high or low as a criterion of evaluation'

  b. *Shinchoo-ga takai-**to**   hikui-**to**-o   hyooka-no   kijun-to   suru.
         height-NOM high-CONJ short-CONJ-ACC evaluation-GEN criterion-as make

It is true that there is a sharp contrast in acceptability between the (a) and (b) examples in (26) and (27). However, (26b) and (27b) can become acceptable once the conjunctive coordinator *to* 'and' is replaced by the disjunctive coordinator *ka* 'or,' as in (28):

(28) a. Kodomo-ga   iru-**ka**   inai-**ka**-o   teema-ni   suru.
         child-NOM   be-CONJ be.not-CONJ-ACC theme-to   make
         'to place it on the agenda whether one has a child or not'

  b. Shinchoo-ga takai-**ka**   hikui-**ka**-o   hyooka-no   kijun-to   suru.
         height-NOM tall-CONJ short-CONJ-ACC evaluation-GEN criterion-as make
         'to regard it as a criterion of evaluation whether one is tall or short'

As the translations in these examples show, these constructions always have the semantics of disjunction rather than conjunction. This contrasts with the fact that what is overtly realized in (1b) and (2b) is a (nominal) conjunctive marker *to* 'and.' Recall here Scalise et al.'s (2009) discussion of what they call exocentric compounds, stating in essence that the *combination* of semantically antonymous verbs or adjectives will inevitably lead to contradiction, so that they cannot remain verbal or adjectival but have to be converted to nouns to name the whole scale. However, if the coordinator of two antonyms is that of *disjunction* rather than *conjunction,* which happens to be overtly realized sometimes as *to* and other times as *ka*, then the problem of semantic contradiction does not occur from the beginning. Hence, this fact also corroborates our claim that the entire phrase that contains mutually antonymous Vs/As has to be nominal, not to avoid the semantic contradiction, as Scalise et al. argue, but because there is a syntactic nominalizer that takes scope over the antonyms. Now, the remaining question is why semantic disjunction of antonymous Vs/As can sometimes be morphologically realized with the syntactic conjunctive marker *to*. Our tentative answer is that in the structures in (18a) and (18b), the empty head of &P, which is semantically *disjunctive*, is morphologically realized with *to* 'and' when each conjunct is nominalized, because the conjunctive marker *to*, as well as the disjunctive marker *ka,* can combine two or more NPs. Here, the categorial/morphological selection can override the semantic selection for some reason. Such a peculiar alternation between the conjunctive and disjunctive markers without changing the meaning is also observed in English, as in (29), where sentential negation requires switching from *and* to *or* as in (29a,b), and maintenance of *and* in the negative context leads to constituent negation:

(29) a. John bought apples and oranges.   (A&B)
    b. John did not buy apples or oranges.   (¬A&¬B)
    c. John did not buy apples and oranges.   (¬ (A&B))

As both interrogatives (of the *whether A or B* type) and the negation license negative polarity items (NPIs), it is not surprising that they behave alike in terms of the peculiar morphological switching. The only difference between (1b) and (2b) in Japanese and (29a,b) in English is that the switching pattern is opposite: While in the former the semantic *or* is expressed with a conjunctive marker, in the latter, the semantic *and* is expressed with a disjunctive marker. There is no *semantic* reason for why the sentential negation of conjunction needs to be expressed with a disjunctive marker as in (29b); (29c) could be semantically ambiguous between the wide- and narrow-scope readings with respect to the negation, given that a quantifier that is potentially scopally ambiguous with respect to negation seldom shows a morphological opposition like the one between (29b) and (29c) (this is natural because quantifier raising in LF should not have an effect on the PF side).[4] Thus, the switching from *and* to *or* in (29b) is fairly likely to be a matter of morphological parameter whose value is to be fixed at various levels of granularity, ranging from a nanoparameter to a macroparameter (Biberauer and Roberts 2012). A nanoparametric approach is also applicable to the question of why the disjunctive marker must be used in (28a,b), but the conjunctive marker can be used in (1b) and (2b), a question to which we cannot give a principled explanation in any event.[5]

---

4 Interestingly, in the Japanese counterpart to (29b), we do not use either *to* 'and' or *ka* 'or', but an additive morpheme *mo* 'also' as in (i), and the use of *to* 'and' in the scope of negation leads to constituent negation:
   (i) John-wa ringo-{**mo/*to/*ka**} mikan-{**mo/*to/*ka**} kawa-nakat-ta.
     John-TOP apple-{also/and/or} orange-{also/and/or} buy-NEG-PAST
     'John did not buy (either) apples or oranges.'
5 The conjunctive *to* in (1b) and (2b) can be replaced with the disjunctive *ka*, without changing the meaning:
   (i) a. Kodomo-no/ga   iru$_V$ **ka**   i-nai$_A$ **ka**   de/niyotte, ...   <VP & AP>
       child-NOM/GEN is   CONJ is-not CONJ   with/depending.on
     b. Koe-no/ga   ookii$_A$ **ka**   chiisai$_A$ **ka**   de/niyotte, ...   <AP & AP>
       voice-GEN/NOM large CONJ small   CONJ   with/depending.on
In Japanese, in the context of choice from alternatives too, the alternation between *to* and *ka* is possible:
   (ii) Ringo-{**to/ka** mikan-{**to/ka** banana-no-naka-kara   hi totu-o   erande-kudasai.
     apple-or/and orange-and/or banana-GEN-inside-from one-CL   choose-please
     'Please choose one from the set including an apple, an orange, and a banana.'

We believe that we managed to make it clear to the readers that the *whether A or B* clause in Japanese can have an &P structure whose head can be empty and that the head of the semantically disjunctive &P can sometimes be morphologically realized as the overt conjunctive coordinator *to* (in which case morphological selection can override semantic selection).

4. The diachronic development of the NACC

Now, let us shift our focus to the diachronic development of the NACC. Allowing the possibility of assuming the two structures (18b) and (20) for cases of juxtaposition entails a structural ambiguity for juxtaposition with a Genitive subject, but juxtaposition with a Nominative subject and coordination with either type of subject remain structurally unambiguous, because a Nominative subject requires CP and an overt coordination requires the functional projection &P, neither of which is present in (20).

It is important to note that of the two types of Genitive subject NACC, the one without an overt coordinator is found in literature written as early as the 11ᵗʰ century, the early Heian period, in the form of *ari-nasi* 'present-absent', as in (30a); by contrast, the version whose structure must be analyzed as in (18a) with an overt coordinator first appeared in literature in the 13ᵗʰ century, the Kamakura period, as in (30b), and we must wait until the late 19ᵗʰ century, the Meiji period, for the Nominative counterpart of (30b) to first appear in the CHJ:

(30) a. Turaki   kokoro-no   ari-nasi-o   mi-mu.   (c1010; *Izumi Sikibu Nikki*)
         hard   mind-GEN   present-absent-ACC   see-AUX
         'I wonder why you can see whether my mind is faithless or not.'
     b. Kore,   chie-no   aru to   naki   to   nari. (c1220; *Uji Shuui Monogatari*)
         this   wisdom-GEN   exist CONJ   absent CONJ MOD
         'This is (the difference in whether) a wisdom is present or not.'

A survey of the CHJ also shows that, putting aside the three strongly idiomatic expressions *ari-nasi* 'present-absent', *yosi-asi* 'good-bad', and *suki-kirai* 'like-dislike', both the type frequency and token frequency of the juxtapositional NACC has gradually increased over the last 1000 years, as shown in Table 1. The CHJ also shows that the coordinational NACC was not found before the Meiji period, and originally it was only compatible with a Genitive subject in the Meiji-Taisho period, but from the Showa period on it has only been compatible with a Nominative subject, as in Table 2. Judging from the corpus data, we may assume that the NACC construction started from the juxtapositional NACC with a Genitive subject in (20), and developed to the coordinational NACC with a Genitive subject and an optional coordinator in (18a), after which it developed from (18a) to the coordinational NACC

with a Nominative subject and an optional coordinator, as in (18b). This developmental path of the syntactic construction is fairly in line with what Ogawa (2014) calls "syntactic constructionalization," which is a general process of diachronic syntactic change defined as in (31) and (32):

Tab. 1: *Juxtaposition of antonymous verbs and adjectives.*

	Heian-Edo	Meiji-Taisho	Showa	
the years of publications	900~1864	1874~1925	1970~80s	1990-2000s
Words compiled (M)	3.39	14.73	10.86	94.05
type frequency (V1&V2)	3	8	10	18
token frequency (V1&V2)	12	66	61	828
type frequency (A1&A2)	2	4	4	13
token frequency (A1&A2)	2	4	5	38
type frequencies in total	5	12	14	31
token frequencies in total	14	70	66	866
PER MIL in total	**4.13**	**4.75**	**6.08**	**9.21**

Tab. 2: *Coordination of antonymous verbs and adjectives.*

	Heian-Edo	Meiji-Taisho	Showa	
the years of publications	900~1864	1874~1925	1970~80s	1990-2000s
Words compiled (M)	3.39	14.73	10.86	94.05
Genitive subject & antonyms				
token frequency	0	10	0	0
～no V1-to-V1{zaru／nai}-to	0	6	0	0
～no A1-to-A2-to (antonymous)	0	4	0	0
PER MIL	**0.00**	**0.68**	**0.00**	**0.00**
Nominative subjects & antonyms				
token frequency	0	0	1	12
～ga V1-to-V1{zaru／nai}-to	0	0	0	3
～ga A1-to-A2-to (antonymous)	0	0	1	2
PER MIL	**0.00**	**0.00**	**0.09**	**0.13**

(31) Syntactic Construction (cf. slightly modified from Ogawa 2014):
     If a morphosyntactic constituent that dominates two or more morphemes (Y_1, . . .,
     no paragraph Y_n, X) (n ≥ 1, X = head) contains at least one variable Y_i, call it a Syntactic
     Construction. Y_i is qualified as a variable only if there are at least two candidates
     no paragraph for substituting Y_i in combination with a particular head X.

(32) Syntactic Constructionalization (cf. slightly modified from Ogawa 2014):[6]

> When a syntactic constituent, which was not a syntactic construction at the earliest stage, becomes a minimal syntactic construction (i.e., with only one variable and one categorizer) at a later stage, and comes to have more than one variable and/or more functional categories than before and possibly enlarges the size of its syntactic constituent in a unidirectional fashion, in a way in accordance with the universal principle of structure building, functional hierarchy, and category selection.

Given (32), and given that the syntactic constructionalization of the NACC is now in progress among the native speakers of Japanese, we predict that certain versions of NACC are more acceptable for younger speakers than for older ones, unlike the standard NGC, which Niikuni et al. (2017) and Ogawa et al. (2017) have shown is more acceptable for older speakers than for younger ones due to what we call "clause shrinking." In the next section, we will explain the result of the experiment we administered, which bears out this prediction.

5. An intergenerational difference in the acceptability of the NACC

To test the validity of our prediction, we administered a large-scale Internet-based survey of acceptability judgments on each of the Nominative and Genitive subjects that occur in sentences like (2), targeting 400 participants ranging in age from their 20s to 60s. We examined the effects of participants' age on the acceptability of these sentences, as well as the differences in acceptability by type of sentences.

5.1. Methods

5.1.1. Participants

Our web-based survey included 567 native speakers of Japanese and was administered in the first half of 2019. All the participants met the following criteria: (i) born in the Tokyo metropolitan area (i.e., born in Tokyo, Saitama, Chiba, or Kanagawa Prefecture), (ii) raised in this area until the

---

6   Ogawa (2014) argues that (32) explains many phenomena involving diachronic syntactic change, such as the development from a stem compound (e.g., *hydrophobia*) to a word compound (e.g., *dog phobia*), from a resultative V-A form (e.g., *push open the door*) to the resultative construction (e.g., *push the door open*), from a lexical V-V compound (*tachi-kiru* 'separate-cut') to a syntactic V-V compound (*yomi-kiru* '(lit.) read-cut (= finish reading)') via the grammaticalization of the second verb in Japanese, and the development of the predicate inversion construction in DP from an apparent compound.

age of 15, and (iii) now living in this area. We finally analyzed the data from 400 participants (Table 3), who correctly answered each of the dummy items described in the Materials and Procedure section. Table 3 shows the age categories and the numbers of participants for each category.

Tab. 3: Participants of the Experiment.

Age (years old)	*N* (female)	Average age (*SD*)
20–29	80 (40)	25.7 (2.8)
30–39	80 (40)	34.8 (2.9)
40–49	80 (40)	44.7 (3.0)
50–59	80 (40)	54.1 (2.7)
60–69	80 (40)	63.8 (2.7)

5.1.2. Materials and procedure

We created 16 sets of experimental sentences, each of which can be assigned to one of the following 2 (Case: Nominative/Genitive) × 2 (Coordinator: absent/present) conditions:

(i) Nominative case/coordinator absent condition
heya-**ga**　　akarui kurai de, sagyoo-no　　siyasusa-wa　　　zenzen tigai-masu.
room-NOM bright dark by operation-GEN easiness-TOP　　totally differ-HON.
'Depending on whether the room is bright or dark, the ease of the operation differs totally.'

(ii) Nominative case/coordinator present condition
heya-**ga**　karui-**to**　　kurai-**to** de, sagyoo-no　　　siyasusa-wa　zenzen tigai-masu.
room-NOM bright-CONJ dark-CONJ by operation-GEN easiness-TOP totally differ-HON
(the intended meaning is identical to (i))

(iii) Genitive case/coordinator absent condition
heya-**no**　　akarui kurai de, sagyoo-no　　　siyasusa-wa　zenzen tigai-masu.
room-GEN bright　dark by operation-GEN easiness-TOP totally differ-HON
(the intended meaning is identical to (i))

(iv) Genitive case/coordinator present condition
heya-**no**　　akarui-**to**　　kurai-**to**　　de, sagyoo-no　　siyasusa-wa　zenzen tigai-masu.
room-GEN bright-CONJ dark-CONJ by operation-GEN easiness-TOP totally differ-HON
(the intended meaning is identical to (i))

Thirty-two filler sentences were also prepared, and the target sentences were distributed over four experimental lists using a Latin square design with conditions counterbalanced across lists. The filler sentences were added to each list, and the orders of the items were individually randomized.

A total of 48 sentences (16 target and 32 filler items) were presented on a page on the Web browser, and participants performed an acceptability-rating task in which they rated each sentence on a 5-point Likert scale ranging

from 0 (*unacceptable*) to 4 (*acceptable*). In addition, two dummy items were also inserted at random positions in the array of sentences. For these items, participants were instructed to make the specified answer: rating "0" for one dummy item and "4" for the other. If a participant made a different answer from what was specified to at least one dummy item, we excluded the participant's data from analysis.

5.2. Results

Taking the rating scores as the dependent variable, we performed linear mixed-effects model analyses with participants and items as random factors (Baayen, Davidson, & Bates, 2008). We included Case (Nominative/Genitive), Coordinator (absent/present), and participants' Age (continuous variable) as fixed effects with interactions between the factors. Case conditions and Noun conditions were deviation-coded, and the continuous variable (Age) was standardized (to *z*-scores). The *R* programming language and the *lmer* function within the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) were used for the analyses. Table 4 presents the results of the statistical analysis.

Tab. 4: Results of the linear mixed-effects model analysis for rating scores.

	β	SE	t	p
(Intercept)	2.564	0.056	46.12	< .001
Case	0.054	0.021	2.52	.012
Coordinator	−0.116	0.029	−4.06	< .001
Age	−0.048	0.045	−1.08	.280
Case × Coordinator	−0.120	0.044	−2.71	.007
Case × Age	−0.049	0.021	−2.32	.021
Coordinator × Age	−0.030	0.029	−1.06	.290
Case × Coordinator × Age	< 0.001	0.044	−0.01	.994
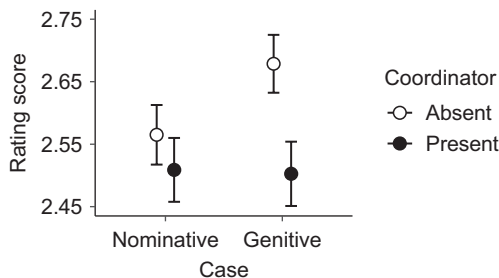
Fig. 1: Mean rating scores for each experimental condition. Error bars indicate standard errors of the mean by participant.

Figure 1 shows the mean rating scores for each experimental condition for all participants. Since the interaction between Case and Coordinator was significant, we tested the simple main effects of Coordinator for each Case condition as well as the effects of Case for each Coordinator condition. While there was no significant main effect of Coordinator in the Nominative Case condition (β = –0.06, SE = 0.04, t = –1.58, p > .1), in the Genitive Case condition we found a significant main effect of Coordinator (β = –0.18, SE = 0.04, t = –4.78, p < .001). The main effect of Case was also significant in the coordinator-absent condition (β = 0.11, SE = 0.03, t = 3.60, p < .001), but not significant in the coordinator-present condition (β = –0.01, SE = 0.03, t = –0.21, p > .1). These results indicate that sentences were judged more acceptable in the Genitive Case/coordinator-absent condition than the other three conditions.

In addition, since the interaction between Case and Age was also significant, we tested the simple main effects of Age for each Case condition. The analyses found no significant main effect of Age in the Nominative Case condition (β = –0.02, SE = 0.05, t = –0.51, p > .1) or in the Genitive Case condition (β = –0.07, SE = 0.05, t = –1.61, p > .1), although the coefficients suggest that younger speakers tend to judge the sentences to be more acceptable in the Genitive Case condition, and this tendency is slightly weaker for the Nominative Case condition. Figure 2 shows the Case × Age interaction plot (vertical axis: the rating scores predicted from the parameter estimates of the final regression model, which is shown in Table 4).
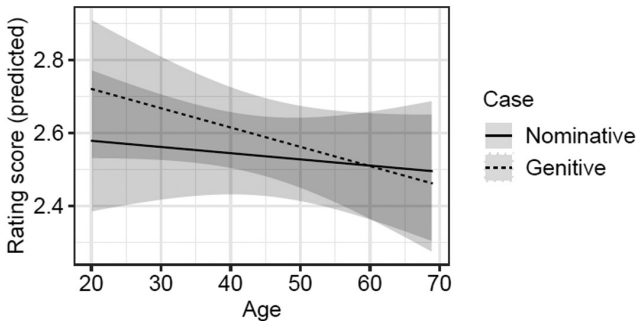
Fig. 2: Rating scores predicted from the parameter estimates of the mixed effects regression model. Shaded areas indicate 95% confidential intervals.

5.3. Discussion

The statistical analysis and its results obtained in the previous subsection show that (i) the Case × Coordinator interaction was significant, meaning that in juxtaposition but not in coordination the Genitive subject tends to be more acceptable than the Nominative subject, (ii) the Case × Age

interaction was significant, meaning that Genitive Case tends to be increasingly acceptable than Nominative Case for the younger speakers than for the older ones, and (iii) the Coordinator × Age interaction was not significant.

The first result is obtained by our assumption that the NACC is more likely to be constructionalized up to (18a) for the younger speakers, but has been constructionalized up to (18b) only for part of the youngest speakers, and that the overt realization of the coordinator *to* 'and' in (18a,b) is optional. In the absence of the overt coordinator, a Genitive NACC will be accepted with the structure in (20) for all the participants, but the Genitive NACC with the overt coordinator will be accepted only for those who accept (18a). Here arose a statistical significance. However, the Nominative NACC has the structure in (18b), which is the most constructionalized of all. For those who allow (18b), the coordinator is optionally realized, but the number is quite few. For those who do not allow (18b), the Nominative NACC is unacceptable, whether there is an overt coordinator or not. Hence, the presence or absence of a coordinator does not cause statistical significance, as far as the Nominative NACC is concerned.

The second result is also obtained because the syntactic constructionalization from (20) to (18a) to (18b) is now in progress. If (18a) is more acceptable for the younger speakers than for the older ones and if the number of those who accept (18b), the structure of the Nominative NACC, is significantly less than the number of those who accept (18a), a structure of the Genitive NACC, then it follows that the younger speakers are more likely to allow a wider variety of Genitive NACCs than the older ones and they are also more likely to accept the Nominative NACC. This is why there will be a much weaker effect of participant's age for the Nominative NACC than for the Genitive NACC.

Third, let us consider why the Coordinator × Age interaction was not significant. This is explained by our assumption on the optionality of an overt coordinator in (18a,b). For those who accept the Genitive NACC with the structure in (18a), it is acceptable, irrespectively of whether there is an overt coordinator or not. The same thing also applies to those who accept the Nominative NACC with the structure in (18b). Hence, for each speaker, acceptability will not differ, whether there is an overt coordinator or not.

6. Conclusion

We have argued that Japanese allows what we call the Nominalization of Antonymous Combination" Construction (NACC), which maps a combination of two non-nominal constituents with or without an overt coordination into a noun phrase. First of all, we argued that the NACC is a construction distinct from the Nominative/Genitive Conversion (NGC) in Japanese. As for the exocentricity of the antonymous combinations, we argued against

Scalise et al.'s (2009) claim that it is based on a universal semantic condition, and proposed that they behave as nominals when each conjunct of the NACC is capped by an empty nominalizer, and that the syntactic size of each conjunct of the NACC started from the root but has been enlarging to a larger category, including TP and even CP, as a result of a common diachronic process of "syntactic constructionalization" (Ogawa 2014). This proposal was supported by the intergenerational differences in the acceptability of the NACC.

References:

Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59. 390–412.

Bader, Markus & Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46. 273–330.

Biberauer, Theresa & Ian Roberts. 2012. Towards a parameter hierarchy for auxiliaries: Diachronic considerations. *Cambridge Occasional papers in Linguistics* 6. 267–294.

Cinque, Guglielmo. 2010. *The syntax of adjectives: A comparative study*. New York: Oxford University Press.

Harada, Sin'ichi. 1971. *Ga-no* conversion and idiolectal variation in Japanese. *Gengo Kenkyu* 60. 25–38.

Hiraiwa, Ken. 2002. Nominative-genitive conversion revisited. *Japanese/Korean Linguistics* 10. 545–558.

Johannessen, Janne Bondi. 1998. *Coordination*. New York: Oxford University Press.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82. 1–26.

Maki, Hideki & Asako Uchibori. 2008. *Ga/No* conversion. In Shigeru Miyagawa and Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 192–216. Oxford: Oxford University Press.

Miyagawa, Shigeru. 2011. Genitive subjects in Altaic and specification of phase. *Lingua* 121. 1265–1282.

Nambu, Satoshi. 2014. *On the use of case particles in Japanese: Corpus and experimental studies*, Ph.D. dissertation, Osaka University.

Niikuni, Keiyu, Yuichi Wada & Yoishiki Ogawa. 2017. Youninsei-no sedaikan-sa-ga simesu gengo henka-no yousou: Syukaku-zokkaku koutai-no baai [Nominative-genitive conversion and language change still in progress: Evidence from acceptability judgments by different age groups], *Cognitive Studies* 24. 395–409.

Ogawa, Yoshiki. 2001. *A unified theory of verbal and nominal projections*. New York: Oxford University Press.

Ogawa, Yoshiki. 2014. Diachronic demorphologization and constructionalization of compounds from the perspective of distributed morphology and cartography. *Interdisciplinary Information Sciences* 20. 121–161.

Ogawa, Yoshiki. 2018. Diachronic syntactic change and language acquisition: A view from nominative/genitive conversion in Japanese. *Interdisciplinary Information Sciences* 24. 91–179.

Ogawa, Yoshiki, Keiyu Niikuni & Yuichi Wada. 2017. Nominative/genitive conversion in Japanese and syntactic clause shrinking now in progress. *A paper presented at LACUS 2017*, held at Hamilton University, Canada.

Scalise, Sergio, Antonio Fábregas & Francesca Forza. 2009. Exocentricity in compounding. *Gengo Kenkyu* 135. 49–84.

Suzuki, Satoko. 2006. Quotative topic markers in Japanese. *International Journal of Language and Communication* 23. 67–96.

Watanabe, Akira. 1996. Nominative-genitive conversion and agreement in Japanese: A cross-linguistic perspective. *Journal of East Asian Linguistics* 5. 373–410.

Yoshiki Ogawa
Tohoku University
Graduate School of Information Sciences
6-3-09 Aramaki-Aoba, Aoba-ku, Sendai, Miyagi, 980-8579, Japan
ogawa@ling.human.is.tohoku.ac.jp

Keiyu Niikuni
Niigata Seiryo University
1-5939 Suido-cho, Chuo-ku, Niigata, Niigata, 951-8121, Japan
keiyu@n-seiryo.ac.jp

Yuichi Wada
Tohoku University
Graduate School of Information Sciences
6-3-09 Aramaki-Aoba, Aoba-ku, Sendai, Miyagi, 980-8579, Japan
yuwada@cog.is.tohoku.ac.jp

Obed Nii Broohm & Chiara Melloni

Mind your tones! The role of tonal morphology in Kwa action nominalization

Abstract: In the typology of West African languages, tone has been noted to play crucial grammatical and lexical roles, but its function in word formation has been less systematically explored and remains to be fully understood. Against this backdrop, the present study seeks to examine the form and function of tonal morphology in the formation of action nominals in four Kwa languages spoken in Ghana, namely Akan, Gã, Lɛtɛ, and Esahie, a relatively unexplored language of the Central Tano subgroup. Relying on data from both secondary and primary sources, we argue that tone raising is an important component of Kwa action nominalization, as it is found across different languages and derivational strategies. Specifically, while across the Kwa languages considered, tone raising tends to be an epiphenomenon of phonological conditioning, sometimes tone is the sole component of the nominalization operation or, as in Esahie, it concurs with the affix to the derivation, hence playing a morphological function.

Keywords: Kwa languages, tonal morphology, action nominalization, affixation, synthetic compounding

1. Introduction

Action Nominalizations (ANs)[1] are typically defined as nouns derived from verbs that preserve the event or state meaning denoted by the verbal base. More specifically, as stated by Porzig (1930–31), they are *Namen für Satzinhalte* 'lit., nouns for sentence contents' because they do not only preserve the meaning but also the arguments of the base predicate. ANs are very common in English, which has a rich array of suffixes for their derivation (e.g. *destruct-ion*, *govern-ment*, *dismiss-al*, *accept-ance*, etc.), but their occurrence, far from being limited to English or other Indo-European languages, represents a phenomenon robustly attested cross-linguistically (Koptjevskaja-Tamm 2005). A great deal of infra-linguistic and cross-linguistic variation is found as to the morphosyntactic means of nominalization and their relation to the corresponding sentence, with some

---

1 Abbreviations: AN(s): Action Nominalization(s); CL: Class Marker; DEF: Definite marker; H: High tone; HAB: Habitual marker; ICV: Inherent Complement Verb; L: Low tone; NMLZ: Nominalizer; NMLZ$_{inst}$: Instrumental Nominalizer; PL: Plural; PROG: Progressive marker; RED: Reduplicant; SG: Singular; TAMP: Tense-Aspect-Mood-Polarity; TBU: Tone Bearing Unit; 1SG: First Person Singular.

nominalization constructions displaying more verb-like traits and some others displaying more noun-like traits (Koptjevskaja-Tamm 1993). Also, a single language can make use of various means of nominalization. For instance, English does not only have several suffixes, but also deploys conversion, i.e. zero derivation (e.g., *(to) change > (a) change*), and prosodic means, i.e. stress alternation (e.g., *(to) incrèase > (an) ìncrease*) for the derivation of ANs.

Kwa languages (Niger Congo) are not an exception to this infra- and cross-linguistic variation and different morphological means are deployed for the formation of ANs. Kwa languages are in fact prone to deverbal nominalizations which behave like standard nouns in distribution and inflectional features, and deploy a variety of morphological tools, crucially including affixation, reduplication and compounding (cf. Adams 2001; Ofori 2002; Kambon 2012; Akrofi Ansah 2012a; Appah 2013; Boamah 2016; Campbell 2017; Asante 2018; Broohm 2019a). However, non-segmental strategies also play a pivotal role in these languages, which have a tone (non-stress) based phonological system, and where both tone and processual morphology are deployed in AN construction. In the typology of West African languages, tone has been noted to play crucial grammatical and lexical roles, ranging from tense-aspect-mood-polarity (TAMP) marking to information structure, alignment marking, and the signaling of lexical contrast (cf. Ward 1936; Dolphyne 1988; Akanlig-Pare and Kenstowicz 2003; Akanlig-Pare 2005; Marfo 2005; Schwarz 2009; Genzel and Duah 2015).

Although action nominalization has been fairly described in the Kwa literature (Akan: Obeng 1981; Appah 2005; Adomako 2012; Gã: Korsah 2011; Campbell 2017; Lɛtɛ: Akrofi-Ansah 2012a; Esahie: Broohm 2019a; 2019b), the role of tone in this word-formation operation has yet to receive adequate attention. The present study, therefore, offers a comparative overview of a few Kwa languages, i.e. especially Akan[2] (Central-Tano), Gã (Gã-Dangme), Lɛtɛ (Guan) and Esahie (Central-Tano),[3] and examines the form and function of tonal morphology in the formation of ANs in these languages.[4] We will focus on Esahie, on which very few studies have been conducted thus far, and we will describe the interplay of segmental and prosodic means deployed in AN derivation. Hence, considering that the languages in question still remain relatively under-described, this study, rather than providing a theoretical formalization, offers a panoramic description of AN formation in

---

2   We draw examples from all three major dialects of Akan, namely Fante, Akuapim, and Asante.

3   Occasionally, we resort to data from other Kwa languages such as Ewe (Gbe).

4   Following Stewart (1989), we shall collectively refer to all these Kwa languages as *Akanic* languages.

Kwa and the role of tone in this word formation operation. In particular, we aim to show that tone, and specifically tone raising, is a constant derivational means to form ANs across the various Kwa languages considered. Although various cases of tone raising could be explained as an effect of phonological conditioning triggered by a high-toned nominalizing suffix, the morphological role of tone in AN formation is enforced by the investigation of Esahie, which also exhibits tone raising on the last syllable of the base verb, despite the low tone on the nominalizing suffix. Further, cases of zero nominalization also argue in favor of a morphological role of tone: across Kwa, zero derivation is consistently characterized by a tonal change in the base verb, so that nominalization is null at the segmental level, but always marked suprasegmentally.

The paper is organized as follows. In section 2, we start with an overview of AN strategies in Akan, Gã, and Lɛtɛ. The role of tonal morphology in Kwa nominalizations is discussed in section 3. Section 4 focuses on Esahie, whose tonal system and nominalization strategies will be scrutinized. Finally, we will draw conclusions in section 5, by highlighting some general facts that emerge from our contrastive study.

2. Action Nominalization in Kwa

Action nominalization has been given a fair deal of attention in Kwa languages. Indeed, various issues have been discussed in the literature regarding the form and function of ANs from different empirical and theoretical perspectives. Mention could be made of works on Akan (cf. Obeng 1981; Appah [2003] 2005; 2013; 2015; 2017; Adomako 2012; Kambon 2012; Kambon et al. 2015; Kambon et al. 2019), Gã (cf. Adams 2001; Korsah 2011; 2016; Campbell 2017), Ewe (cf. Ofori 2002; Ameka 2006), Lɛtɛ (cf. Akrofi-Ansah 2012a), Lelemi (Boamah 2016), and very recently, Nkami (Asante 2018) and Esahie (Broohm 2019a; 2019b).

It is instructive to note that due to the relative under-description of Kwa languages in general, the focus of most studies has been descriptive rather than theoretical. The goal has primarily been to describe among others the strategies employed in the formation of ANs, the syntactic distribution and the semantic properties of ANs, in the ultimate interest of language documentation. We acknowledge that the present analysis draws inspiration from data and insights from the literature, particularly for Akan (Appah 2005; 2013; Adomako 2012; Kambon 2012), Gã (Korsah 2016; Campbell 2017), and Lɛtɛ (Akrofi Ansah 2009; 2012a).

As to strategies available for the formation of ANs, affixation, reduplication, and compounding constitute the most acknowledged mechanisms (cf. Obeng 1981; Adams 2001; Ofori 2002; Appah 2003; Akrofi Ansah

2012a; Boamah 2016; Asante 2018; Broohm 2019a). For instance, in addition to affixation and compounding, Gã is also noted to employ strategies of "processual" morphology, i.e. vowel lengthening (Campbell 2017).

In recent times, one issue that has been at the forefront of the discussion on action nominalization in Kwa has to do with the relation between compounding and nominalization. Kwa languages, as originally observed by Koptjevaskaja Tamm (1993), deploy a strategy of AN construction that could be defined as "possessive-incorporating": the verb argument structure is realized by means of an (optional) possessive modifier, in the case of the external argument, and a form of incorporation/compounding strategy for the internal argument (as in English *John's/his coffee-making*). Since the realization of the internal argument is often compulsory, this typological characterization of Kwa ANs points to a fascinating interplay between compounding and nominalization, as the former operation invariably feeds into the latter.

Finally, tone also plays a crucial role in AN formation: typically, in the presence of affixation and compounding, there is a change in the tonal melody of the base verb, but interestingly, sometimes, tonal morphology can manifest as the only (non-segmental) strategy of AN in some Kwa languages (Akan: Appah 2005; Adomako 2012; Appah et al. 2017; Lɛtɛ: Akrofi-Ansah 2012a; Gã: Campbell 2017; Esahie: Broohm 2019a; see section 2.2.).

2.1. An overview of AN strategies across Akanic and other Kwa languages

Action nominalization in Kwa typically involves the deverbalization of a verbal stem, but the nominalization operation may take the form of affixation, reduplication, (N-V) compounding and some processual morphology. In comparative terms, affixation is the most productive and cross-linguistically most applicable mechanism for the derivation of ANs in Kwa. In what follows, we provide an overview of how these strategies are employed across these languages.

Affixation. For most Kwa languages, ANs are derived via the attachment of a nominalizing affix to a verbal base. Let us consider the examples from Akan and Gã in Table 1, drawn largely from the literature (Akan: Appah 2003; Adomako 2012; Gã: Campbell 2017).[5]

---

5   The examples reported in the table show that ANs may be derived from stative bases, too. In these cases, the nominal preserves the *Aktionsart* characterization of the base. Hence, the term AN should be loosely intended to refer to deverbal nominalizations preserving the eventive or stative meaning of the base predicate.

Tab. 1: Action Nominalization via Affixation.

Language	Dialect[6]	Verbal Base	Action Nominal
Akan		Prefixation	
		bìsà 'ask/consult'	*à-bíꜜsá* '(spiritual) consultation'
	Fante	*kàé* 'remember'	*ǹ-káꜜé* 'remembrance'
	Akuapim	*wú* 'die'	*ò-wú* 'death'
	Asante	*dɔ́* 'weed'	*à-dɔ́* 'act of weeding'
	Asante	*sùrò* 'fear	*è-sùró* 'fear'
			(Appah 2003)
	Asante	*hìá* 'need/want'	*ò-hìá* 'poverty'
	Asante	*táń* 'hate'	*ɔ̀-táꜜń* 'hatred'
	Asante	*pɛ́* 'wish/desire'	*ɔ̀-pɛ́* 'will'
		pàgyà 'lift'	*m̀-páꜜgyá* 'upliftment'
		pàtà 'compensate'	*m̀-páꜜtá* 'compensation'
		hyìrà 'bless'	*ǹ-hyìrá* 'blessing'
			(Adomako 2012)
		Suffixation	
	Asante	*yàrè* 'fall sick'	*yàrè-ɛ́* 'sickness'
	Asante	*sɔ̀rè* 'adore'	*sɔ́rꜜé-ɛ́* 'adoration'
	Asante	*fɛ̀rè* 'be shy'	*fɛ́ꜜré-ɛ́* 'shyness'
Gã		Prefixation	
		mãlé '(to) lie'	*à-mãlé* '(a) lie'
		Suffixation	
		yè 'eat'	*yé-ꜜ**l̃i*** 'eating'
		bè '(to) fight'	*bè-ꜜ**i*** '(a) fight'
			(Campbell 2017)
		li 'to mock at someone'	*li-**mɔ́*** 'act of mockery'

Processual Morphology. For Gã, in particular, non-property verbs are nominalized by lengthening their final vowel, as illustrated in (1), where the difference between the input verb and the output nominal crucially lies in the length of the terminal vowel.

(1) a. *jù*   'steal'   *jùù*   'act of stealing'
     b. *wɔ*   'sleep'   *wɔɔ*   'act of sleeping'
     c. *jò*   'dance'   *jòò*   'act/style of dancing'   (Campbell 2017: 138)

---

6   Note that, for words that have the same form cross-dialectally in Akan, no dialectal information is provided.

However, vowel lengthening as a nominalization operation is not very productive in Kwa.

Reduplication. This strategy is highly productive in the morphological system of Akan (cf. Adomako 2012; Marfo and Osam 2018) and Ewe (cf. Ameka 1999; Ofori 2002). Let us consider the examples in Table 2 adapted from Ofori (2002) and Adomako (2012).

Tab. 2: Action Nominalization via reduplication

Language	Verbal Base	Action Nominal
	Total Reduplication	
Ewe	*xɔ̀* 'redeem'	*xɔ̀~xɔ̀* 'redemption'
	dzò 'fly'	*dzò~dzò* '(act of) flying'
	lɔ̃ 'love'	*lɔ̃~lɔ̃* 'love'
	śi 'to escape'	*si~śi* 'escape'
	vɔ̃́ 'to be afraid'	*vɔ̀~vɔ̃́* 'fear'
	Partial Reduplication	
	ɓíá 'ask'	*ɓá~ɓíá* 'questioning/question'
	fìà 'mutter'	*fà~fìà* '(act of) muttering'
		(Ofori 2002: 173–179)
Akan (Asante)[7]	*káń* 'read'	*à-kèn~káń* NMLZ-RED-read 'act of reading'
	bám 'embrace'	*à-bèm~bám* NMLZ-RED-embrace '(act of) embracing someone'
		(Adomako 2012: 52)

In Table 2, a verbal root is either partially or totally reduplicated to form an AN. In Ewe, for instance, the nominal *xɔ̀xɔ̀* 'redemption' is formed by reduplicating the verb *xɔ̀* 'redeem'. In Akan too, *abèmbám* '(act of) embracing' is formed from the verb *bám* 'embrace'. The difference between Ewe and Akan is that reduplication alone is enough for nominalization in Ewe, while Akan requires the use of dedicated nominalizers (e.g., the prefix *à-* in Table 2) in addition to reduplication. By implication, while nominalization in Ewe can be viewed as the result of the reduplication operation *per se*, this does not hold for Akan.[8]

---

7 See Adomako (2012), Osam et al. (2013), Marfo & Osam (2018) for more on reduplication in Akan.

8 Reduplication involving no affixation in Akan, if any, is a marked case.

Compounding. As noted in section 2, there is an interesting relationship between compounding and nominalization in Kwa. To be precise, compounding operations in Kwa are by default also nominalization operations (Dolphyne 1988; Akrofi Ansah 2012a; Appah 2013; 2015; 2016; Lawer 2017; Broohm 2019a). This is particularly interesting if we consider the nominalization of transitive verbs, which ultimately results in a form of synthetic compounding (see Melloni 2020 for an overview). In the nominalization process, the verb's internal argument becomes the first part of the complex AN synthetic compound (see Koptjevskaja-Tamm 1993 for more on the typology of nominalizations). It appears that the compounding operation is blind to the syntactic category of the input elements. Let us examine the data in Table 3 below.

Tab. 3: Action Nominalization via compounding

Internal structure	Akan (Appah 2016)	Dangme (Lawer 2017)	Lɛtɛ (Akrofi-Ansah 2012a)
	Base VP		
[V+NP]$_{VP}$	*bɔ òsé* make outcry 'jubilate'	*ngɔm yo* receive woman 'marry'	*bùè ń-dámfʋ̀* take PL-friend 'befriend'
	AN Synthetic Compound		
a. [N+V]$_N$ b. [[N+V]+suff]$_N$	*òsé!-bɔ́* outcry-make 'jubilation'	*yo-ngɔm* woman-receive 'marriage'	*ń-dámfʋ̀-bú!é* PL-friend-take 'friendship'

(Broohm 2019a: 239–245)

The first observation to make from Table 3 is, that, notwithstanding the syntactic category of their inputs, their outputs always bear a nominal syntactic category. The second observation involves synthetic compounds of the form [N+V]$_N$ or [[N+V]+suff]$_N$, which instantiate ANs. These compounds have been argued to be derived via a re-ordering of elements within a verb phrase (VP). The compounding (and, by implication, the nominalization) of such phrases may or may not involve overt (segmental) affixes. The Akan synthetic compound *òsé!bɔ́* [outcry-make] 'jubilation', for example, does not involve any overt segmental nominalizing affix. This leads us to introduce the role of tonal morphology in Kwa nominalizations.

3. Tonal morphology in Kwa nominalizations

The last nominalization strategy we deal with involves cases with no overt affixation, often described as *zero-derivation* or *conversion* in the literature on Indo-European languages. However, different from actual cases

of conversion in non-tonal languages, this type of AN is prosodic in that transposition is signaled via tone, a prominent phonological feature of Kwa languages. Hence, rather than attaching a segmental nominalizing affix, this nominalization process relies on the supra-segmental property of tone to derive ANs. Before describing tonal melody in AN formation, section 3.1. will briefly introduce tones in Kwa languages.

3.1. The role of tones in Kwa

In Kwa languages, characterized by the presence of more or less complex tone systems, tone has been noted to play crucial lexical and grammatical roles (cf. Christaller 1933; Boadi 1974; Dolphyne 1988; Dakubu 2002; Marfo 2005; Akrofi- 2003; Schwarz 2009; Frimpong 2009; Obiri-Yeboah 2013; Genzel and Duah 2015; Korsah and Murphy 2019). For the use of tone in signaling lexical and grammatical contrast, consider the following hackneyed Akan examples (Akan has two basic tones: High and Low tones, the former produced with relatively high pitch, the latter with relatively low pitch):

(2) a. *pàpá* 'father'
     b. *pàpà* 'fan'
     c. *pápá* 'good'   (Dolphyne 1988: 52)

In (2), we notice that the form *papa* has three potential meanings, depending on the tone melody it bears. It may be produced with a Low-High melody as in (2a), a Low-Low melody as in (2b), or a High-High melody as in (2c), to encode different meanings. Therefore, the meaning of a phonological word in Kwa languages is not only a function of the sound segments and their sequencing, but also of the pitch patterns associated with them.

Tone can also be used grammatically for TAMP marking. In (3), the difference between the form of the stative verb in habitual or progressive aspect is a matter of tone.

(3) a. *Abɔfra   nó   kòtó   hɔ́   (da   biara)*
         child   DEF   squat.HAB there day   every
         'The child squats there (everyday).'
     b. *Abɔfra   nó   kòtò   hɔ́*
         child   DEF   squat.PROG there
         'The child is squatting there.'

It is worth noting that the role of tone in TAMP marking in Kwa languages goes beyond signaling the difference between habitual and progressive aspect. Indeed, Dolphyne (1988) contends that in Akan, the grammatical function of tone outweighs the lexical function.

A related issue in Kwa tonology is the question of what phonological unit of a word is the tone-bearing unit (TBU), the mora or the syllable? For Akan, the fact that tone alteration or shift affects the entire syllable and not just a mora has led to a widely accepted view that the syllable, rather than the mora, is the TBU (Stewart 1965; Dolphyne 1988; Abakah 2002; 2005a, 2005b; Kügler 2016).[9] The same holds for many other Kwa languages (Gã: Dakubu 2002; Lɛtɛ: Akrofi-Ansah 2003; 2012b; Esahie: Frimpong 2009; Ewe: Motte 2013; Gua: Obiri-Yeboah 2013; Tutrugbu: Gborsi 2015; Essegbey 2019). Hence, we can safely assume that the syllable is the TBU in the Kwa languages of this study.[10]

Beyond the question of the TBU, another pertinent issue relates to *tone sandhi*.[11] According to Laver (1994: 476), "tone sandhi refers to the allotonic variation in the phonetic realization of tonemes due to contextual effects exercised by neighbouring tonemes." In other words, tones may be altered as a result of their interaction with tones of nearby syllables. Tone sandhi is particularly interesting since Akan, for instance, is a terraced-tone language whose lexical High tones are subject to a lowering process after low tones (Welmers 1959). In terraced-tone languages, this downtrend is triggered in the context of words with a *High-Low-High* (H-L-H) tone sequence. Once there is an underlying /H-L-H/ tonal sequence, its second H tone is realized with a pitch lower than that of the first H tone. This has been termed *downstepping*. We may further distinguish between *automatic downstepping* (downdrift) where the L tone trigger is overt, and *non-automatic downstepping,* where the L tone trigger is covert. Either way, downstepped H tones are marked with a superscript exclamation mark (ꞌ). Hence, in the Akan word *páꞌpá* 'good', the second H tone is (non-automatically) downstepped. Nevertheless, the downstepped H tone is an allotone of the H tone.

3.2. Tonal strategies

Tone Raising (without affixation).   Consider the Akan and Lɛtɛ nominalizations in Table 4: they are derived solely via a prosodic change in the tonal

---

9   Syllables in Kwa come in various shapes: they may be open syllables with a (CV, CVV, CVCV, CCV, V) structure, closed syllables with a (CCVC, CVC, VCC) structure, or may simply be a syllabic consonant (typically /m/ and /n/). The nucleus of the syllable is either a vowel or a syllabic consonant. This implies that vowels and syllabic consonants constitute syllables and two adjacent vowels constitute two distinct syllables. As such, they bear the tone (or tone marking) of the syllable.

10 In this paper, all relevant syllabic TBUs are written in bold.

11 Tone sandhi has also been referred to as *tonal assimilation* (see Abakah 2005b).

melody of the input verb. In order to get the bare form of the verb, which is crucial since we are making a case for tonal nominalization, we use the form of the verb in the imperative mood, which is the bare form for most Akanic languages (cf. Boadi 2008; Akrofi Ansah 2009; Agyemang 2016; Abunya 2018; Sakyi 2019).

Tab. 4: Nominalizing role of tone in Kwa languages.

Language	Verbal Base	Corresponding AN
Akan	*dùà* 'cultivate'	*dùá* 'tree/cultivation'
	kàsà 'speak'	*káˈsâ* 'language/speech'
	sòmà 'send'	*sòmá* 'errand'
	sèrèẁ 'laugh'	*séˈrêẃ* 'laughter'
	nàǹtsèẁ 'walk'	*náńˈtsêẃ* 'walking'
		(Appah 2005: 2)
Lɛtɛ	*gyɪ̀* 'eat'	*gyɪ́* 'eating'
	wùò 'descend'	*wúˈó* 'act of descending'
	nà 'walk'	*ná* 'act of walking'
		(Akrofi Ansah 2012a: 7)

From Table 4, we notice that, while the verbs bear a L or L-L tone as in the Lɛtɛ examples *nà* 'walk' and *wùò* 'descend', respectively, or a L-L-L-(L) tone as in the Akan example *sèrèẁ* 'laugh', their corresponding ANs, on the other hand, bear either a high tone as in *ná* '*walking*', a H-H tone as in *wúˈó* 'act of descending' or a H-H-H tone as in *séˈrêẃ* 'laughter'. The second H in the ANs *wúˈó* 'act of descending' and *séˈrêẃ* 'laughter' are both downstepped. This prosodic derivation of ANs, therefore, manifests itself in the form of *tone raising*, where the tone of the final syllable of the verb is raised from low to high. Undoubtedly, there are cases where all the tones of the verb are raised in the AN, with a leftward high tone spread affecting the whole verb stem (as in *sèrèẁ* 'laugh' vs. *séˈrêẃ* 'laughter'). However, since this is not ubiquitous, we would restrict ourselves to posit that the tone of *at least* the ultimate syllable of the base, is raised from a low to a high tone, if it is not underlyingly high.[12]

Tone raising + affixation. It is interesting to note that, even with affixed-derived ANs, as earlier discussed, the tone raising operation is still required in addition to the affixation operation. One would have observed that the

---

12 Where the final syllable of the verb is underlyingly high, there is no tone raising.

data earlier presented on affixed-derived ANs captures this fact. In what
follows (in Table 5), we present data to illustrate this point.

Tab. 5: AN derivation via Affixation + Tone raising

Language	Dialect	Verbal Base	Action Nominal	
Akan			Prefixation	
		bǐsà 'ask/consult'	*à-*	*à-bǐ'sá* '(spiritual) consultation'
		prà 'sweep'		*à-prá* 'sweeping'
	Asante	*sùrò* 'fear'	*è-*	*è-sùró* 'fear'
		pàgyà 'lift'	*m̀-*	*m̀-pá'gyá* 'upliftment'
		pàtà 'compensate'		*m̀-pá'tá* 'compensation'
		hyìrà 'bless'		*ǹ-hyìrá* 'blessing'
			Suffixation	
	Asante	*yàrè* 'to fall sick'	*-έ*	*yàré-'έ* 'sickness'
	Asante	*sɔ̀rè* 'adore'		*sɔ́'ré-έ* 'adoration'
	Asante	*fɛrè* 'to be shy'		*fɛ́'ré-έ* 'shyness'
Gã			Suffixation	
		yè 'eat'	*-lĭ*	*yé-lĭ* 'act of eating'

Table 5 shows ANs formed via (derivational) affixation. We observe that,
in addition to the affixation operation, tone raising is still required for a
well-formed AN. Hence, beyond the attachment of the nominalizing prefix
[*n-*] to the Akan verbal base *hyìrà* 'bless', the tone of the base is changed
from a L-L to a L-H sequence to derive the *ǹhyìrá* 'blessing'. A similar pro-
sodic change is observed in Gã (affixed-derived) *yé-li* 'act of eating', after
attachment of the nominalizing suffix [*-lĭ*]. Action nominalization in this
case is morpho-prosodic.[13]

   Tone raising + synthetic compounding. The final class of data we examine
involves (AN) synthetic compounds. Here too, in addition to the re-ordering
of the elements in the base VP, the tone of the verb in the compound is raised
along the lines of the pattern discussed earlier.

---

13 It is instructive to note that, as far as nominalizer-triggered tone raising is
   concerned, there is another suffix [*-mɔ*] in Gã, which has a less consistent prosodic
   behavior, as pointed out to us by an anonymous reviewer. We are grateful to the
   reviewer for drawing our attention to this. Further research on the properties of
   this nominalizer in Gã is needed to understand its prosodic pattern.

Tab. 6: AN-derivation via synthetic compounding + tone raising

	Underlying VP	**Action Nominal**
Akan	*bàà~bàè ànɔ́* RED-open mouth 'to engage in verbal exchanges'	*ànò-bààbá︎ˈé* mouth-RED-open '(act of) verbal exchanges (Appah 2013: 395)
Lɛtɛ	*bùè èsúmì* do work 'work'	*èsúmí-ˈbúé* work-do 'act of working' Akrofi-Ansah (2012: 8)
Gã	*yè òmɔ́* eat rice 'eat rice'	*òmɔ́-ˈyé-li* rice-eat-NMLZ '(act of) eating rice' Korsah (2011: 41)

Consider the examples in Table 6. The Akan and Lɛtɛ examples do not involve any kind of overt affixation; instead, the synthetic compound appears to derive from a re-ordering of elements in a VP, in addition to the usual prosodic change (i.e. tone raising). In the Gã example, however, synthetic compounding involves overt suffixation, coupled with tone raising, with the ante-position of the noun stem playing the role of the verb internal argument, as in (standard) synthetic compounding. The crucial difference between Gã, on the one hand, and Akan and Lɛtɛ, on the other hand, is that, in Gã, AN obligatorily requires both overt affixation and tone raising.

In keeping with what has been observed for other African languages such as Edo (Adéníyì 2010) and Tee (Anyanwu and Omego 2005), the data discussed above, presents evidence in support of the argument that tone plays a crucial role in the formation of (deverbal) nominals. The common pattern arising across the languages considered is a regressive high tone spreading, where the ultimate syllable of the verb (or all syllables) anticipates by assimilation the high tone of the nominalizing suffix. It is instructive to point out that a similar tone raising effect has been observed for Nkami (Guan), another Akanic (Kwa) language, where the attachment of a nominalizing prefix results in a raising in the tones of all TBUs in the base (Asante 2018). However, contrary to the directionality of the tone spreading as earlier discussed, Nkami ANs exhibit a progressive high tone spread, since the trigger operator is a prefix. Consequently, in ANs where there is no overt suffix, the high tone could be analyzed as a residue of a deleted syllable (i.e. a derivational suffix) bearing a high tone: this (floating)[14] high tone could hence take the role of a transpositional nominalizing affix in these cases and,

---

14 Following Goldsmith (1991), we define 'floating' the tones that are not linked to a vowel/syllable. They are nonetheless associated to the last syllable of the verb in the suffixation operation, where the inherent low tone of the last syllable of the base verb is substituted by the high tone.

arguably, act as the sole means of nominalization. This is illustrated below with data from Akan.

	Input	**Derivation**	**Output**
(4) a.	*yàrè*	*yàrè-ɛ́*	*yàréɛ́*
	'be sick'	sick-NMLZ	'sickness'
b.	*dùà*	*dùà-ó*	*dùá*
	'cultivate'	cultivate-NMLZ	'cultivation/tree'

On the one hand, as shown in (4a), the attachment of the nominalizing suffix *-ɛ́* to the base *yàrè* 'be sick' results in the tone raising in the final TBU of the base, due to a regressive spread of the high tone of the nominalizing suffix *-ɛ́* in the derivation of the output *yàréɛ́* 'sickness'. On the other hand, in (4b), where there is apparently no overt segmental nominalizing suffix to account for the tone raising effect exhibited in the output *dùá* 'cultivation/tree', we posit that a floating high tone associated with a (nominalizing) zero morph is what triggers the tone raising needed to distinguish the verb *dùà* 'cultivate' from the noun *dùá* 'cultivation/tree'. As shown in the derivation, the nominalizing zero suffix, represented as null '*ǿ*' in (4b), bears a (floating) high tone. Indeed, the possibility for tones to survive even after the loss or deletion of the segments (i.e. TBUs) they are originally associated with – technically called *tonal stability*, forms the basis for the influential analysis of tones as autonomous segments (Goldsmith 1976, 1990).

4. AN in Esahie

This section will be concerned with another Akanic language, Esahie, which has been fairly unexplored thus far, especially as far as word formation is concerned. Most of the data reported here were obtained through the elicited production of 35 language consultants, selected from across various Esahie speaking communities, while the prosodic analysis was performed with the speech analysis software Praat.

The illustration of Esahie AN does not only serve a documentation purpose but is also very relevant for a better understanding of the role of tone in Kwa AN formation. As shown in the previous sections, tone may be the only formal marking of nominalization, but in most cases the alteration of the tone pattern of the base verb can be explained in terms of a leftward assimilation which anticipates the high tone of the suffix. High tone spread, however, cannot be the explanation for the tone raising pattern we systematically find in Esahie ANs, because the nominalizing suffix bears a low tone. Therefore, Esahie data speaks in favor of a morphological function of tone in nominalization.

The section is organized in two parts: section 4.1. is a short description of the language and its tonal system, while a comprehensive assessment of AN is laid out in section 4.2.

4.1. An introduction to the Esahie language and its tonal system

Esahie (ISO 639-3: sfw) is a Kwa (Niger-Congo) language that belongs to the Northern Bia family of the Central-Tano subgroup (Dolphyne and Dakubu 1988) and is mainly spoken in Southern Ghana and parts of the Ivory Coast.

Like Akan, Esahie is a tonal language that distinguishes between two basic contrastive tones: a high tone (relatively high pitch) marked with an acute accent (´), and a low tone (relatively low pitch) marked with a grave accent (`) (cf. Frimpong 2009). Tone in Esahie plays both lexical and grammatical roles. In its lexical role, tone is used phonemically to show differences in meaning between otherwise identical words. For instance, in Esahie, the words *bo* and *gye* have two meanings depending on their tonal melody (see (5) and (6)). Thus, the meaning of a phonological word in Esahie does not only depend on the sound segments, but also on the pitch patterns they are associated with, similar to the Akan examples in (2).

(5) a. *bó* 'beat'
     b. *bò* 'buttocks'
(6) a. *gyé* 'defecate'
     b. *gyè* 'tooth'   (Broohm 2019b: 127)

In its grammatical role, tone in Esahie can be used to signal or alter the tense, aspect, mood, and polarity of verbs. For example, tone can be employed in distinguishing between the habitual aspect and the progressive aspect of Esahie verbs. Esahie habitual form of verbs is marked by a low tone on monosyllabic stems and L-H tone on the first and second syllables in disyllabic stems respectively, as in (7a) illustrating a monosyllabic verb. On the other hand, the progressive form is marked by a high tone for monosyllabic stems and their pronoun, and H-H-H tonal melody on disyllabic stems and their pronouns (Frimpong 2009), as in (7b) (see (3) for a comparison with Akan).

(7) a. *mé-kɔ̀*
         1SG-go.HAB
         'I go'
     b. *mè-búkyé*
         1SG-open.PROG
         'I open'

In addition to these functions, tone also plays a crucial morphemic role in Esahie nominalization.

4.2. AN formation in Esahie

As seen in section 2, Kwa AN formation makes use of segmental and prosodic patterns, and Esahie too exhibits an array of morphological and

tonal strategies in AN formation which, as we shall see, involves affixa-
tion and/or compounding, and systematically resorts to tone raising of the
last syllable of the base verb. In this section, we will present an overview
of the most common means of AN formation, focusing on the role of tone
raising and its interplay with affixation and N-V compounding. Lastly, we
will argue in favor of some (limited) cases of zero derivation in Esahie ANs.

Affixation. Esahie ANs are derived from monosyllabic and disyllabic
roots through affixation. With the affixation strategy, a nominalizing suffix
is attached to the verbal base, typically an unergative verb: the suffix [*-lɛ̀*]
appears to be the most regular and productive nominalizing affix in Esahie
and comes with three allomorphs [*-lɛ̀*], [*-rɛ̀*] and [*-nɛ̀*]. While [*-rɛ̀*] occasion-
ally occurs as a free variant of [*-lɛ̀*], [*-nɛ̀*] only appears in contexts where the
last vowel in the base verb has a nasality feature. Therefore, the distribution
of [*-lɛ̀*]/[*rɛ̀*] and [*-nɛ̀*] is phonologically conditioned. This is illustrated below
in (8).

(8)   [*-lɛ̀*] when the last vowel of the base is oral as in *dwùdwó-lɛ̀* 'language'

-lɛ̀ ——— [*-nɛ̀*] when the last vowel is nasal/nasalized as in *nĩngĩnĩ-nɛ̀* 'tickling'

        [*-rɛ̀*] a free allomorph of [*-lɛ̀*] as in *wònzɛ́-rɛ̀* 'pregnancy'

Now, let us proceed to analyze the internal structure and formation of Esahie
ANs.

		Input	**Output**
(9)	a.	*sũ*	*e-sṹ-nɛ̀*
		cry	SG-cry-NMLZ
			'(act of) crying'
	b.	*gó*	*e-gó-lɛ̀*
		dance	SG-dance-NMLZ
			'(act of) dancing'
	c.	nwã́fĩ	nwã́fĩ-nɛ̀
		run	run-NMLZ
			'(act of) running'
	d.	*nàtɛ̀*	*nàtɛ́-lɛ̀*
		walk	walk-NMLZ
			'(act of) walking'

The prefix *e-/ɛ-* may be also found with ANs, as in example (9a-b). Contrary
to what has been proposed for Akan, where the prefix may be the only seg-
mental marker of nominalization (see Table 5), we argue that this prefix
is not a derivational prefix but rather functions as a declension marker.
Declension markers appear to have a lexically determined distribution in
Esahie: they are often null (zero prefixes) but when they are overt they typi-
cally take the form of a vowel and only appear at the beginning of the word

in Esahie, which exhibits a residual declension system signaling number concord within the DP (Broohm 2017).

Tone raising. As observed for the other Kwa languages examined in the previous sections, suffixation is concomitant with a conditioned change in the underlying tonal melody of the base verb. The modification in the tonal melody of the disyllabic base verb is shown in the examples (9c–d), where the prosodic change occurs at least in the ultimate syllable of the disyllabic base(s). The change in tonal melody systematically involves tone raising in all those cases where the last syllable is not underlyingly high.

The tone raising pattern is not exceptional across the Kwa languages here considered, but the Esahie data are more challenging. Different from the other Kwa languages considered thus far, where the suffix bears high tone, the underlying tone of the *-lὲ* nominalizer is low. Therefore, apparently, there is no independent prosodic explanation that could account for the tone raising pattern in Esahie AN formation. Given that this same suffix (with a similar low tone) is found in other Kwa languages such as Nzema and Gã, we speculate that this suffix in Esahie was, possibly, historically borrowed or inherited from other Kwa languages. The historical account is plausible given that Nzema, in particular, is genetically and areally close to Esahie. Under this view, the *-lὲ* suffix would constitute a sort of pan Proto-Kwa nominalizing operator. Consider the following examples from Nzema and Gã.

Tab. 7: -lὲ as Proto-Kwa nominalizing operator

Language	Input	Output
Nzema	*wù* 'die'	*ὲ-wù-lὲ* CL-die-NMLZ 'death'
		Ndako (2011: 114)
Gã	*lǐ* 'to mock at someone'	*lǐ-lὲ* mock-NMLZ 'act of mockery'

Since the hypothesis of the suffix being historically inherited is in need of further corroborating data, we reserve its verification for future research.

To conclude, it is worth observing that this prosodic pattern is specific of AN formation, as it is not found with other cases of nominalization through suffixation. For instance, it is not attested in instrument nominalizations, as in (10), where the L-H tone suffix has no effect on the tonal melody of the base verb.

(10) zà   n-zà-lèέ
    'hang'   PL-hang-NMLZ_inst
                'a stick used to stake yam plant [so that it climbs around]'

Compounding and zero derivation. As in the other Kwa languages, action nominalization in Esahie is possessive-incorporating, i.e. the verb internal argument of (most) monotransitive eventive verbs is expressed obligatorily through noun incorporation, by which an AN turns up being realized as a synthetic compound (see section 2.1.; Broohm and Melloni, forthcoming). In these cases, indeed, [N-V] compounding is coupled with a segmental strategy of suffixation through [-lè] or its allomorphs. Crucially, synthetic compounds are characterized by the same tone raising pattern discussed above for unergative verbs, as shown in Table 8.

Tab. 8: Esahie AN-formation: Synthetic compounding + tone raising

Underlying VP	AN
bɔ̀ ndìrè	*ndìrè-bɔ́-lɛ̀*
hit weeds	weed-hit-NMLZ
'to weed'	'act of weeding'
sὲkyὲ dùmàà	*dùmàà-sὲkyé-lὲ*
destroy name	name-destroy-NMLZ
'to defame'	'(act of) defamation'
dì awùé	*awùé-lǐ-lὲ*
ICV death	death-ICV-NMLZ
'to murder'	'(act of) murder(-ing)'
hyὲ ɛhɔ̀ɛ̀n	*ɛhɔ̀ɛ̀n-hyé-lὲ*
ICV hunger	hunger-ICV-NMLZ
'to fast'	'(act of) fasting'
bɔ̀ mbáé	*mbáé-bɔ́-lὲ*
ICV prayer	prayer-ICV-NMLZ
'to pray'	'(act of) praying/prayer'

With respect to the other Kwa languages, however, Esahie seems to be more consistent as to the morphological and prosodic patterns employed in AN formation. First of all, suffixation is (mostly) necessary in ANs and is realized by [-lὲ]; prefixation with [è-] only appears in monosyllabic intransitive verbs and with an inflectional value. Also, contrary to Akan and other Kwa languages, zero suffixation/conversion is never attested with intransitive verbs (cf. examples from Akan in Table 4):

(11) a. *wònzὲ*   'to impregnate'   **wònzέ* (N)
    b. *kùrò*   'to love'   **kùró* (N)

However, cases of zero suffixation may be found when the base is a mono-transitive verb, whose object is expressed as the first member of a [N-V]$_N$ compound, as in (12).

		Output Nominal	Input VP	
(12)	a.	*nyame**sóm*** 'God worship'	*sòm*	*nyame*
			worship	God
	b.	*mogya**frá*** 'blood mixing, incest'	*frà*	*mogya*
			mix	blood
	c.	*eyiagỳ**iná*** 'sunshine'	*gỳinà*	*eyia*
			stand	sun
	d.	*nzɛm-bî'**sá*** 'questioning/question'	*bìsà*	*nzɛm*
			ask	matter

Superficially, the complex forms in (12) may appear as exocentric N-V compounds; however, issues of semantics and tonal melody point to the opportunity of analyzing these forms as instances of ANs, more specifi-cally, as synthetic compounds with a zero suffix. First, the meaning of these formations is eventive or stative, depending on the meaning of the base verb; further, like more 'standard' ANs, polysemy and meaning extensions are very frequent. Therefore, the zero-affixed N-V compounds cover the same semantic spaces as those affixed with *-lὲ*.[15] Second, even in the absence of *-lὲ*, it should be noticed that the N-V compounds in (12) display the same tone raising pattern attested for the (analogous) suffixed forms in Table 8. Therefore, the existence of an overt affix with a similar function and the tonal features of the compounds in (12) support the hypothesis of a zero-suffixation operation. Hence, while in Esahie zero suffixation (or conversion) is impos-sible with nominalization of intransitive verbs, it has limited productivity with monotransitive verbs. Despite their low productivity, the occurrence of synthetic compounds with a zero affix points to the crucial role of tonal morphology in Esahie nominalization: the high tone on the last verb syllable represents a derivational means for the formation of ANs, and demonstrates the role of tonal morphology as a derivational (specifically, transpositional) operation in Esahie, too, as in the other Kwa languages here considered.

5. Conclusion

Far from being an exhaustive assessment of AN formation in Kwa languages, the present study aimed at highlighting some common patterns across the

---

15 Like the English *-ing* and other nominalizing affixes, the Esahie nominalizing affix [*-lɛ*] is semantically multifunctional as it derives both eventive and resultative nominals (Grimshaw 1990).

derivational strategies adopted, and at examining the role of tones and tonal morphology, i.e. an issue not systematically investigated thus far, especially from a crosslinguistic perspective.

We wish to conclude our overview by emphasizing some general facts emerging from the empirical observation of several classes of nominalizations. First, AN involves a common pattern of tonal modification across the Kwa languages considered, despite their different genealogical affiliation and the various (segmental) means of word formation attested - including affixation, compounding and reduplication. Second, segmental and non-segmental morphology seem to go hand in hand in these languages: suffixation triggers a modification of the tonal contour of the base verb and this prosodic alteration follows a common pattern, which we have analyzed as a high tone spread (i.e. tonal extension/assimilation by anticipation) from the nominalizing suffix, bearing a high tone, leftward. This pattern is specifically attested in Akan, Gã and Lɛtɛ. However, although possibly being the result of a prosodic effect, tone apparently has a morphological value in AN formation, being in some cases the sole component of the nominalization operation, as proved by the many cases of 'zero derivation' (especially in Akan and Lɛtɛ), whereby nominalization is only marked suprasegmentally, i.e. by the raised tonal melody of the verb. Third, Esahie is arguably the most interesting case among the languages scrutinized: a L-to-H tone raising is systematically found in ANs notwithstanding the low tone borne by the suffix. Crucially, the high tone on the last verb syllable cannot be understood as an epiphenomenon of prosodic conditioning, i.e. the effect of a leftward tone assimilation by anticipation. Instead, we argued that a (floating) high tone, in association with the suffix, plays a morphemic role in the V-to-N transposition. Finally, cases of true conversion in AN are not found across the Kwa languages considered: (overt) affixes can be missing from an AN but some tonal cue is always present and marks the deverbal derivation as nominal.

Acknowledgments

Author contribution

This article is the result of the joint work of the co-authors, whose names are listed alphabetically. For academic purposes, OB takes responsibility for sections 2 and 3; CM takes responsibility for sections 1, 4, 5. OB is responsible for the Esahie data collection and conducted the prosodic analysis with the software Praat.

References

Abakah, Emmanuel N. 2002. The low tone in Akan. In Theda Schumann, Mechthild Reh, Roland Kießling & Ludwig Gerhardt (eds.), *Aktuelle Forschungen zu afrikanischen Sprachen: Sprachwissenschaftliche Beiträge zum 14. Afrikanistentag, Hamburg, 11–14 Oktober,* 2000, 193–210. Köln, DE: Köppe.

Abakah, Emmanuel N. 2005a. Tone in Akan nominals. In Mary E. Kropp Dakubu & Emmanuel K. Osam (eds.), *Studies in the Languages of the Volta Basin* (Vol. 3), 193–218. Legon, GH: Linguistics Department, University of Ghana.

Abakah, Emmanuel N. 2005b. Tone rules in Akan. *Journal of West African Languages* 32(1–2). 109–134.

Abunya, Levina N. 2018. *Aspects of Kaakye Grammar*. Accra, GH: University of Ghana dissertation.

Adams, George. 2001. *Nominalization Process in Ga*. Accra, GH: University of Ghana *MA Thesis*.

Adéníyì, Harrison. 2010. Tone and Nominalization in Edo. *California Linguistics Notes*, 35(1). 1–22.

Adomako, Kwasi. 2012. Verbal Nominalization as a Derivational Process: The Case of Akan. *Ghana Journal of Linguistics* 1(2). 43–64.

Agyeman, Nana A. 2016. *A descriptive grammar of Efutu (southern Ghana) with a focus on serial verb constructions: a language documentation study*. London, UK: University of London dissertation.

Akanlig-Pare, George & Michael Kenstowicz. 2003. Tone in Buli. *MIT Working Papers on Endangered Languages* 4, 1–33. Revised version in *Studies in African Linguistics* 35. 55–96.

Akanlig-Pare, George. 2005. *Buli Tonology: A Non-Linear Approach*. Accra, GH: University of Ghana dissertation.

Akrofi Ansah, Mercy. 2003. Function of Tone in Larteh. In Mary E. Kropp Dakubu & Emmanuel K. Osam (eds.), *Studies in the Languages of the Volta Basin* (Vol.1), 197–203. Legon, GH: Linguistics Department, University of Ghana.

Akrofi Ansah, Mercy. 2009. *Aspects of Lɛtɛ (Larteh) Grammar*. Manchester, UK: University of Manchester dissertation.

Akrofi Ansah, Mercy. 2012a. Action Nominalization in Lɛtɛ. *Ghana Journal of Linguistics* 1(1). 3–13.

Akrofi Ansah, Mercy. 2012b. Compound Formation in Lɛtɛ (Larteh). *The Journal of West African Languages* 39(2). 115–124.

Ameka, Felix. 1999. The typology and semantics of complex nominal duplication in Ewe. *Anthropological Linguistics* 41(1). 75–106.

Ameka, Felix. 2006. Ewe serial verb constructions in their grammatical context. In Alexandra Y. Aikhenvald & Robert M.W. Dixon (eds.), *Serial verb constructions: A crosslinguistic typology* (Vol. 2), 124–143. Oxford: Oxford University Press.

Anyanwu, Ogbonna & Christie Omego. 2015. Aspects of Deverbal Nominalization in Tee. In Ruth Kramer et al. (eds.), *Selected Proceedings of the 44th Annual Conference on African Linguistics*, 1–10. Somerville, MA: Cascadilla Proceedings Project.

Appah, Clement K. I, Reginald A. Duah, Obadele B. Kambon. 2017. Akan Noun-Verb nominal compounds: The exocentric synthetic view. *Language Sciences* 64. 1–15.

Appah, Clement K. I. 2003. *Nominal derivation in Akan: A descriptive analysis*. Trondheim, NO: Institute of Language and Communication studies (NTNU) *MA Thesis*.

Appah, Clement K. I. 2005. Action Nominalization in Akan. In Mary E. Kropp Dakubu & Emmanuel K. Osam (eds.), *Studies in the Languages of the Volta Basin* (Vol. 3), 132–142. Legon, GH: Linguistics Department, University of Ghana.

Appah, Clement K. I. 2013. *Construction morphology: Issues in Akan complex nominal morphology*. Lancaster, UK: Lancaster University dissertation.

Appah, Clement K. I. 2015. On the syntactic category of Akan compounds: A product-oriented perspective. *Acta Linguistica Hungarica* 62(4). 361–394.

Appah, Clement K. I. 2016. Akan Verb-Noun Compounds. *Italian Journal of Linguistics* 28(2). 3–24.

Appah, Clement K. I. 2017. On holistic properties of morphological constructions: the case of Akan verb-verb nominal compounds. *Acta Linguistica Hafniensia* 49(1). 12–36.

Asante, Rogers K. 2018. Nominalization in Nkami. *Journal of Linguistic Association of Nigeria* (JOLAN) 21(3). 99–126.

Broohm, Obed N. 2017. Noun Classification in Esahie. *Ghana Journal of Linguistics* 6(3). 81–122. https://10.4314/gjl.v6i3.4

Broohm, Obed N. 2019a. *Issues in Esahie nominal morphology: from inflection to word-formation*. Verona, IT: University of Verona dissertation.

Broohm, Obed N. 2019b. Lexical and Clausal Nominalization in Esahie: A descriptive account. *Journal of West African Languages* 46(2). 125–145.

Broohm, Obed N. & Chiara Melloni. Submitted. *Action Nominalization: A view from Esahie (Kwa)*. Verona, Italy: University of Verona manuscript.

Boadi, Lawrence A. 1974. Focus-marking in Akan. *Linguistics* 140. 5–57.

Boadi, Lawrence A. 2008. Tense, Aspect and Mood in Akan. In Felix K. Ameka & Mary K. Dakubu (eds.), *Aspect and Modality in Kwa Languages*, 9–68. Amsterdam & Philadelphia: Benjamins.

Boamah, Sophia. 2016. *Nominal derivation in Lelemi. A descriptive analysis*. Accra, GH: University of Ghana *MA thesis*.

Campbell, Akua A. 2017. *A grammar of Ga*. Houston, TX: Rice University dissertation.

Christaller, Johan G. 1933. *Dictionary of the Asante and Fante language called Tshi (Twi)*. Basel: Evangelical Missionary Society.

Dakubu, Mary K. 1988. *The Languages of Ghana*. London, UK: KPI Ltd.

Dakubu, Mary K. 2002. *Ga phonology (Language monograph n.6)*. Institute of African Studies: University of Ghana.

Dolphyne, Florence A. & Mary E. K. Dakubu. 1988. The Volta-Comoe Languages. In Mary K. Dakubu (ed.), *The Languages of Ghana*, 50–90. London, UK: Kegan Paul International.

Dolphyne, Florence A. 1988. *The Akan (Twi-Fante) language: Its sound systems and tonal structure*. Accra, GH: Ghana Universities Press.

Essegbey, James. 2019. *Tutrugbu (Nyangbo) Language and Culture (Grammars and Sketches of the World's Languages, vol. 8)*. Leiden, NL: Brill.

Frimpong, Victoria. 2009. *Aspects of Esahie Phonology: An Autosegmental Analysis*. Accra, GH: University of Ghana MA Thesis.

Gborsi, Divine K. 2015. *Aspects of Tùtrùgbù (Nyagbo) Phonology*. Accra, GH: University of Ghana MA Thesis.

Genzel, Susanne & Reginald A. Duah. 2015. H tone insertion on verbs in *na*-focus sentence in Akan. Paper presented at Morphosyntactic triggers of tone: New data and theories, Leipzig University, 12–13 June.

Goldsmith, John. A. 1976. *Autosegmental phonology*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology* (Vol. 1). Basil: Blackwell.

Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, MA: MIT Press.

Kambon, Obadele B. 2012. *Serial verb nominalisation in Akan*. Accra, GH: University of Ghana dissertation.

Kambon, Obadele B., Emmanuel K. Osam & Nana A. Amfo. 2015. *Studies in African Linguistics* 44 (2). 75–99.

Kambon, Obadele B., Reginald A. Duah & Clement K. I. Appah. 2019. Serial verb nominalization in Akan: The question of the intervening elements. In E. Clem, P. Jenks & H. Sande (eds.), *Theory and description in African linguistics: Selected papers from ACAL 47*, 361–386. Berlin, DE: Language Science Press.

Koptjevskaja-Tamm, Maria. 1993. *Nominalizations*. London: Routledge.

Koptjevskaja-Tamm, Maria. 2005. Action nominal constructions. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The World Atlas of Language Structures*, 254–257. Oxford: Oxford University Press.

Korsah, Sampson & Andrew Murphy. 2019. Tonal reflexes of movement in Asante Twi. *Natural Language & Linguistic Theory*. 1–59. https://doi.org/10.1007/s11049-019-09456-9

Korsah, Sampson. 2011. *Transitivity and inherent complementation in Gã*. Accra, GH: University of Ghana MA thesis.

Korsah, Sampson. 2016. On deriving inherent complement verb constructions. In Kate Bellamy, Elena Karvovskaya, Martin Kohlberger & George Saad (eds.), *Proceedings of ConSOLE XXIII 2015*, 52–69. Leiden, NL: Leiden University Linguistics Centre.

Kügler, Frank. 2016. Tone and Intonation in Akan. In Laura J. Downing & A. Rialland (eds.), *Intonation in African Tone Languages*, 89–130. Berlin: de Gruyter.

Laver, John. 1994. *Principles of phonetics*. Cambridge, UK: Cambridge University Press.

Lawer, Richard A. 2017. *Compounding in Dangme*. Accra, GH: University of Ghana MA dissertation.

Marfo, Charles O. & Emmanuel K. Osam. 2018. Optimal inflections as asymmetry between nominal and verbal reduplications in Akan. *Ghana Journal of Linguistics* 7(2). 33–51.

Marfo, Charles O. 2005. Akan Focus and Topic Constructions and the Prosody-Syntax Interface. *Journal of West African Languages* 32(1–2). 45–59.

Melloni, Chiara (2020). Subordinate and synthetic compounds in morphology. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. http://dx.doi.org/10.1093/acrefore/9780199384655.013.562

Motte, Mercy. 2013. *Vowel Hiatus Resolution in Ewe*. Trondheim, NO: Norwegian University of Science and Technology (NTNU) MA thesis.

Ndako, Stephen K. 2011. *The Syntax and Semantics of Possessive Structures in Nzema*. Winneba, GH: University of Education Winneba MA dissertation.

Obeng, Samuel G. 1981. *Nouns and Nominalization in Akan with Special Reference to the Twi Dialects*. Accra, GH: University of Ghana BA Essay.

Obiri-Yeboah, Michael. 2013. *Aspects of Gua (Gwa) phonology*. Accra, GH: University of Ghana MA thesis.

Ofori, Kafui G. 1988. *Nominalization in Ewe (The Processes)*. Accra, GH: University of Ghana MA thesis.

Osam, Emmanuel K., Charles O. Marfo & Kofi Agyekum. 2013. The morphophonology of the Akan reduplicated verb-form. *Journal of Language and Linguistic Studies* 9(2). 45–56. http://www.jlls.org/vol9no2/45-56.pdf

Porzig, Walter. 1930. Die Leistung der Abstrakta in der Sprache. Blätter für deutsche Philosophie 4: 6677. [repr. in: Hugo Moser (ed.), *Das Ringen um eine neue deutsche Grammatik. Aufsätze aus drei Jahrzehnten (1929–1959)*, 255–268. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969].

Sakyi, Joana P. 2019. *Modality and Evidentiality in Akan: A Corpus-Based Study*. Antwerp, BE: University of Antwerp dissertation.

Schwartz, Anne. 2009. Tonal focus reflections in Buli and some Gur relatives. In Sabine Zerbian, Laura J. Downing & Frank Kugler (eds.), *Special issue on tone and intonation from a typological perspective, Lingua* 1(19). 950–972.

Stewart, John M. 1965. The typology of the Twi tone system. *Bulletin of the Institute of African Studies*. 1–27.

Stewart, John M. 1989. Kwa. In J. Bendor-Samuel (ed.), *The Niger-Congo languages*, 216–245. Lanham: University Press of America.

Ward, Ida C. 1936. *An introduction to the Ibo language*. London: Heifer.

Welmers, William E. 1959. Tonemics, morphotonemics, and tonal morphemes. *General Linguistics* 4. 1–9.

Obed Nii Broohm
Department of Language and Communication Sciences
Kwame Nkrumah University of Science and Technology, Kumasi
KNUST Campus, Kumasi, Ghana
obed.broohm@knust.edu.gh

Chiara Melloni
Dipartimento Culture e Civiltà,
Università di Verona
viale dell'Università 4
37129 Verona, Italy
chiara.melloni@univr.it

Terrance Gatchalian

Deverbal nominalizations in Ktunaxa

Abstract: This paper presents an overview on deverbal nominalizations from Ktunaxa, a language isolate spoken in eastern British Columbia, Canada. Deverbal nominalizations are formed uniformly with a left-peripheral nominalizing particle *k* (Morgan 1991). However, they do not form a single homogenous class with respect to various syntactic properties. These properties are illustrated with novel data, showing that deverbal nominalizations fall into at least two classes, which are analyzed here as nominalization taking place at either *v*P or VP, where *v*P-nominalizations include the external argument and VP-nominalizations do not. Evidence for this division comes from how possession is expressed, the interpretation of the passive (and passive-like constructions), and the licensing of verbal modifiers. As both classes of deverbal nominalizations are constructed uniformly with the nominalizing particle, these properties are derived syntactically from the size of the verbal constituent being nominalized.

Keywords: Ktunaxa; Deverbal nominalization; possession; argument structure; passive

1. Introduction

Ktunaxa (isolate, also Kutenai) is a language spoken in Eastern British Columbia, Canada, and in northern Idaho and Montana, in the United States. In British Columbia it is spoken by 31 fluent speakers across four communities (Dunlop, Gessner, Herbert, & Parker 2018). The language reported on here comes from fieldwork conducted with two speakers in the ʔaq̓am First Nations near Cranbrook, BC, and with one speaker in Vancouver, BC.

There are several early descriptions of the Kuntaxa language from Canestrelli (1894), Boas (1926) and Garvin (1948a, b, c, d). More recently, there has been an effort to document and analyse various grammatical properties of Ktunaxa (Mast 1988; Dryer 1991, 1992, 1994, 2002; Laturnus 2011; Blamire 2011; Tammpere, Birdstone & Wiltschko 2012; McClay & Birdstone 2015; McClay 2017; Bertrand 2019), though there has been little attention payed to the grammar of nominalizations. In addition to the above works, the most substantial description of Ktunaxa to date is in Morgan (1991). While Morgan does include brief discussions on subordination and nominalization in what he calls *k*-forms (Morgan 1991:124, and especially chapter 4), there is no formal discussion of their syntax. This investigation provides the first steps in filling this gap by presenting novel data bearing on the internal syntax of deverbal nominalizations and accounting for their properties.

K-forms include questions (1a), subordinate clauses (1b), relative clauses (1c), and nominalizations (1d). In each of the examples below, the bracketed constituent is marked on the left-edge with the morpheme *k*. Examples throughout are presented in a slight variation from practical, community used orthography.[1] Unless otherwise noted, the examples in this paper were elicited by the author through grammaticality judgements on constructed examples and elicited speech with contexts supported through the use of storyboards (Burton & Matthewson 2015).

(1) K-forms in Ktunaxa[2]

    a.   qapsin kin ʔik

          qapsin   k-hin   ʔik

          what   K-2.SBJ eat

          'What did you eat?'

    b.   sukitmunapni niʔis k sukiɬ ʔik ɬsaːn

          sukitmun-ap-ni   niʔis k sukiɬ ʔik ɬsaːn

          make.happy-1.OBJ-IND DEM K well   eat John

          'It makes me happy that John eats well.'

    c.   sukaxniʔsi k̓ikɬis k ʔitkin aːn

          sukaxniʔ-s-i   k-ʔik-iɬ-s   k ʔitkin aːn

          good.taste-OBV-IND K-eat-PASS-OBV K make Anne

          'The food that Anne made tastes good.'

    d.   wiɬqaʔni k'it'iq'

          wiɬqa-ni   k-ʔit'iq'

          big-IND   K-stretch.INTRANS

          'The sweater is big.'

This paper investigates the syntax of deverbal nominalizations, such as (1d). Viewed from the outside, these forms are nominal and have the same distribution as other full DPs. However, there is some variation in their internal syntax. Morgan (1991: 305) presents one example of this: (2a) shows a deverbal nominalization with a possessor expressed with the verbal subject clitic, and (2b) shows one whose possessor is expressed with the usual nominal possessor marker. While there are a handful of such examples in Morgan's grammar, there is no discussion about why this variation arises and the examples are presented in passing.

---

1   Specifically, IPA symbols are used rather than orthographic <ȼ> and <ɫ> to represent the alveolar affricate /ts/ and the lateral fricative <ɫ>. In orthography, *k* is written as variably as a stand-alone particle or as part of the following word, generally when it precedes the subject proclitic or a pre-verb. Morgan invariably analyzes *k* as a proclitic (Morgan 2011:34–38).

(2) Possession of deverbal nominalizations

  a. ku ʔiknała   b. ka k'it'iq'

     k-hu   ʔik-nała   ka   k-ʔit'iq'

     NMLZ-1.SBJ eat-1.PL   1.poss nmlz-stretch

     'our food' (Morgan, 1991:305)   'my sweater'

The above contrast will be used to argue for the presence of verbal material embedded within the nominalization. The central claim I will make in this paper is that the *k* morpheme in nominalizations heads a category-changing *n*P (Marantz 1997; Wiltschko 2014) taking some level of verbal structure as its complement. Specifically, I propose that there are (at least) two sizes of *k*-form deverbal nominalizations in Ktunaxa, corresponding to VP-nominalizations and *v*P-nominalizations. As we will see, the picture that emerges is one where the varying properties of the nominalizations are derived entirely from differences in the embedded verbal structure. Under this view, the nominalizing n^0 is strictly a category-changing head.

2. Verbal structure overview

Ktunaxa has a generally free word order across major constituents, with the unmarked word order being VOS (Morgan 1991: 367). McClay (2017) presents evidence that word order is sensitive to focus, with SOV order also being common. Within constituents, however, there is a strict order of morphemes. This section overviews the basic syntactic properties of the verbal complex which are relevant for the discussion to follow. For a more complete discussion, see Bertrand (2019). The basic verb template is given in (3).

(3)   Basic Verbal Template

     SUBJ.PERS-(*pre-verbs*)-VERB.ROOT-(PASS/OBJ)-(SBJ.NUM)-OBV-IND

The leftmost element in the verb phrase is the subject pro-clitic, which appears with the exponents in Table 1. These pro-clitics express the person of the subject, while plural number of the subject, as well as the person and number of the direct object, is indicated as a separate suffix on the verb (see Morgan 1991: 242–244, Bertrand 2019 for discussion).

Tab. 1: Subject pro-clitics and suffixes

Person	SUBJ.PERS	SUBJ.NUM
1	hu=	-nała
2	hin=	-kił
3	∅	

Prototypically, the subject proclitics realize the external argument, which I assume is introduced in the specifier of v^0, which take a VP complement (Bertrand 2019).[3] In using the label vP, I would like to focus on its capacity to introduce external arguments (Chomsky 1995; also see VoiceP of Kratzer 1996). There is a vast literature on the various functions and articulations of v and extended verbal projections, and I do not wish to make specific claims about the verbalising or Case-checking functions of v^0, which require further investigation into the argument structure of Ktunaxa.[4] Additionally, whether roots are rather category-neutral √Ps and v^0 is a verbalizer (Marantz 1997) is a choice that is orthogonal to our concerns in this paper. As such, I will assume that verb roots are verbal VPs, use vP as a cover term for a verbal functional projection which introduces the external argument.

Immediately following the subject proclitic are a series of optional pre-verbal modifiers called *pre-verbs* (Dryer 2002; Morgan 1991 refers to them as (derived) adverbs). Pre-verbs are the only elements that may intervene between the subject pro-clitic and the verb root. They generally end in *-(i)ł* and convey a variety of meanings from temporal information to focus related information (McClay 2017: 76–86, Blamire 2011). Multiple pre-verbs are permitted (Dryer 2002).

Following the verb stem is PASS, which hosts the passive marker, *-i(ł)*, but also includes other voice-related morphemes such as the indefinite subject morpheme *-nam* which will be discussed below in greater detail in Section 3.2. I assume that these morphemes are realizations of v^0 (Bertrand 2019) in the sense discussed above, and will serve as important tools for developing the diagnostics for nominalizations.

The verbal obviative morpheme, *-(ʔi)s*, often functions as a switch-reference marker, and I assume following Bertrand (2019) that it is

---

2   The following abbreviations are used: DEM: demonstrative, IND: indicative, INST: instrumental, INTRANS: intransitive, NMLZ: nominalizer, OBJ: object, OBV: obviative, PASS: passive, PL: plural, POSS: possessive, SBJ: subject.

3   The subject proclitics do not exclusively mark external arguments. There has not yet been an in-depth investigation into the structure of unaccusative verbs, but forms such as *hu sahanni* 'I was bad' (Morgan 1991: 357) show that stative meanings typical of unaccusative verbs are marked with the same proclitic.

4   Specifically, questions such as whether Ktunaxa is a v/Voice-bundling language or not (Pylkkanen 2008, Harley 2017) requires further investigation, but the rich system of valency suffixes, which I do not touch upon in this paper (but see Morgan 1991: 290–308 for a catalogue of such suffixes), suggests that v and Voice are not bundled. This is further suggested by the co-occurrence between transitivizers and passivizers in forms such as *hun upi-ł-nałat-ił-ni* '1.SG die-TR-1. PL-PASS-IND' "We all got killed" (Morgan 1991: 301).

introduced as I^0. As a verbal category, it appears to be sensitive to relational properties between events in a discourse. Specifically, Bertrand, Birdstone, & Wiltschko (2017) proposes that it marks disjointedness between events in terms of event participants, event time, and event location.

The rightmost element glossed as IND is the indicative marker, *-(n)i*. It occurs on all matrix contexts and is banned from all embedded contexts including all *k*-forms discussed here. I follow Bertrand (2019) in assuming that this morpheme instantiates C^0.
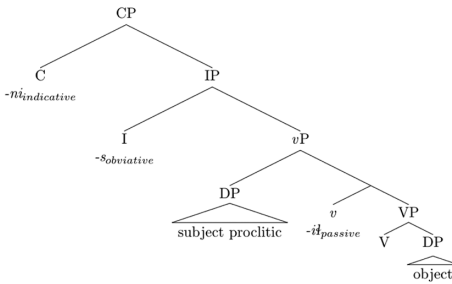
Fig. 1: Simplified clause structure of Ktunaxa.

The structure above is linearized via head-movement of the verb root through the functional projections dominating it (Bertrand 2019; cf. Dékány 2018 on an assessment of head-movement approaches to linearization). Note that each head corresponds to a suffix, and that the linear order of each suffix corresponds to the hierarchical position of its head, in accordance with the Mirror Principle (Baker 1985). The subject argument is raised to the specifier of the head occupied by the verb, generally Spec,CP in indicative clauses.[5]

3. Syntax of *k*-forms

Externally, *k*-forms have the distribution of nominal arguments as can be seen in (1) above. This section demonstrates that the internal properties of nominalizations are heterogeneous, even when restricted to deverbal nominalizations. The variation in these properties will be analyzed as the result of the point in the derivation at which nominalization occurs.

---

5   I do not have an account for whether this is for Case reasons or for phonological reasons. A syntactic account ultimately requires movement of subjects to Spec,CP in matrix indicative clauses, but permits them to remain low when the CP projection is absent (as I show is the case for nominalizations). A more fully articulated view of clausal syntax is necessary to resolve this question.

3.1. Possession of nominalizations

The first major distinction within the deverbal nominalizations arises from the realization of possessors. In one class of nominalizations, the possessor is realized with the normal nominal possessive morpheme, which is illustrated in the partial paradigm below. (4) shows the possessive on an inherent nominal, (5) shows the possessive on a nominalization. The full table of exponents is given in Table 2.

(4) Possession on inherent nominals

   a. ka xaʔɬt͡sin   b. xaʔɬt͡sinis   c. xaʔɬt͡sinʔis
       ka   xaʔɬt͡sin   xaʔɬt͡sin-nis   xaʔɬt͡sin-ʔis
       1.POSS   dog   dog-2.POSS   dog-3.POSS
       'my dog'   'your dog'   'his/her/its/their dog'

Tab. 2: Nominal possessive morphology

	SINGULAR	PLURAL
1	**ka=N**	**ka=N-naɬa**
2	**N-nis**	**N-nis-kiɬ**
3.PROX	**N-ʔis**	
3.OBV	**N-ʔis-ʔis**	

(5) Possessions on nominalizations

   a. ka k'it'iq' hanuhusni
       ka   k-ʔit'iq'   hanuhus-ni
       1.POSS   NMLZ-stretch   be.red-IND
       'My sweater is red.'

   b. k'it'iq'nis hanuhusni
       k-ʔit'iq'-nis   hanuhus-ni
       NMLZ-stretch-2.POSS   be.red-IND
       'Your sweater is red.'

   c. k'it'iq'ʔis f͡saːn hanuhusni
       k-ʔit'iq'-ʔis   f͡saːn   hanuhus-ni
       NMLZ-stretch-3.POSS   John   be.red-IND
       'John's sweater is red.'

In the other class of nominalizations, the possessor is instead realized with the subject proclitic. In such examples, the subject proclitic appears to the right of the *k*-morpheme with no additional possessive morphology, as in (6).

(6) Possession with the subject proclitic

    a.   ku q'umnimu hanuhusni

k-hu	q'umni-mu	hanuhus-ni
NMLZ-1.SBJ	sleep-INS	be.red-IND

       'My pyjamas are red.'

    b.   kin q'umnimu hanuhusni

k-hin	q'umni-mu	hanuhus-ni
NMLZ-2.SBJ	sleep-INS	be.red-IND

       'Your pyjamas are red.'

    c.   kq'umnimu hanuhusni

k-∅	q'umni-mu	hanuhus-ni
NMLZ-3.SBJ	sleep-INS	be.red-IND

       'His/her pyjamas are red.'

The examples in (7) show that expressing possession with nominal possession morphology as in (5) above is not possible with these forms.

(7) Nominal possessive morphology not possible

    a.   *ka q'umnimu hanuhusni

ka	q'umni-mu	hanuhus-ni
1.POSS	sleep-INS	be.red-IND

       Intended: 'My pyjamas are red.'

    b.   *kq'umnimu(ʔ)nis hanuhusni

k-q'umni-mu-nis	hanuhus-ni
NMLZ-sleep-INS-2.POSS	be.red-IND

       Intended: 'Your pyjamas are red.'

    c.   kq'umnimuʔis hanuhusni

k-q'umni-mu-ʔis	hanuhus-ni
NMLZ-sleep-INS-3.POSS	be.red-IND

       Intended: 'His/her pyjamas are red.'

Similarly, example (8) shows that this strategy of exponence with the subject proclitic is not available for the forms in which the possessor is expressed with the nominal possessive morpheme (compare with (5)).

(8) Subject proclitic not possible

    a.   *ku ʔit'iq' hanuhusni

k-hu	ʔit'iq'	hanuhus-ni
NMLZ-1.SBJ	stretch	be.red-IND

       Intended: 'My sweater is red.'

b. *kin ʔit'iq' hanuhusni

kin	ʔit'iq'	hanuhus-ni
NMLZ-2.SBJ	stretch	be.red-IND

Intended: 'Your sweater is red.'

c. *k'it'iq' ʦaːn hanuhusni

k-∅	ʔit'iq'	ʦaːn	hanuhus-ni
NMLZ-3. SBJ	stretch	John	be.red-IND

Intended:'John's sweater is red.'

The generalization that emerges is that deverbal nominalizations fall into two classes, based on how the possessor is expressed. What drives this division? Consider some of the verbs whose nominalization expresses possession with the verbal subject proclitic.

Tab. 3: *Possession Strategies for various verbs*

NOMINAL POSSESSION		SUBJECT PRO-CLITIC POSSESSION	
K-NOMINALIZATION	CORRESPONDING VERB	K-NOMINALIZATION	CORRESPONDING VERB
k'it'iq' 'sweater'	ʔit'iq' 'stretch.INTRANS'	kq'umnimu 'his/her/its pyjama'	q'umni 'to sleep'
kamak'ʦi 'orange (fruit)'	hamak'ʦi 'be orange'	kyawkɫifʦ 'his/her/its bed'	yawkɫifʦ 'to lie on top'
kanuhusnana 'apple'	hanuhus 'be red'	kiʔik 'his/her/its food'	ʔik 'to eat'
kawisxu 'banana'	hawisxu 'hang'	kqaːxniːmuɫ 'his/her/its apron'	qaːxni 'to cover'

In Table 3, there are two key take-aways. First, the referent of these nominalizations corresponds generally to the theme or other internal argument of the verb from which the nominalization is derived (on the instrumental suffix *-mu*, see section 3.3.). The possessor, when expressed as a subject pro-clitic, can be construed as the agent of the embedded event, or syntactically, the external argument of the embedded verb introduced by v (Chomsky 1995, cf. VoiceP of Kratzer 1996). That is, for a form like *kiʔik* 'food', a more faithful paraphrase would be "thing that he/she/it eats", with the agent of eating being interpreted as the nominal's possessor. This suggests an analysis such as Figure 2, where such a possessor is introduced in Spec,vP as in a clausal context. The internal argument *e* serves as the referential argument, or R-argument (Williams 1981), for the whole nominalization.

```
              nP
         ┌────┴────┐
         n         vP
         k      ┌───┴────┐
       NMLZ    hu        │
             1.SBJ   ┌───┴────┐
                     v        VP
                          ┌───┴───┐
                          V       e
                         ʔik
                       'to eat'
```
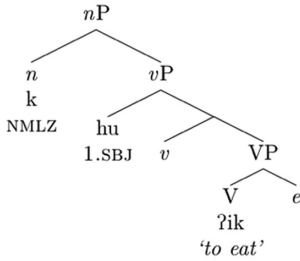
Fig. 2: Possessor is introduced in Spec,vP.

This analysis also predicts that the nominalizations of intransitive (unaccusative) verbs in which the possessor cannot be understood as the agent of the embedded event will not allow the possessor to be expressed with the subject proclitic. This prediction is borne out, as can be seen in (5, 8) above, as well as below in (9).

(9) Possessor cannot be the agent of the embedded eventuality

a.	ka	kamakʼʦi	b.	ka	kanuhusnana
	ka	k-hamakʼʦi		ka	k-hanuhus-nana
	1.POSS	NMLZ-be.orange		1.POSS	NMLZ-be.red-DIM
	'orange (fruit)'			'apple'	

These cases of nominalization are straightforwardly analyzed as a VP-nominalization, with the possessor introduced in a nominal projection parallel to the possession of an inherent nominal, external to the nominalizer. Crucially, these nominalizations exclude the external argument introducing vP. This is modelled here as a PossP projection, with the possessor in Spec,PossP. Again, the referential argument of the whole nominalization is the internal argument *e* of the verb.

```
                PossP
           ┌──────┴──────┐
          ka             │
        1.POSS    ┌──────┴──────┐
                 Poss           nP
                           ┌────┴────┐
                           n         VP
                           k     ┌───┴───┐
                                 V       e
                               ʔitʼiqʼ
                           'to stretch.INTRANS'
```
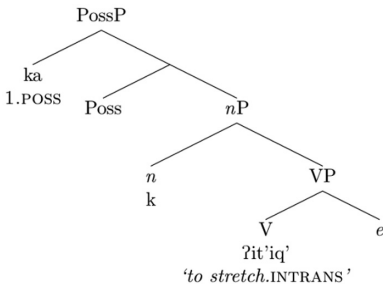
Fig. 3: Possessor is introduced in Spec,PossP.

The above discussion draws a picture of two mutually exclusive classes of nominalizations, distinguished by the level at which the nominalizer is introduced. This recalls Abney's (1987) analysis of English gerunds, where the

difference in where the eventive agent is expressed in *of-ing*, POSS-*ing*, and
ACC-*ing* gerunds. In Ktunaxa, however, the crucial variation lies not in where
the semantic agent of the embedded verb is expressed, but rather how the
grammatical subject of the embedded verb is interpreted: in nominalizations
the external argument can be interpreted as the possessor.

There are, however, a small number of forms which permit both possessive
strategies, given in (10).

(10) (i) Possessive Morpheme and Proclitic permitted
    a.   ka k'ałmałuma haqmaxunisni

ka	k-ʔałmałuma	haqmaxusn-is-ni
1.POSS	NMLZ-have.deep.voice	scare-2.OBJ-IND

        'My deep voice scares you.'
    b.   ku ʔałmałuma haqmaxunisni

k-hu	ʔałmałuma	haqmaxusn-is-ni
NMLZ-1.SBJ	have.deep.voice	scare-2.OBJ-IND

        'My deep voice scares you.'

Following the analysis above, the possessor is introduced after the
nominalization in Spec,PossP (10a) or before the nominalization in Spec,*v*P
(10b). While I have no account for why these verbs specifically permit
nominalization at both *v*P and VP, my consultant noted a slight interpretive
difference in the forms in (10) – (10a) is somewhat more direct and could
potentially be seen as insulting or derogatory in the third-person, an inter-
pretation which is absent in cases such as (10b).

The second key take-away from Table 3 is that the k-nominalizations
which take subject pro-clitic possessors are all interpreted as possessed by
a contextually salient third-person. Recall that the third-person subject pro-
clitic is phonologically null. While these forms appear morphologically sim-
ilar to the k-nominalizations which take nominal possessive morphology,
they are distinguishable semantically in that the third-person possessor inter-
pretation is obligatory. To derive an interpretation where the subject pro-
clitic possessor k-nominalization is unpossessed, additional morphology is
required. We turn to this immediately in the next section.

3.2. Passives, Indefinite Subjects, and Unpossessed k-forms

As noted, nominalizations which include the subject proclitic are obligato-
rily interpreted as possessed. This is unsurprising, given that the third-person
subject proclitic is phonologically null, and that these nominalizations were
proposed to include *v*P and the external argument.

(11) Possessive interpretation obligatory
      kiʔik
      k-ø-ʔik
      NMLZ-3.SBJ-eat
      'his/her/their food'
      #'food'

An "unpossessed" interpretation requires additional morphology. This takes
the form of the passive morpheme, -(i)ɬ. Presence of the passive morpheme
precludes the appearance of overt subject proclitics (12b).

(12) "Unpossessed" interpretation requires the passive
      a. k'ikiɬ   b. *ku ʔikiɬ
          k-ʔik-iɬ   k-hu   ʔik-iɬ
          NMLZ-eat-PASS   NMLZ-1.SBJ   eat-PASS
          'food/someone's food'   Intended: 'my food'

The scare quotes on "unpossessed" are due to (13), which shows that the
passive does not straightforwardly refer to a referent with no possessor. If this
were the case, there would be no reason why nominal possession couldn't
apply regularly to these derived nominals. In other words, the passive is
doing more than "removing" the external argument from the derivation.[6]

(13) No nominal possessive on passive nominalizations
      a. *ka k'ikiɬ
          ka   k-ʔik-iɬ
          1.POSS   NMLZ eat-PASS
          Intended: 'my food'

---

6   There are, however, some forms which occur with the passive and the nominal
    possessive morpheme. These are potentially problematic under the analysis
    developed here. One possibility is that these are lexicalized vP$_{Pass}$ nominalizations.
    Morgan (1991) discusses a contrast between lexicalized nominalizations and
    ad-hoc nominalizations, though he does not present any diagnostics or examples
    in reference to this contrast. I leave accounting for these forms and their
    relationship to those discussed in the body of this paper to further research.
    (i) Passive nominalizations, but allows nominal possessive morpheme

    a. ka kaqkfimuɬ hanuhusni
        ka   k-haqkfi-mu-ɬ   hanuhus-ni
        1.POSS   NMLZ-swim-INSTR-PASS   be.red-IND
        'My swimsuit is red'
    b. *ku haqkfimu hanuhusni
        ku   haqkfi-mu   hanuhus-ni
        K-1.SBJ   swim-INSTR   be.red-IND
        Intended: 'My swimsuit is red'

Bertrand (2019) analyses the passive morpheme *-(i)ł* as a v^0 head which, as we have seen above, introduces the external argument in its specifier. The presence of the passive is in these nominalizations is predicted, then, from their inclusion of vP. I assume that passive v^0 is distinct from active v^0 in that, rather than introducing an external argument, it existentially closes it (see Breuning 2013). The ungrammaticality of (13) follows from this: the passive does not exclude the external argument but rather existentially quantifies over it and makes its position unavailable for the introduction of other referents. Possession with nominal possessive morphology is not available because these forms are already possessed, but with an existentially-bound indefinite, introduced by passive v^0.
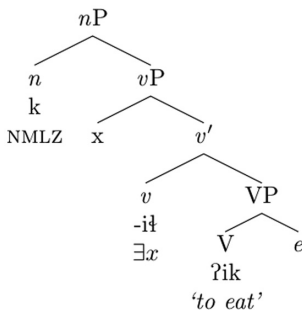
Fig. 4: *The passive in vP-nominalizations.*

The passive in Ktunaxa is restricted to transitives. In the case of nominalization of intransitive verbs, there is an indefinite morpheme which functions similarly to the passive. I assume that these are also instances of v^0 and existentially close the external argument in the same way as the passive.

(14) Indefinite subject marker for intransitive vP-nominalizations

　　a. kyawkłifsnam　　　　　　　　　b. *ku yawkłifsnam
　　　　k-yawkłifs-nam　　　　　　　　　　k-hu
　　　　NMLZ-lie.on-INDEF　　　　　　　　NMLZ-1.SBJ
　　　　'bed'　　　　　　　　　　　　　　Intended: 'my bed'

　　c. *ka kyawkłifsnam
　　　　ka　　　　　　　kyawkłifs-nam
　　　　1.POSS　　　　　lie.on- INDEF
　　　　Intended: 'my bed'

In a derived sense, then, vP-nominalizations are inalienable. Due to the inclusion of the external argument introducing vP, there is necessarily some argument which is interpreted as the possessor, whether that argument be expressed by the subject proclitic or existentially closed by the passive.

For VP nominalizations, an indefinite possessor marker is possible which is identical to the indefinite subject marker of intransitives.[7] Note, however, that in these examples the indefinite marker is not obligatory for the unpossessed reading – rather, the indefinite marker here is introduced after the nominalization in $Poss^0$.

(15) VP nominalizations with the indefinite morpheme marking possession

    a. k'it'iq'   b. k'it'iq'nam
        k-ʔit'iq'   k-ʔit'iq'-nam
        NMLZ-stretch   NMLZ-stretch-INDEF.POSS
        'sweater'   'someone's sweater'
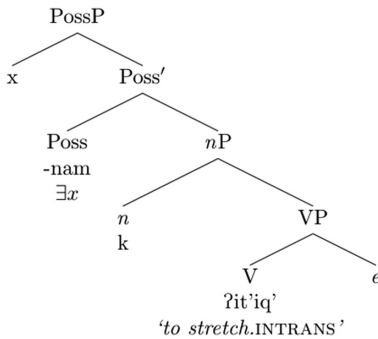
Fig. 5: *-nam introduced above nP interpreted as indefinite possessor.*

3.3. Instrumental readings

Returning briefly to some forms which express their possessor with the subject proclitic, there are several forms which include an additional instrumental morpheme, *-mu*, as in (16a). These are used quite productively in mechanical parts, such as car parts, or tools. Note that all these nominalizations naturally induce an instrumental reading, given the presence of the instrumental morpheme. Crucially, this morpheme is verbal, as can be seen in (16b).

(16) Instrumental morpheme in nominalizations

    a. kq'umnimuł   b. q'umnimu
        k-q'umni-mu-ł   q'umni-mu
        NMLZ-sleep-INST-PASS   sleep-INST
        'pyjamas'   'to sleep with (something)'

---

7 Thanks to an anonymous reviewer for suggesting the following line of analysis.

This shows that the referent of *k*-form deverbal nominalizations is not restricted to theme arguments but can be modulated by the presence of other verbal arguments. In other words, the assignment of the R-argument is not mediated by the nominalizing head but rather by the embedded verbal structure.

This resembles reference assignment in Blackfoot clausal nominalization, which does not have a consistent reference and whose referent is necessarily internal to the nominalized constituent (Bliss 2014). The argument that acts as the referent depends on the cluster of properties exhibited by the embedded clause. Blackfoot, however, has multiple types of nominalization and in fact has two dedicated nominalizers with consistent referent assignment: "abstract" or process nominalization with *-hsin*, and instrumental nominalization with *-a'tsis* (Bliss 2014; Wiltschko 2014; Ritter 2014). In Ktunaxa, nominalizations are all accomplished through *k*, which nominalizes its complement (nominalization-via-complementation in Wiltschko 2014).

I will leave a formal account of nominalization referent-assignment in Ktunaxa along the lines of Bliss (2014) for Blackfoot to future work. Informally, the presence of verbal argument-introducing morphemes (see Morgan 1991: 309–314 for more on these "involvement suffixes") modulates which argument functions as the referent for the nominalization. Given the position of the instrumental morpheme specifically, one possible line of analysis is that the highest verbal argument is the referent. This raises certain difficulties, however. Despite being present in *v*P nominalizations, external arguments are not candidates for the referent for cases discussed here, as in *kiʔik* 'food'. Further, Ktunaxa does have agent nominalizations, such as *k'anam* 'hunter' (< *aʔnam* 'to hunt'), which appear to contradict any analysis where the external argument is categorically excluded from referent assignment. I leave this for future research.

3.4. Preverbal Modifiers

The final piece of evidence for the contrast between *v*P and VP nominalizations is the presence of verbal modifiers, known as pre-verbs verbs (Dryer 2002; Blamire 2011; Morgan 1991: 33 refers to them as derived adverbs). These pre-verbs, as expected from their name, occur linearly before the main verb stem but following the subject proclitic. They serve various semantic functions, including subject focus (McClay 2017: 76–86) and contributing temporal or aspectual information. For a more elaborate description, see Dryer (2002).

Syntactically, they are a verbal category. A precise characterization of their syntactic position, and whether they are best analyzed as adjuncts or

instances of high verbal functional projections, is uncertain. However, what is crucial for the purposes of this paper is that they are restricted to verbal environments, and thus serve as a useful diagnostic for verbal material. Consider the forms below, which are the nominalizations discussed in (10) that allow either the subject proclitic or the nominal possessive strategy for expression possession.

(17) Covariation of possession strategy and preverb compatibility

a. Tanwaɬ huɬpaɬni **k isiɬ aɬmaɬumas fšans**

Tanwaɬ	huɬpaɬ-ni	k	isiɬ	aɬmaɬuma-s	fšans
Daniel	hear-IND	NMLZ	very	deep.voice-OBV	John

'Daniel heard John's very deep voice'

b. * Tanwaɬ huɬpaɬni **k isiɬ aɬmaɬumaʔis fšan**

Tanwaɬ	huɬpaɬ-ni	k	isiɬ	aɬmaɬuma-ʔis	fšan
Daniel	hear-IND	NMLZ	very	deep.voice-3.POSS	John

Intended: 'Daniel heard John's very deep voice'

When the possessor is expressed as the subject proclitic, which is null in (17a) but co-occuring with the overt DP *fšan* 'John', the result is grammatical. When the possessor is expressed through the nominal possessive morpheme (17b), the result is ungrammatical. This is predicted by the analysis above. If preverbs are high clausal adjuncts, at least higher than vP, then the lack of this projection in (17b) precludes preverbs by cutting off the clausal spine before they are licenced.[8]

4. Conclusion

I have shown that Ktunaxa deverbal *k*-nominalizations fall into two classes based on syntactic evidence from the realization of possessors, the presence of passives and other verbal morphology on the embedded verb, and the presence of verbal modifiers. These classes of deverbal nominalizations appear with a cluster of properties that suggest the presence of an external argument introducing a vP projection. Possessors introduced within this projection surface with the same morphology as verbal subjects. The passive is another instance of v^0 and forms "unpossessed" vP nominalizations – these

---

8   In the above examples (16), the obviative morpheme appears. While this suggests that the nominalization takes place much higher, at IP (see Bertrand 2019 for the obviative morpheme as I^0), I will leave this issue for further research as I lack the necessary data beyond the obviative morpheme, as well as lacking an account of the obviation system itself. An alternative analysis is that the form in (16a) is a subordinate clause. What is crucial for our purposes is the presence or absence of vP, which the preverbs allow us to determine.

are analyzed as being possessed by an indefinite third-person. Finally, additional argument-introducing verbal morphemes and verbal modifiers in nominalizations co-occur with the vP-level property of expressing possession with the verbal subject proclitic.

These properties have not been discussed in the description on Ktunaxa's nominalization system in Morgan (1991). The data and discussion presented here forms the first steps toward a more detailed picture of the Ktunaxa's ubiquitous k-forms. This emerging picture from deverbal nominalizations is that the k morpheme serves only to nominalize its complement. Under Wiltschko's (2014) typology of recategorization strategies, Ktunaxa's k is the realization of the nominalization via complementation head n^0, which takes various functional projections (here, VP and vP) as complements. The syntactic properties of these nominalizations are derived entirely, then, from the internal structure of n's complement.

An important remaining question is how other k-forms (presented in (1)) are to be analyzed under such a proposal. If k-forms are all nPs, then how are we to treat questions, subordinate clauses, and relative clauses? In her work on the verbal functional projections of Ktunaxa, Bertrand (2019) analyzes k as a realization of C^0 due to its role in subordination and its complementary distribution with respect to the indicative marker. However, an analysis of these constructions as fundamentally nominalizations, taking seriously the fact that k is present in all these constructions, could provide a unified analysis of these various constructions, as well as contribute to the large body of literature on the relationship between nominalizations and clausal constructions (see Introduction and papers in Zariquiey, Shibatani, & Fleck 2019, Comrie & Estrada-Fernández 2012). The view from deverbal nominalizations provides the first look at the possible complements of the Ktunaxa nominalizer and the properties these constructions yield and will form a foundation for investigating the structure of k-forms more broadly.

References

Abney, Steven. 1987. *The English noun phrase in its sentential aspect*. Cambridge, MA: MIT dissertation.

Baker, M. (1985). The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* 16(3): 373–415.

Bertrand, Anne. 2019. Drawing a map to the Ktunaxa clause: evidence from verbal morphology. In Emily Sadlier-Brown & Avery Ozburn (eds.), *UBc Qualifying Papers 4* (2015–2017), 1–20. Vancouver, BC: University of British Columbia Working Papers in Linguistics.

Blamire, Emily. 2011. The ordering of preverb strings in Ktunaxa. In John Lyon & Joel Dunham (eds.), *Proceedings of ICSNL XLVI*, 32–42. Vancouver, BC: University of British Columbia Working Papers in Linguistics.

Bliss, Heather. 2014. Assigning reference in clausal nominalizations. In Ileana Paul (ed.), *Cross-linguistic Investigations of Nominalization Patterns*, 85–117 Amsterdam: Benjamins.

Boas, Franz. 1926. Additional Notes on the Kutenai Language. *International Journal of American Linguistics* 4(1). 85–104.

Breuning, Daniel. 2013. By phrases in passives and nominals. *Syntax* 16. 1–41.

Burton, Strang & Lisa Matthewson. 2015. Targeted Construction Storyboards in Semantic Fieldwork. In Ryan Bochnak and Lisa Matthewson (eds.), *Methodologies in Semantic Fieldwork*, 135–156. Oxford: Oxford University Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Comrie, Bernard & Zarina Estrada-Fernández (eds.). 2012. *Relative clauses in the languages of the Americas: A typological overview*. Amsterdam: Benjamins.

Dékány, Éva. 2018. Approaches to head movement: A critical assessment. *Glossa: A Journal of General Linguistics* 3(1). 65.

Dryer, Matthew. 1991. Subject and inverse in Kutenai. In J. E. Redden (ed.), *Papers from the American Indian Languages Conference, held at the University of Santa Cruz, July and August, 1991. Occational Papers in Linguistics* 16, 183–202. Carbondale, IL: Dept. of Linguistics.

Dryer, Matthew. 1992. A comparison of obviation systems of Kutenai and Algonquian. In William Cowan (ed.), *Papers of the 23rd Algonquian Conference*, 119–163. Ottawa, ON: Carleton University.

Dryer, Matthew. 1994. The discourse function of the Kutenai inverse. In Talmy Givón (ed.), *Voice and Inverstion*, 65–99. Amsterdam: Benjamins.

Dryer, Matthew. 2002. A comparison of preverbs in Kutenai and Algonquian. In David Pentland (ed.), *Proceedings of the 30th Algonquian Conference*, 63–94. Winnipeg, MB: University of Manitoba.

Dunlop, Britt, Suzanne Gessner, Tracy Herbert, & Aliana Parker. 2018. *Report on the status of B.C. First Nations languages*. 3rd edn. Brentwood Bay, BC: First People's Cultural Council.

Garvin, Paul L. 1951a. Kutenai I: Phonemics. *International Journal of American Linguistics* 14(1). 37–42.

Garvin, Paul L. 1951b. Kutenai II: Morpheme Variations. *International Journal of American Linguistics* 14(2). 87–90.

Garvin, Paul L. 1951c. Kutenai III: Morpheme Distributions (Prefix, Theme, Suffix). *International Journal of American Linguistics* 14(3). 171–187.

Garvin, Paul L. 1951d. Kutenai IV: Word Classes. *International Journal of American Linguistics* 17(2). 84–97.

Harley, Heidi. 2017. The "bundling" hypothesis and the disparate functions of little v. In Roberta D'Alessandro, Irene Franco, & Angél J. Gallego (eds.), *The Verbal Domain*, 3–28. Oxford: Oxford University Press.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck & Laurie Zaring (eds), *Phrase Structure and the Lexicon*, 109–137. Dordrecht: Kluwer.

Laturnus, Rebecca. 2011. Future expressions in Ktunaxa. In John Lyon & Joel Dunham (eds.), *Proceedings of ICSNL XLVI*, 157–163. Vancouver, BC: University of British Columbia Working Papers in Linguistics.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4: 201–225.

Mast, Susan. 1988. *Aspects of Kutenai Morphology*. Pittsburgh, PA: University of Pittsburgh MA thesis.

Morgan, Lawrence. 1991. *A description of the Kutenai language*. Berkeley, CA: University of California dissertation.

McClay, Elise Kedersha. 2017. *Focus in Ktunaxa*. Vancouver, BC: University of British Columbia MA thesis.

McClay, Elise & Violet Birdstone. 2015 Evidence for question formation by direct wh-movement in Ktunaxa. In Natalie Weber, Erin Guntly, Zoe Lam, & Sihwei Chan (eds.), *Papers for ICSNL 50*, 89–102. Vancouver, BC: University of British Columbia Working Papers in Linguistics.

Pylkkanen, Liina. 2008. *Introducing Arguments*. Cambridge, MA: The MIT Press.

Ritter, Elizabeth. 2014. Nominalizing Inner Aspect: Evidence from Blackfoot. In Ileana Paul (ed.), *Cross-linguistic Investigations of Nominalization Patterns*, 25–50 Amsterdam: Benjamins.

Tammpere, Laura, Violet Birdstone & Martina Wiltschko. 2012. Independent pronouns in Ktunaxa. In Joel Dunham, John Lyon, & Natalie Weber (eds.), *Proceedings of ICSNL XLVII*, 325–339. Vancouver, BC: University of British Columbia Working Papers in Linguistics.

Williams, Edwin. 1981. Argument structure and morphology. *The Linguistic Review* 1: 81–114.

Wiltschko, Martina. 2014. Patterns of Blackfoot nominalizations. In Ileana Paul (ed.), *Cross-linguistic Investigations of Nominalization Patterns*, 189–214. Amsterdam: Benjamins.

Zariquiey, Roberto, Shibatani, Masayoshi & Fleck, David (eds.). 2019. *Nominalization in the languages of the Americas*. Amsterdam: Benjamins.

Terrance Gatchalian
University of British Columbia
tgatchalian@alumni.ubc.ca

Rossella Varvara

Constraints on nominalizations: Investigating the productivity domain of Italian *-mento* and *-zione*

Abstract: The paper investigates the different productivity domains (Rainer 2005) of two Italian event denoting suffixes, *-mento* and *-zione*. These suffixes share the same eventive semantics, they are both productive and thus can be seen as rivals in the formation of event nominalizations. The aim is to obtain a better understanding of the constraints that play a role in the selection of one affix over the other. By means of a logistic regression model the contribution of different features of the base verb is investigated. The analysis is conducted on a dataset of 678 nominalizations extracted from a section of Midia, a diachronic balanced corpus explicitly built for morphological research (Gaeta 2017). Results show that the frequency, the inflectional class and the number of characters of the base verb as well as the presence of the prefix *a-* significantly contribute to the definition of the different domains, only partially confirming previous findings.

Keywords: constraints on productivity, competition, derivational morphology, action nouns, deverbal nominalizations, corpus linguistics, quantitative analysis

1. Introduction

Competition among affixes (also known as *rivalry*) is a common phenomenon in morphology (see a.o. Rainer et al. 2019), both in inflection and word-formation. It can be described as the availability of multiple patterns to express a certain concept. Research on this topic usually focuses on understanding which form is preferred by a speaker or by a speech community and what are the reasons underlying this choice.

The competing patterns may differ in their degree of productivity[1]: one could be more available to form neologisms in the present-day language than the other. Or they could be productive in different domains, i.e. different subsets of words to which the pattern applies (Rainer 2005). The pattern's domain can be defined through the features (called *constraints* or *restrictions*) that a potential base should possess. An example comes from English nominalizations in *-ation*: contrary to the *-al, -ance, -ment* or *-ure* nominalizations, they apply to suffixed verbs in *-ize* and *-ify* (e.g. *adultification, aristocratization,* Plag 2003: 63, Bauer et al. 2013: 196–202).

---

1   For an overview of the notion and measures of *productivity* see Bauer (2001, 2005) and Baayen (2009).

Constraints can be of different nature, as they can concern phonological, morphological or semantic aspects of the base. An example of a morphological condition has just been described above for the English suffix -*ation*, whereas we can cite as a case of phonological constraint the preference of the English suffix -*eer* for bases ending in [t] (e.g. *musketeer*, *profiteer*, *racketeer*, Adams 1973: 175-178). A syntactic restriction is at play in the preference of the suffix -*able* for transitive base verbs: *visitable* vs *goable, *observable* vs *lookable (Rainer 2005: 348). At the semantic level, the Spanish relational suffix -*uno* is mostly attached to base nouns referring to animals (e.g. *vaca* 'cow', *vacuno* 'relating to cow'). Among others, Rainer (2005) and Gaeta (2015) offer overviews of the different types of constraints a pattern may present, with further examples from multiple languages.

The present paper focuses on the differences in the productivity domain of two Italian competing patterns, i.e. nominalizations in -*mento* and -*zione*. They belong to a specific class of nominalizations, here called *event-denoting deverbal nominalizations* (henceforth EDN). The term nominalization indicates both the process and the result of "turning something into a noun" (Comrie and Thompson 2007: 334), but in this context we restrict our analysis to cases in which the base of the process is a verb and the resulting nominalization refers to an event (in the broadest sense[2]).

In many languages, more than one affix is available to form an EDN. In English, for example, the suffixes -*al* (*arrival, approval*), -*ance* (*resistance, attendance*), -*ing* (*reading, learning*), -*ation* (*regulation, consultation*), and -*ment* (*recruitment, development*) can all be used to form event nominalizations. These can be seen as constituting a single paradigmatic cell of semantic derivation (Booij and Lieber 2004). In Italian, the language under investigation in this work, multiple suffixes are available to exploit this function as well: -*zione* (*venerazione* 'veneration'), -*mento* (*annegamento* 'drowning'), -*tura* (*spuntatura* 'trim'), -*aggio* (*smontaggio* 'dismanteling'), -*ata* (*sbirciata* 'peek'), -*nza* (*permanenza* 'permanence, stay'). Moreover, event nouns may be formed also by means of conversion (or *zero derivation*): *aumento* ('increase'), *viaggio* ('trip'). Among all these patterns, the -*mento* and -*zione* have been selected for this first investigation since they are the most productive. The study aims at understanding if some morphological properties of the base verbs are relevant in the selection of one affix instead of the other in the formation of EDNs.

---

2   With the term "event" I refer to every kind of eventuality (Bach 1986), including states. Thus, event nominalizations may denote activities, achievements, accomplishments and states (following the terminology proposed in Vendler 1957). The same class of derived nominals has been frequently called *action nouns* (or *nomina actionis,* Comrie 1976; Comrie and Thompson 2007; Koptjevskaja-Tamm 1993, 2006).

The analysis is conducted by considering all the formations attested in a corpus from a specific period of time (from 1841 to 1947), thus investigating the constraints on *realized productivity* (Baayen 2009), rather than on *expanding* or *potential productivity*. The realized productivity (also known as *extent of use*) measures the number of complex words a morphological process produced in the past. Conversely, expanding and potential productivity are seen as measures of the expansion of the class in the near future, i.e. how much the morphological processes are expected to be used to form neologisms. In this study, the analysis is conducted on "past" formations, but future work can investigate the productivity constraints considering only neologisms in the dataset.

The paper is structured as follows. In the next section (§2), I introduce previous findings on the differential constraints of *-mento* and *-zione*. In section 3, I present the methodology applied, i.e. regression modelling based on corpus data, and the data sampling. In section 4, I list the variables considered as predictors and provide their descriptive statistics. In section 5, the regression modelling and its results are presented. Section 6 offers a discussion of the main findings. Section 7 draws conclusions and directions of future research.

2. Previous works on Italian *-mento* and *-zione*

Numerous works have focused on assessing the productivity degree of the suffixes *-mento* and *-zione* (Thornton 1988; Iacobini & Thornton 1992; Gaeta & Ricca 2002; Fiorentino 2008; Štichauer 2009; Varvara 2019), and they frequently considered also other EDNs. Similarly, the problem of the stem form to which they attach has often been discussed (Scalise 1983; Thornton 1990–1991, 2015; Gaeta 2004). The issue of the competition between these two Italian suffixes has been addressed by Scalise (1983: 207–208), Melloni (2007: 70–71), and in more depth by Gaeta (2002 2004: 327 and ff). They propose numerous constraints on the productivity of *-zione* and *-mento* suffixes. For an easier description, we can divide them in phonological, morphological and semantic constraints.

2.1. Phonological constraints

Gaeta (2004) observes that *-zione* is used to derive action nouns for some of the few Italian monosyllabic verbs (ex. *dizione* 'diction', from *dire* 'to say', *stazione* 'station', from *stare* 'to stay', *dazione* 'dation' from *dare* 'to give'), whereas *-mento* attaches to bases that are at least bisyllabic. Second, he notices a euphonic restriction for *-zione*, which usually does not follow bases that end in /ts/ plus a vowel (e.g. **deprezzazione* from the base verb *deprezzare* 'depreciate').

2.2. Morphological constraints

Previous works do not recognize a specific association with the inflectional class of the base verb; however, parasynthetic verbs from the third conjugation seem to prefer *-mento* (*impigrimento* 'the act of becoming lazy'). Moreover, previous analyses note a slight preference of simple base verbs for *-mento* (e.g. *biascicamento* 'munching'), whereas base verbs formed by conversion prefer *-zione* (e.g. *datazione* 'dating').

Various associations with base verbs that present some specific affixes are listed (Gaeta 2004: 327–331). First, they highlight an association of the suffix *-mento* with parasynthetic verbs formed with the prefixes *ad-* and *in-* (*ammanettamento* 'handcuffing', *inacidimento* 'souring'). Second, prefixed verbs with *s-* are more correlated with *-mento*. More specifically, this prediction is linked to the two meanings that the prefix *s-* can bring in Italian: a negative value (e.g. *fiorire* 'to bloom' vs *sfiorire* 'to wither') and an intensifier one (e.g. *gridare* 'to shout' vs *sgridare* 'to scold'). Nouns in *-mento* are formed from prefixed bases with either of these two meanings, whereas *-zione* attaches mainly to bases in which the prefix *s-* has a negative value. Thus, the verbs with the intensifier *s-* are associated with *-mento* derivatives. Third, the suffix *-zione* is more frequently associated with prefixed verbs with *e(s)-* (e.g. *eruzione* 'eruption') or *de-* (e.g. *decomposizione* 'decomposition'), no matter if the bases are parasynthetic verbs or simple prefixed ones. Fourth, suffixed verbs with *-ific-* and *-izz-* seem to be more frequently associated with derivatives in *-zione (*e.g. *laicizzazione* 'secularization', from *laicizzare* 'to secularize', and *nazificazione* 'nazification', from *nazificare*), even if some derivatives in *-mento* are attested (*volgarizzamento* 'translation into vernacular'). Note, however, that this constraint seems in contradiction with the euphonic restriction we have listed in 2.1., i.e. *-zione* derivatives tend to avoid base verbs ending in /ts/ plus a vowel. Fifth, suffixed verbs in *-eggi-, -acchi-/-ucchi-, -(er)ell-, -ett-, -icch-* prefer *-mento* to form nominalizations (*fronteggiamento* 'confrontation', *saltellamento* 'hopping', *scoppiettamento* 'crackling', *mordicchiamento* 'nibbling'). Lastly, bases with the suffix *-iv-* select the *-zione* suffix (*attivazione* 'activation').

2.3. Semantic constraints

Gaeta (2002: 215 and ff) also identifies a difference in the semantics of the resulting nominalizations. He draws two conclusions: 1- *-mento* derivatives show the simple derivational meaning of 'the act of V' (where V is the base verb) more frequently compared to those in *-zione* (which have thus a higher degree of polysemy); 2- among all the other possible readings, *-zione* shows a high number of derivatives with an additional resultative meaning of 'what has been V-ed'.

3. Methodology

Even if the observations made by Gaeta (2004) were based on quantitative data extracted from the DISC online dictionary and from one year of the *La Stampa* newspaper, the strength of these associations was not assessed, since no statistical test was conducted. As noted by Bonami and Thuilier (2019: 6) "descriptive statistics does not allow one to determine whether the tendencies observed in a sample are robust enough that one can exclude their being due to chance: neither do they allow one to conclude on the relative role of highly correlated properties of the base, such as phonological and morphological characterizations of their shape". For these reasons, raw counts of occurrences in a corpus are not enough to assess the correlation among phenomena; moreover, since there are multiple possible factors at play, a multivariate statistical model is more suitable for this kind of investigation than single monofactorial tests.

Statistical approaches are nowadays widespread in linguistics, and also specifically in the study of the rivalry between affixes. Arndt-Lappe (2014) applies an analogical model to investigate the rivalry between the English suffix *-ity* and *-ness*. Her findings show that the model can predict the preference patterns by the phonological characteristics of the two base-final syllables and by the syntactic category of the base. Varvara (2017) focuses on the competition between Italian nominal infinitives and the whole class of event-denoting nominalization suffixes, by applying a regression analysis to evaluate possible constraints. Bonami and Thuiller (2019) focus on the French suffixes *-iser* and *-ifier*, and they highlight how multiple factors may play a role at the same time.

Similarly to the latter two works, the present study applies logistic regression modelling, inspecting the domain of application of the Italian suffixes *-mento* and *-zione*. In binary logistic regression, the model estimates the probability of a predicted event whose outcome is binary (0 or 1). In our case, the predicted event is the suffix used to form a nominalization from a base verb. This is our dependent variable (also called response) and it has the binary outcome *-mento* vs *-zione*. Given the data observed, the model will assess the role of different predictors (or independent variables, i.e. the constraints investigated) in the selection of the suffix used.

The analysis is conducted on data extracted from the MIDIA corpus[3], a diachronic corpus of 7,5 million tokens. Even if the corpus size is quite small, it has the advantage of being balanced through genres. For the present work, only the period[4] from 1841 to 1947 has been considered as a

---

3   www.corpusmidia.unito.it
4   The subcorpus selected contains 1.667.928 tokens.

first step, but future work will extend the analysis to the whole corpus and compare the results diachronically. From the texts available for this time span, all the occurrences of words in *-mento* and *-zione* have been automatically extracted. As a second step, a manual check was done to remove all the typos and, following a procedure similar to that adopted by Gaeta and Ricca (2002: 233–237), to also remove:

1- simple, not morphologically complex nouns that accidentally end in *-mento* or *-zione* (e.g. *elemento* 'element', *cemento* 'cement');

2- opaque derivations whose semantic relation with the bases is no more transparent (e.g. *stazione* 'station', derived from *stare* 'to stay');

3- denominal nouns (e.g. *tunnellamento*), since the aim of the study is to assess the relevant properties of the base verbs in the suffix selection.

The resulting dataset consisted of 678 items, 249 nouns in *-mento* and 429 nouns in *-zione*. Each lemma was thus annotated with a set of 7 features, i.e. possible constraints on pattern productivity. These variables are listed in the next section.

4. Predictors and descriptive statistics

In this section I describe each variable and present some descriptive statistics. The set of variables considered as predictors are:

• Frequency of the base (continuous variable);
• Frequency of the derived term (continuous variable);
• Ratio of the derivative frequency to the base frequency (continuous variable);
• Length in characters of the base verb (continuous variable with values from 6 to 15);
• Inflectional class of the base verb (categorical, with three levels: *-are*, *-ere*, *-ire)*;
• Number of total derivational processes of the derivative (continuous variable with values from 1 to 4);
• Other affixes present on the base verb (categorical, with 14 levels).

4.1. Control variables

The first variable considered is the frequency of the nominalization in the corpus MIDIA. The frequency distributions of the two categories are slightly different, with *-zione* derivatives in general more frequent than *-mento* ones. Tab. 1 reports their token frequency in the corpus (providing minimum and maximum values, median and mean).

Tab. 1: Frequency distributions for -mento and -zione nominals.

	Min.	Median	Mean	Max
-mento	1	4	17.58	409
-zione	1	7	19.64	422

In addition, I take into consideration the frequency of the corresponding base verb. The base verbs for the two groups show a similar frequency distribution, as reported in Tab. 2. In 27 cases, both a *-mento* and a *-zione* derivative was attested for the same verb.

Tab. 2: Frequency distributions of corresponding base verbs.

	Min.	Median	Mean	Max
-mento	1	74	76.2	152
-zione	1	71	70.41	151

Furthermore, the ratio of the EDN frequency to the base frequency (known as *relative frequency*) is considered. Previous work has indeed highlighted that relative frequency is more correlated to morphological processing, productivity and semantic transparency, compared to absolute frequency (see Hay 2001; Baayen 2009). Higher relative frequency seems to be related to faster processing, higher semantic transparency, and higher productivity. Tab. 3 reports the values for the two EDN categories: nominalizations in *-zione* have higher relative frequency than those in *-mento*; on average, the frequency of *-zione* EDNs is equal to the frequency of the corresponding base, whereas the *-mento* EDNs have on average a lower frequency than that of the base.

Tab. 3: Frequency distribution of ratio of EDN frequency to base verb frequency.

	Min.	Median	Mean	Max
-mento	0.0066	0.0758	0.5245	26.2500
-zione	0.0068	0.1467	1.0870	52.3333

4.2. Length in characters of the base verb

Previous accounts (Gaeta 2004; Melloni 2007) found a correlation between monosyllabic bases and *-zione* derivatives. For this reason, the length of the base verb (in the form of the infinitive[5]) is considered as a predictor. In our sample, verbs vary from a minimum of 6 characters to a maximum of 15 (median=9, mean[*-mento*]= 8.976, mean[*-zione*]= 9.177). A more detailed overview is given in Table 4.

---

5   The length is computed in terms of number of characters. However, further analysis may consider the number of phonemes.

Tab. 4: Number of characters of corresponding base verbs.

	Number of characters of the base verb									
	6	7	8	9	10	11	12	13	14	15
-mento	11	30	48	74	49	28	8	0	0	1
-zione	13	45	87	114	86	57	18	7	2	0

4.3. Inflectional class of the base verb

Italian verbs are classified in three main conjugations, depending on the infinitive ending: first conjugation in *-are* (e.g. *mangiare* 'to eat'), second conjugation in *-ere* (*bere* 'to drink'), third conjugation in *-ire* (*sentire* 'to hear'). The distribution of these three classes is reported in Table 5, together with expected values[6]. Previous works (§2) have noticed a relation between the *-mento/-zione* rivalry and the conjugation of the base verb. Specifically, it has been argued that parasynthetic base verbs from the third conjugation more frequently derive EDNs in *-mento*. If we compare observed and expected values, we note that *-mento* derivatives from the third conjugation are higher than expected. This may be due specifically to parasynthetic verbs or to the whole third conjugation. Moreover, the suffix *-mento* associates more frequently than expected with verbs from the second conjugation too, whereas *-zione* associates more with the first conjugation. We will test the significance of these correlations in section 5.

Tab. 5: Inflectional class of base verbs.

	Observed values		Expected values	
	-mento	-zione	-mento	-zione
-are	144	350	181.42	312.58
-ere	46	33	29.01	49.99
-ire	59	46	38.56	66.44

4.4. Morphological complexity

The presence of other morphological processes was considered in two ways: first, by computing the total number of morphological processes; second, by considering the specific affix present. This information was taken, whenever possible, from Derivatario[7] (Talamo, Celata, and Bertinetto 2016), a freely available digital lexicon of morphologically complex Italian words. When the attested lemma was not available in this resource, the annotation was carried out by hand.

---

6   Expected values are computed by means of a chi-squared test.

7   http://derivatario.sns.it

4.4.1. Total number of morphological processes

The total number of morphological processes attested in our sample ranges from 1 to 4 processes. Tab. 6 reports the distribution of EDNs (type frequency) for each number of processes observed. For example, there are 58 nominalizations in *-mento* that show two morphological processes; *-zione* EDNs are instead 102 for this level. One morphological process indicates that the derivative shows only the nominalization process, i.e. it does not contain any other derivational affix besides *-mento* or *-zione*. This is the case of *fondamento* ('foundation'), which is directly derived from the base verb *fondare* ('to found'). The word *armonizzazione* ('harmonization') has undergone two derivational processes: first the formation of the denominal verb *armonizzare* ('to harmonize'), derived from the noun *armonia* ('harmony') by means of the suffix *-izz-*; then, the denominal verb transformation into a deverbal noun by means of the nominalizing suffix *-zione*. An example of a derivative with three derivational processes is *immatricolazione* ('enrolment'): starting from the noun *matricola* ('freshman'), the verb is formed by parasynthesis[8] (e.g. conversion combined with the inchoative prefix *in-*); then, the noun is derived with the suffix *-zione*. For our sample, the maximum number of derivational processes is 4. The word *ristabilimento* ('reinstatement') is an example where the verbal base *stare* ('to stay') is turned into an adjective by means of *-bile*, which is then converted into a verb (*stabilire* 'to establish'), modified by the iterative prefix *re-*, and lastly turned into an EDN by the suffix *-mento*. Table 6 reports additionally the expected values of EDNs in each category (computed by means of a chi-squared test). As can be seen, *-zione* derivatives are more frequent than expected when no other derivational process is present, whereas *-mento* nominalizations occur more than expected when 3 or 4 derivational processes are present. The significance of this difference will be tested in the regression model.

Tab. 6: Number of morphological processes in EDNs.

	Observed values		Expected values	
	-mento	-zione	-mento	-zione
1	127	273	146.90	253.10
2	58	102	58.76	101.24
3	56	54	40.40	69.60
4	8	0	2.94	5.06

---

8   Parasynthesis indeed counts as two derivational processes in the total count.

4.4.2. *Presence of other affixes*

The last variable taken into consideration is the nature of other affixes (whenever present). Tab. 7 reports the different affixes attested in the sample with the type frequency of the two nominalization patterns. Following Talamo, Celata, and Bertinetto (2016), some affixes are split in two groups based on their semantics: e.g. 1de- indicates the prefix *de-* when occurring with a reversative reading (like in *detassare* 'to untax'); 2de- the same prefix with a causative meaning (*depurare* 'to purify'). With the label *a-* I refer to every prefix formed by the sequence *a* plus a consonant (e.g. *abbassare* 'to lower' *avvicinare*, 'to place near'). In the rest of the paper I will call this whole class *prefix a-*. Some affixes are attested only with one nominalization, but the values in some cases are really low and it would not be possible to generalize on these few occurrences. As it will be explained in the next section, in the regression modelling the affixes that have zero formations with one EDN will be aggregated in one level called *other affixes*. Sampling zeros would cause indeed infinite estimates.

Tab. 7: Presence of other affixes.

Affix	Observed values		Affix	Observed values	
	-mento	-zione		-mento	-zione
1 de-	2	4	-eggiare	3	0
1 in-	0	1	-ificare	0	22
1 s-	3	1	-izzare	0	11
2 de-	3	8	-nte	0	1
2 in-	15	14	No other affix	147	310
2 s-	7	1	pre-	0	3
a-	39	12	pro-	3	2
-bile	2	0	ri-	16	8
co-	1	5	trans-	1	3
con-	2	9	-zione	1	0
dis-	3	7			

5. Multifactorial statistical analysis

In summary, we have 7 independent variables in our statistical analysis. Two of them are categorical variables, whereas the other 5 are continuous variables. The response of our model is a binary variable 0/1, where 0 corresponds to the suffix *-mento* and 1 to the suffix *-zione*. All the frequency variables were log-transformed and scaled on their mean. The analyses are performed with the software R (R Core Team 2015), and by means of the *glm* function (with family type equal to binomial).

In order to determine the model that best fits our observations and to keep only significant variables, I proceed with backward selection of the variables, using the *step* function. Given the most complex model (with all the variables together), this function compares it to all the possible alternative models removing one variable at a time and evaluates the best-fitting model based on likelihood ratio tests and AIC[9] values (i.e., the lower, the better). It tries to take into account as much variance as accounted for by the complex model, while removing predictors that do not contribute to the regression equation.

In Tab. 8, the result of this procedure is summarized: in the first row the starting complex model is defined, whereas in the second row the final model is reported.

Tab. 8: Model selection.

	Model	Deviance	AIC
full model	Freq. EDN + Freq. Base + Rel. freq. + Conjugation + Affixes + Num. of deriv. processes + Base length	732.46	774.46
best model	Freq. EDN + Conjugation + Affixes + Num. of deriv. processes + Base length	733.50	771.5

As shown, the frequency of the base and the relative frequency were not improving the model significantly and were removed. Thus, the predictors of the final model are thus the frequency of the derived word, the inflection class of the base verb, the other affixes present on the base, the total number of derivational processes and the length in characters of the base verb. The probability of deriving a *-zione* derivates is computed based on these variables.

Note, however, that from the original dataset I removed 8 levels of the factor *Other affixes* (specifically 1in-, -bile, -eggiare, -ificare, -izzare, -nte, pre-, -zione), because they presented *sampling zeros* (Agresti 2003: 138). As can be noted in Tab. 7, in our sample these affixes occur only with one of the two nominalizations, and thus show a value of zero in the other cell. Sampling zeros may cause computation problems in the model (with infinite estimates occurring for that level). For this reason, I aggregated these 8 levels into one called *others*. We cannot know if these zeros indicate a true correlation or are either due to sparse data. We can only note that, given our sample and the values reported in Tab. 7, base verbs formed with the suffixes *-eggiare*, *-zione* and *-bile* show only nominalizations in *-mento*,

---

9   Akaike Information Criterion. It is a badness of fit indice and for this reason we pick the model with the lowest index.

whereas base verbs formed with the affixes *1in-, pre-, -ificare, -izzare* and *-nte* present only nominalizations in *-zione*. Further evidence from a larger corpus is needed to confirm these tendencies.

The summary of the final model is reported in Tab. 9. The first column reports the predictors of the model, with a row for each level of the categorical factors. In the second column, the sign of the estimated coefficient[10] indicates the direction of the effect: a positive estimate indicates an association between the factor and the nominalization in *-zione*; a negative one an association with *-mento*. Column 5 shows the significance of a predictor (i.e., its p-value), which indicates how much it contributes to the distinction. The significance level of the p-value is set to < 0.05.

Tab. 9: Summary of the final model.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.372	1.211	1.134	0.256889	
scale(log(Fq_EDN))	0.427	0.098	4.364	1.28e-05	***
conjugation: ere	-1.695	0.291	-5.818	5.97e-09	***
conjugation: ire	-1.421	0.252	-5.631	1.79e-08	***
affix: others	0.272	0.982	0.277	0.781851	
affix: 1S	-1.691	1.546	-1.094	0.274038	
affix: 2DE	0.630	1.133	0.556	0.578213	
affix: 2IN	-0.506	0.970	-0.522	0.601763	
affix: 2S	-2.042	1.392	-1.467	0.142301	
affix: A	-1.908	0.948	-2.013	0.044111	*
affix: CO	0.686	1.446	0.475	0.635039	
affix: CON	0.932	1.184	0.787	0.431039	
affix: DIS	-0.125	1.143	-0.110	0.912667	
affix: No affix	-0.805	0.947	-0.850	0.395277	
affix: PRO	-0.753	1.300	-0.579	0.562748	
affix: RI	-1.215	1.002	-1.213	0.225253	
affix: TRANS	0.315	1.502	0.210	0.833783	
number of derivational processes	-0.728	0.219	-3.327	0.000877	***
length in characters of the base verb	0.172	0.070	2.441	0.014645	*

---

10 Estimates are expressed in log odds, an alternate way to express probability, whose values range from $-\infty$ to $+\infty$.

The frequency of the derived term, the inflectional class of the base verb and the number of its derivational processes are all significant predictors (p < 0.001) for the distinction between *-zione* and *-mento* nominalizations.

Specifically, a one-unit increase in frequency is associated with an increase in the log odds of the derivative being a *-zione* nominalization in the amount of 0.427. This finding confirms the trend we already noted in Tab. 1 (p.84) as significantly supported by the statistical data analysis. Nominalizations in *-zione* are more frequently used than those in *-mento*, probably because of their higher productivity.

5.1. Inflectional class of the base verb

With regard to the effect of the inflectional class, we observe that base verbs ending in *-ere* and *-ire* are both associated with the decrease in the log odds for *-zione* nominalizations. In other words, verbs from these conjugation classes tend to derive EDNs with the suffix *-mento*, rather than *-zione*; by contrast, verbs from the first conjugation are more associated with *-zione* nominalizations. Fig. 1 represents this effect[11]. This finding may be interpreted as a correlate of what has been previously observed about parasynthetic verbs in *-ire*, which form nominalizations in *-mento* (§4.3, p.85).
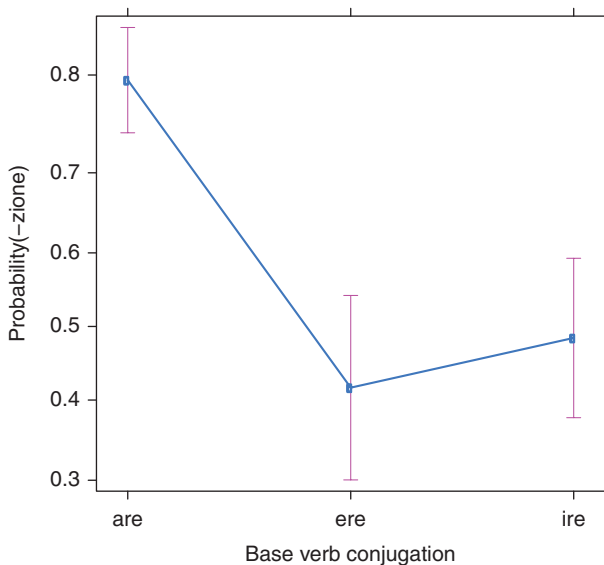
Fig. 1: Inflectional class effect plot.

---

11 Effect plots have been drawn using the R package *effect* (Fox et al. 2019).

Indeed, a quick analysis of verbs in *-ire* forming *-mento* derivatives reveals that 16 of them (out of 59) are parasynthetic verbs, whereas only one parasynthetic verb in *-ire* forms a *-zione* nominalization. However, the role of parasynthetic verbs may be marginal, considering a further association between *-mento* and the Italian second inflectional class of verbs (ending in *-ere*). There is probably something more linked to the base verb conjugation that drives the choice of the nominalizing suffix.

5.2. Total number of morphological processes and length of the base

The total number of morphological processes on the EDN is also significant: a higher number is linked to a decrease of log odds for *-zione* nominalizations. A higher number of morphological processes is thus associated with *-mento* derivatives, fewer processes are associated with *-zione* (Fig. 2). This finding contradicts previous works that stated a slight preference of simple base verbs for *-mento* EDNs.
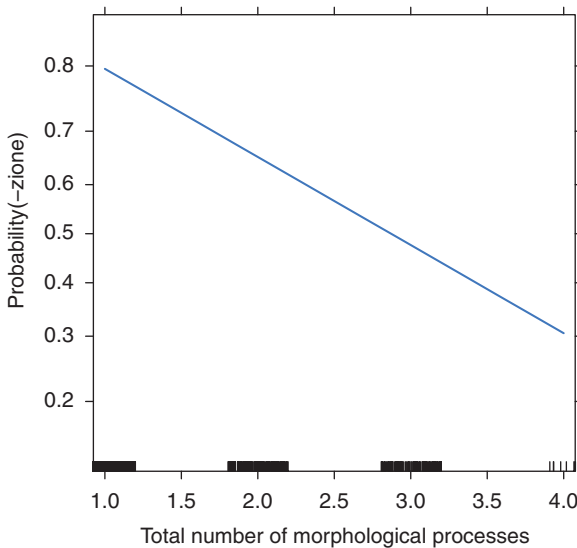
Fig. 2: Number of morphological processes effect.

The number of characters of the base verb is significant as well ($p < 0.05$), but its effect goes in the opposite direction with respect to the number of morphological processes: longer bases are associated with *-zione* EDN (Fig. 3). It is interesting to note that longer bases are thus not correlated with a higher number of derivational processes: *-zione* EDNs are formed from longer bases, but *-mento* EDNs present a higher number of morphological processes. This phenomenon may be linked to the treatment of parasynthetic verbs:
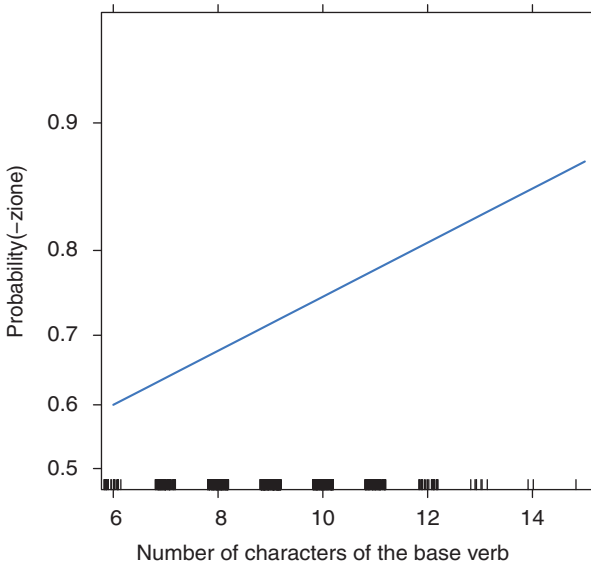
Fig. 3: Effect of the number of characters of the base.

parasynthesis, indeed, has been counted as two morphological processes and EDNs annotated as showing 4 morphological processing (mainly *-mento* EDNs) are indeed all derived from parasynthetic verbs. Nominalizations in *-mento* seem once more associated with this category of verbs.

Still, when considering the number of characters of the base, it should be noted that we do not have monosyllabic verbs (e.g. *dire* 'to say') in our sample and thus we cannot assess the hypothesis (expressed in previous works, see §2.1.) that *-zione* is preferred for these verbs. However, we note that our analysis reveals an opposite tendency, i.e. *-zione* EDNs are associated with longer bases.

5.3. Presence of other affixes

Lastly, the model in Tab. 9 shows that only one level of the variable OTHER AFFIX is significant ($p < 0.05$), i.e. the presence of the prefix *a-*. It reduces the log odds of having a *-zione* derivative, and thus increases those of *-mento* EDNs. Since no other affixes contributed to the model, I decided to simplify the model by considering only the presence (or absence) of the prefix *a-* as factor (instead of the whole list of affixes). The new factor (which I will call *Prefix A*) has three levels: presence (if the base verb shows this affix), no affix

(if the base verb is a simple non-derived verb), other affix (if the base verb has additional affixations besides the prefix *a-*). I then repeated the analysis with a stepwise selection of the significant variables (Tab. 10).

Tab. 10: Model selection considering only the prefix a- *as additional affix.*

	Model	Deviance	AIC
full model	Freq. EDN + Freq. Base + Rel. freq. + Conjugation + Prefix A + Num. of deriv. processes + Base length	750.15	770.15
best model	Freq. EDN + Conjugation + Prefix A + Num. of deriv. processes + Base length	751.29	767.29

The final model contains the same variable as the model from the previous setting (Tab. 8 and Tab. 9). The frequency of the base and the relative frequency were removed in the final model because they were not significantly improving the fit. The summary of this model is reported in Tab. 11.

Tab. 11: Summary of the final model considering only the prefix a- *as additional affix.*

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.780	0.871	-0.896	0.37049	
Freq. EDN	0.427	0.096	4.437	9.14e-06	***
conjugation -*ere*	-1.824	0.283	-6.448	1.13e-10	***
conjugation -*ire*	-1.504	0.247	-6.083	1.18e-09	***
Prefix A: different affix	1.637	0.395	4.150	3.32e-05	***
Prefix A: No affix	1.119	0.481	2.324	0.02012	*
number of deriv. processes	-0.753	0.206	-3.662	0.00025	***
Base length	0.206	0.067	3.055	0.00225	**

All the effects go in the same direction as in the previous setting; moreover, their significance increased, and the AIC lowered. The effect of the prefix *a-* is confirmed, as the plot in Fig. 4 shows.
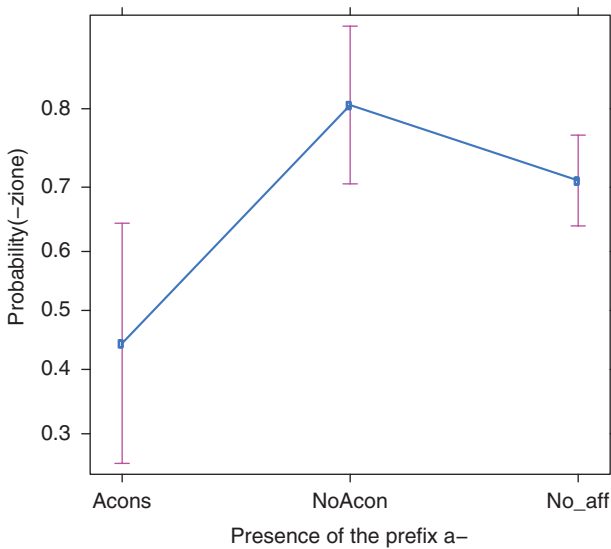
Fig. 4: Effect of the prefix a-.

6. Discussion

The present study contributes to the understanding of the intricate matter of the rivalry between two Italian event-denoting suffixes. The study has focused on the possible constraints that shape the domain of application, by investigating seven possible features. The multivariate statistical analysis has been shown to be useful in the evaluation of the contribution of the different features and it has only partially confirmed previous claims on the topic.

Specifically, the analysis confirms an association between prefixed verbs in *a-* and *-mento* nominalizations, as well as a relation between this suffix and verbs from the third conjugation. By considering these findings together, we can draw two considerations.

First, most of *a-* prefixed verbs (43 out of 51) belong to the first inflectional class. This is an interesting fact if we consider that our analysis also reveals an association between *-zione* EDNs and the first conjugation (and on the other side, an association between *-mento* and the second and third conjugations). This means that even if *-zione* derivatives show higher probabilities to be formed from verbs of the first conjugation, *-mento* nominalizations undermine this association when the verb shows the prefix *a-*.

Second, previous work stated that there is a relation between *-mento* and verbs from the third conjugation that were formed by parasynthesis specifically. Our analysis reveals an association between *-mento* and the whole category of third conjugation verbs, although a qualitative inspection indicates

not only that the majority of third conjugation verbs that form a *-mento* EDN are parasynthetic verbs, but also that prefixed verbs in *a-* from the first conjugation (forming *-mento* derivatives) involve parasynthesis. It is thus possible indeed that parasynthesis is the underlying constraint that drives the choice of *-mento* EDNs. Further work, on a new sample of data, should test this hypothesis by considering parasynthesis as an independent and codified factor in the analysis.

We also observed a relation between verbs from the second conjugation and the *-mento* suffix. The presence of parasynthetic verbs is marginal here, but further work should better assess whether parasynthesis or other factors specific to the conjugation may also play a role in this relation.

With regard to the presence of other affixes, the prefixes *de-, in-* and *s-* were not significant in our study, contrary to what was claimed in the past literature (e.g. Gaeta 2004). The contribution of some affixes could not be assessed due to the small sample and the presence of sampling zeros. However, simple type frequencies (Tab. 7) show that verbs ending in *-ificare* and *-izzare* form only *-zione* derivatives, as claimed in past works. The statistical significance of this association should be assessed on the basis of a larger sample. Indeed, considering the verbs in *-ificare*, we observe that they could form *-mento* EDNs in principle. For example, *purificamento* (from *purificare*, "to purify") is not attested in our sample but is listed in the Treccani online dictionary[12] and it counts around 1400 results in a google search. It is obviously less frequent than its *-zione* counterpart (*purificazione)*, which produces more than 2 million results on google, but further work should test this statistically.

The analysis reveals an association between the total number of morphological processes the derivative has undergone and the EDN suffix: *-zione* nominalizations are associated with fewer morphological processes, whereas *-mento* ones to a higher number. This result contradicts previous claims that mention a slight association between *-mento* and simple base verbs but is in line with what we have seen about parasynthesis: *-mento* derivatives tend to be related to parasynthetic verbs and, since parasynthesis is counted as two morphological processes (i.e. the prefix and the conversion process), it can be responsible for the higher number of processes in *-mento* derivatives. Interestingly, the effect of the number of morphological processes and the effect of verb length go in the opposite direction. Nominalizations in *-zione* have longer bases, even if they show fewer morphological processes.

---

12 http://www.treccani.it/vocabolario/purificamento/

Obviously, there are many other possible features that may constrain the productivity of the two nominalizations considered. I hope to address some in future work.s

7. Conclusion

In the present work, I investigated the role of base constraints in the pattern selection for event-denoting nominalizations. I focused on two suffixes available in Italian, namely *-mento* and *-zione,* and carried a statistical analysis on corpus data. I found that different morphological features of the base verbs (specifically, the length in characters, the inflectional class, and the presence and number of other affixations) influence the use of one pattern over the other. Nominalizations in *-zione* are associated with verbs from the first conjugation, but not if the base verb is prefixed in *a-*. In these cases, *-mento* is preferred. Nominalizations in *-mento* are also preferred for verbs from the second and third conjugations. Moreover, *-zione* base verbs are longer than those forming *-mento* EDN, even if *-mento* derivatives show a higher number of morphological processes. I hypothesize that all these findings may be linked to verbs formed by means of parasynthesis, but future work should test this interpretation.

The results here observed are only a first step towards solving this intricate puzzle, and further research is needed to overcome the limitations of this study. In future work, I intend to expand on the sample coverage, by using a larger corpus, and to employ further factors in the modelling, which may be relevant for this case of affix rivalry. Moreover, I would like to include a diachronic perspective, comparing results from different time periods, and to conduct an additional analysis restricted to neologisms[13], in order to evaluate constraints on *potential* rather than *realized* productivity. Lastly, even if the two suffixes considered are the most productive, future research should take into account other EDN patterns as well.

At a general level, the present study shows that productivity constraints do not represent strict rules with binary outcomes, but they rather emerge as preferences with a graded effect. For instance, stating that verbs from the second conjugation are more frequently associated with the suffix *-mento* is

---

13 A first attempt in this sense was already conducted: the original dataset was restricted considering only EDNs occurring once. Previous studies (e.g. Baayen and Renouf 1996, Plag 2003) suggested indeed that hapax legomena are a good approximation of neologisms. Results of the regression model on these data are in line with those reported above, with the difference of the disappearance of the effect for the length of the base verb. I do not report this analysis here because of the scarsity of data: there were indeed only 120 observations left.

different from arguing that they necessarily select this suffix. Indeed, many cases which would contradict such rigid arguments can be found even in our small sample. Discovering tendencies should not be seen as having a minor impact in research, since they enrich our comprehension of language and language processing. Word formation is a complex process and a complex explanation is what we would expect.

Acknowledgments

References

Adams, Valerie. 1973. *An Introduction to Modern English Word-Formation*. London: Longman.

Agresti, Alan. 2003. *Categorical data analysis*. Hoboken, NJ: John Wiley & Sons.

Arndt-Lappe, Sabine. 2014. Analogy in suffix rivalry: The case of English -ity and -ness. *English Language & Linguistics* 18(3). 497–548.

Baayen, R. Harald & Antoinette Renouf. 1996. Chronicling The Times: Productive lexical innovations in an English newspaper. *Language* 72. 69–96.

Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Lüdeling, Anke & Merja Kytö (eds.), *Corpus Linguistics. An International Handbook*. Vol. 2, 899–919. Berlin: de Gruyter.

Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 9(1). 5–16.

Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2005. Productivity: theories. In Štekauer, Pavol & Rochelle Lieber (eds.), *Handbook of word-formation*, 315–334. Dordrecht: Springer.

Bauer, Laurie, Rochelle Lieber, & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Booij, Geert & Rochelle Lieber. 2004. On the paradigmatic nature of affixal semantics in English and Dutch. *Linguistics* 42(2). 327–357.

Bonami, Olivier, & Juliette Thuilier. 2019. A statistical approach to rivalry in lexeme formation: French-iser and-ifier. *Word Structure* 12(1). 4–41.

Comrie, Bernard & Sandra A. Thompson. 2007. Lexical nominalization. In Timothy Shopen (ed.), *Language Typology and Syntactic Description vol. III: Grammatical Categories and the Lexicon*, 334–381. Cambridge: Cambridge University Press.

Comrie, Bernard. 1976. The syntax of action nominals: A cross-language study. *Lingua* 40(2). 177–201.

Fiorentino, Giuliana. 2008. Action nouns and the infinitive in Italian. *Romanische Forschungen* 120(3). 3–28.

Fox, John, Sanford Weisberg, Michael Friendly, Jangman Hong, Robert Andersen, David Firth & Steve Taylor. 2019. "*Package 'effects'*." ftp://gnualpha.uib.no/pub/cran/web/packages/effects/effects.pdf (accessed 1 February 2020).

Gaeta, Livio. 2002. *Quando i verbi compaiono come nomi. Un saggio di Morfologia Naturale*. Milano: Franco Angeli.

Gaeta, Livio. 2004. Derivazione nominale deverbale. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*, 314–351. Tübingen: Niemeyer.

Gaeta, Livio. 2015. Restrictions in word-formation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds*.), Word-Formation. An International Handbook of the Languages of Europe*, Berlin & New York: de Gruyter.

Gaeta, Livio. 2017. Nomi deverbali in MIDIA. In Paolo D'Achille & Maria Grossman (eds*.), Per la storia della formazione delle parole in italiano. Un nuovo corpus in rete (MIDIA) e nuove prospettive di studio*, 221–241. Milano: Franco Cesati.

Gaeta, Livio & Davide Ricca. 2002. Corpora testuali e produttività morfologica: i nomi d'azione italiani in due annate della Stampa (1996-1997). In Roland Bauer & Hans Goebl (eds.), *Parallela IX. Testo-variazione- informatica/Text-Variation-Informatik. Atti del IX Incontro italoaustriaco dei linguisti, Salzburg, 1-4 novembre 2000*, 223–249. Wilhelmsfeld: Egert.

Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39(4), 1041–1070.

Iacobini, Claudio & Anna Maria Thornton. 1992. Tendenze nella formazione delle parole nell'italiano del ventesimo secolo. In Bruno Moretti, Dario Petrini & Sandro Bianconi (eds.), *Linee di tendenza nell'italiano contemporaneo. Atti del XXV Congresso della Società di Linguistica Italiana (SLI), Lugano, 19–21 settembre 1991*, 25–55. Roma: Bulzoni.

Koptjevskaja-Tamm, Maria. 1993. *Nominalizations*. London: Routledge.

Melloni, Chiara. 2007. *Polysemy in word formation: the case of deverbal nominals*. PhD Thesis. Verona: University of Verona.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

R Core Team. 2015. *R: A language and environment for statistical computing*.

Rainer, Franz. 2005. Constraints on productivity. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*, 335–352. Dordrecht: Springer.

Rainer, Franz, Francesco Gardani, Wolfgang U. Dressler & Hans C. Luschützky (eds.). 2019. *Competition in Inflection and Word-formation*. Cham: Springer International Publishing.

Scalise, Sergio. 1983. *Morfologia lessicale*. Padova: Clesp.

Štichauer, Pavel. 2009. Morphological productivity in diachrony: The case of the deverbal nouns in *-mento*, *-zione* and *-gione* in Old Italian from the 13th to the 16th century. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Selected Proceedings of the 6th Décembrettes*, 138–147. Somerville, MA: Cascadilla Proceedings Project.

Talamo, Luigi, Chiara Celata & Pier Marco Bertinetto. 2016. DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure* 9(1), 72–102.

Thornton, Anna Maria. 1988. *Sui nomina actionis in italiano*. Phd Thesis. Pisa: University of Pisa.

Thornton, Anna Maria. 1990. Sui deverbali italiani in-mento e-zione (I). *Archivio glottologico italiano* 75(2). 169–207.

Thornton, Anna Maria. 1991. Sui deverbali italiani in-mento e-zione (II). *Archivio Glottologico Italiano* 76(1). 79–102.

Thornton, Anna Maria. 2015. Paradigmatically determined allomorphy: the "participial stem" from Latin to Italian. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation. An International Handbook of the Languages of Europe*, 780–802. Berlin & Boston: de Gruyter.

Varvara, Rossella. 2017. *Verbs as nouns: Empirical investigations on event-denoting nominalizations*. PhD Thesis. Trento: University of Trento.

Varvara, Rossella. 2019. Misurare la produttività morfologica: i nomi d'azione nell'italiano del ventunesimo secolo. In Bruno Moretti, Aline Kunz, Silvia Natale & Etna Krakenberger (eds.), *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana (Berna, 6–8 settembre 2018)*, 187–201. Milano: Officinaventuno

Vendler, Z. 1957. Verbs and times. *The Philosophical Review* 66(2), 143–160.

Rossella Varvara
Università degli Studi di Firenze
DILEF
via della Pergola 58–60
50121 Firenze, Italia
rossella.varvara@unifi.it

Marine Wauquier, Nabil Hathout & Cécile Fabre

Semantic discrimination of technicality in French nominalizations

Abstract: French suffixations in *-age*, *-ion* and *-ment* are considered roughly equivalent, yet some differences have been pointed out regarding the semantics of the resulting nominalizations. In this study, we confirm the existence of a semantic distinction between them on the basis of a large scale distributional analysis. We show that the distinction is partially determined by the degree of technicality of the denoted action: *-age* nominals tend to be more technical than *-ion* ones. We examine this hypothesis through the statistical modeling of technicality. To this end, we propose a linguistic definition of technicality, which we implement using empirical, quantitative criteria estimated in corpora and lexical resources. We show to what extent the differences with respect to these criteria adequately approximate technicality. Our study indicates that this definition of technicality, while amendable, provides new perspectives for the characterization of action nouns.

Keywords: action nouns, derivation, technicality, distributional semantics, statistical modeling

1. Introduction

Among the various French derivational processes available to coin deverbal action nouns, the suffixations in *-age* (*remplir* 'to fill' → *remplissage* 'filling'), *-ion* (*réduire* 'to reduce' → *réduction* 'reduction') and *-ment* (*allonger* 'to lengthen' → *allongement* 'lengthening') are the most productive, and are often said to be rival. Several semantic differences have been put forward to explain the suffix rivalry between them.

Some works point towards the nature of the base verb arguments. Kelling (2001) and Martin (2010) claim that verbs involved in *-age*, *-ion* and *-ment* suffixations differ in the agentivity degree of their subjects, while Fradin (2014) argues for a difference of concreteness of the referent denoted by their objects. Other works focus on the semantic properties of the base verb or on the nominalizations themselves. Dubois (1962) and Martin (2010) discuss the preference of the suffix *-age* for verbs that denote physical actions, and the use of *-age* nominalizations for industrial processes. They also acknowledge that the suffix *-ion* is more frequent in the scientific terminology. As for the suffix *-ment*, it is said to be unmarked at the ontological level – i.e. it does not select a domain-specific reading as *-age* does – (Martin 2010), and its nominalizations tend to denote attitude or psychological states (Dubois 1962).

In the light of the diachronic evolution of the suffix *-age*, Fleischman (1990), cited by Uth (2010), makes a claim similar to Dubois's that *-age* nominalizations are strongly associated to the industry domain. Fleischman states that the number of *-age* nominalizations increased in the 19th century, brought about by the industrial revolution and the growing need to designate new technologies and technical processes – which was later demonstrated by Uth (2010). According to Fleischman (1990), French borrowed part of its terminology from English, which extensively used the *-age* suffix, previously borrowed from French. Although the borrowing hypothesis has not been proved, the suffix *-age* is still very productive in French. This raises a question: do lexicalized and neological *-age* nominalizations still tend to be more technical than *-ion* and *-ment* nominalizations?

In this study, we investigate the specialization of derived action nouns in terms of technicality, and more specifically the hypothesis that *-age* action nouns have a higher degree of technicality. We show that this notion plays a significant role in the distinction between nominalizations in *-age*, *-ion* and *-ment*.

We first use a distributional semantic model of a contemporary French corpus to corroborate the existence of a difference in the degree of technicality of French nominals (Section 2). We then propose a definition of technicality and infer a set of linguistic corollaries that we translate into empirical criteria evaluated from corpora and lexical resources (Section 3). Finally, we assess the capacity of these criteria to gauge noun technicality and discriminate the three suffixes (Section 4). The results suggest that our definition of technicality can partially contribute to characterize *-age*, *-ion* and *-ment* deverbal action nouns insofar as *-age* nominalizations tend to be more technical than *-ion* and *-ment* nominalizations. This work is exploratory in nature and provides foundations for a new perspective on the semantic distinction between *-age*, *-ion* and *-ment*.

2. Distributional discrimination of action nouns

Our study of the notion of technicality and its linguistic expression stems from observations, made by means of distributional semantic, that confirm the existence of distributional differences between French nominalizations in *-age*, *-ion* and *-ment*. This section is dedicated to the presentation of these initial observations. The semantic distributional model we use was created with Word2Vec (Mikolov et al. 2013). Our experimental set-up also includes Lexeur, a lexical resource designed for the comparison of properties of words that belong to the same derivational family. We first describe these two resources before presenting the hypothesis and the first observations.

2.1. Lexeur

Our study is based on Lexeur, a French morphological database[1] which contains 5.974 derivational subfamilies of agent nouns suffixed in *-eur*.

Lexeur subfamilies are made up of: (i) an entry, namely a masculine noun in *-eur* (*abatteur* 'slaughterer', *camionneur* 'truck driver', *prédateur* 'predator'); (ii) one or several feminine equivalents (*abatteuse* 'female slaughterer' or 'harvester', *camionneuse* 'female truck driver', *prédatrice* 'female predator'); (iii) its verbal or nominal base if it exists (*abattre* 'to cut down' or 'to slaughter', *camion* 'truck'; *prédateur* does not have an attested base lexeme in French); (iv) a list of morphologically related verbs (*sélectionner* 'select' for *sélecteur* 'selector') and (v) a list of morphologically related nouns (*rectorat* 'board of education' for *recteur* 'superintendent of schools') when the entry is denominal or does not have an attested base lexeme; and (vi) a list of nominalizations of the base or nominals related to the entry (*abattage* 'slaughter', *prédation* 'predation'). All the fields except the first may be empty. The nouns in *-eur* were initially extracted from the *Frantext* corpus and from the French dictionary *Trésor de la Langue Française* (Dendien and Pierrel 2003). They were supplemented with words collected from the Web by annotators. All lexemes of the resource are associated with a morphosyntactic description in the Multext-Grace format. An excerpt of Lexeur is shown in Table 1.

Tab. 1: Five entries from Lexeur.[2]

MascAgt	sculpteur/ Ncms	inflammateur/ Ncms	inflammateur/ Ncms	autostoppeur/ Ncms	chiropracteur/ Ncms
FemAgt	sculpteuse/ Ncfs sculptrice/ Ncfs	inflammatrice/ Ncfs	inflammatrice/ Ncfs	autostoppeuse/ Ncfs	chiropractrice/ Ncfs
Base	sculpter/ Vmn----	enflammer/ Vmn----	inflammer/ Vmn----	autostop/ Ncms	
Nominals	sculpture/ Ncfs sculptage/ Ncms	inflammation/ Ncfs	inflammation/ Ncfs		chiropraxie/Ncfs

1 The resource was created in 2001 in the ERSS lab of CNRS (now CLLE lab) and the University of Toulouse Le Mirail (now University of Toulouse – Jean Jaurès) and is distributed under a creative commons license in a tabular separated format (tsv) and xml on the REDAC repository (http://redac.univ-tlse2.fr).

2 The fields Associated Verbs and Associated nouns are omitted because they are empty in all five entries. Masculine nouns are tagged Ncms, feminine nouns, Ncfs and infinitive verbs Vmn----.

Table 1 shows the diversity of the derivational families included in Lexeur: some columns are complete such as *sculpteur* 'sculptor', others are not, such as *chiropracteur* 'chiropractor' (with no identified base) or *auto-stoppeur* 'hitchhiker' (which has a nominal base but no other related nominal). 78% of the derivational subfamilies in the resource have a verb base, 14% are derived from a noun, and 8% are not associated with a base of any kind (such as *chiropracteur*). Some subfamilies include several nominals, such as *sculpteur* 'sculptor' (*sculpture*, *sculptage* 'sculpture'), while others don't have any, such as *auto-stoppeur* 'hitchhiker'. Moreover, some nouns in *-eur* have several bases and are part of several subfamilies (*inflammateur* 'igniter' may be derived from *inflammer* or *enflammer*, both equivalents of 'to ignite'). For these nouns, each base yields one derivational subfamily (i.e. one record).

2.2. Word2Vec

Distributional semantics is based on the hypothesis, known as the Distributional Hypothesis, that words that are semantically similar share similar distributions in corpora (Harris 1954; Firth 1957). Distributional Semantic Models (DSMs) represent words by vectors, called word embeddings, computed from all the contexts of those words in a corpus. Words with similar contexts are represented by close vectors and words with distinct contexts are represented by distant vectors. Being mathematical objects, word embeddings can be added, subtracted or multiplied in order to form analogies or compute the meaning of a phrase (Lenci 2018; Boleda 2020). We can also easily compute a distance between two vectors; the values range from 0 (no proximity) to 1 (strict equality).

The DSM we use in our study is built from the 2018 French edition Wikipedia. This corpus contains about 900 million words and has a large and diverse vocabulary that covers many domains and is in line with the vocabulary found in Lexeur. The corpus was lemmatized with Talismane (Urieli 2013). In order to limit the effects of the inherent instability of DSMs (Pierrejean 2020), we computed and concatenated 5 matrices into one unique DSM[3]. Each of these matrices is created by a new Word2vec run with all parameters set to their default values: CBOW architecture, Negative Sampling algorithm, a frequency threshold of 5, a window size of maximum 5, and 100 dimensions.

---

3   The DSMs were built using the OSIRIM computing platform that is administered by the IRIT computer science lab and supported by the National Center for Scientific Research (CNRS), the Région Midi-Pyrénées, the French Government, and the European Regional Development Fund (ERDF).

2.3. Methodology

Once the model is built, a proximity measure can be computed between each pair of words in the corpus. We can also identify the nearest neighbors of each word, i.e. the words that are the most distributionally similar (and the most semantically similar, according to the Distributional Hypothesis mentioned in Section 2.2.). For example, the nearest neighbors of *laminage* 'rolling' in our DSM are *forgeage* 'forging' (0.87), *formage* 'shaping' (0.85), *étirage* 'stretching' (0.81) and *extrusion* 'extrusion' (0.81), all denoting other steel or materials deformation processes.

The DSM can also be used to study distributional similarity at the level of noun classes by summarizing the information that is available at the level of the individual words. The idea is to represent each class by a prototypical representation of all the nouns in *-age*, *-ion* and *-ment*. If a semantic distinction exists between the three suffixations, as mentioned in Section 1, it should be reflected in the nouns coined by each suffixation, and be visible in the DSM. More specifically, a class of nouns may be represented by the average vector of the vectors of all its members (Kintsch 2001). The semantic content associated with this average vector is then characterized by its nearest neighbors (Marelli and Baroni 2015), since they are the words that are the most similar to the prototype of the class, and can therefore be considered to be representative instances of the class.

To construct these prototypical representations of *-age*, *-ion* and *-ment* nominalizations, we first need to select the members of each class. The nouns are extracted from Lexeur in the 'nominals' field of subfamilies with verb bases. At this stage, we have 4266 distinct deverbal nominals, including respectively 1687 nominalizations suffixed in *-age*, 1357 in *-ion* and 1222 in *-ment*. We decided to make a preliminary filtering before computing the *-age*, *-ion* and *-ment* centroids in order to build a more homogeneous distributional class for each suffix. We performed some manual filtering to exclude the nouns that: (i) are weakly related to the base verb indicated in Lexeur (*reportage* 'report' with respect to *reporter* 'to postpone'; 'to transfer') or not related (*pleurage* 'wow and flutter' with respect to *pleurer* 'to cry'), on a semantic and morphological basis; (ii) have no eventive meaning, such as *diction* 'diction' (from *dire* 'to say'). We also decided to limit the impact of polysemy. The models created by Word2Vec are not contextual and do not provide separate representations for the different meanings of polysemous words. In other words, forms like *garage*, meaning either the place where one parks a car or the action of parking, are represented by a single word embedding, aggregating the distributional information of the various meanings it covers. In order to limit the noise induced in the resulting representations, we excluded such nouns when we considered that the eventive meaning is less frequent than the non-eventive one – as is the case with *garage*. Finally, so as to ensure the coherence of the

data we use through this work, and because of the frequency threshold mentioned in Section 2.2., we systematically discarded the nouns with a frequency lower than 5 in the French *Wikipedia2018* corpus.

After the clean-up, we are left with 1.828 nouns, among which 629 are suffixed in *-age*, 750 in *-ion* and 449 in *-ment*. They are used to build the *-age*, *-ion* and *-ment* average vectors (henceforth AV$_{AGE}$, AV$_{ION}$ and AV$_{MENT}$). Their first 50 nearest neighbors are given in Table 2. Neighbors which are not coined by the suffix of the class are highlighted in bold.

Tab. 2: 50 nearest neighbors of -age, -ion *and* -ment *average vectors (AV$_{AGE}$, AV$_{ION}$ and AV$_{MENT}$).*

AV$_{AGE}$	*usinage* – 'machining' – *polissage* 'polishing' – *meulage* 'grinding' – *piquage* 'stitching' – *perçage* 'piercing' – *sablage* 'sand blasting' – *pliage* 'folding' – *remplissage* 'filling' – *salage* 'salting' – *soufflage* 'blow molding' – *démoulage* 'unmolding' – *décapage* 'scraping' – *sertissage* 'crimping' – *dégraissage* 'degreasing' – *assemblage* 'assembly' – *séchage* 'drying' – *rinçage* 'rinse' – **soudure** 'welding' – *enrobage* 'coating' – *nettoyage* 'cleaning' – **extrusion** 'extrusion' – *ponçage* 'sanding' – *compactage* 'compacting' – *broyage* 'crushing' – *malaxage* 'kneading' – *soudage* 'welding' – *façonnage* 'shaping' – **recuit** 'recooking' – *remontage* 'reassembly' – **rechargement** 'recharging' – *affûtage* 'sharpening' – *sciage* 'sawing' – *gonflage* 'inflation' – *tamisage* 'sieving' – *égouttage* 'draining' – *clouage* 'nailing' – **chargement** 'loading' – *calibrage* 'calibration' – *formage* 'shaping' – *brossage* 'brushing' – *réglage* 'tuning' – *traçage* 'tracing' – *cintrage* 'bending' – *lavage* 'washing' – *brasage* 'brazing' – *chromage* 'chrome plating' – *trempage* 'soaking' – *serrage* 'tightening' – *appareillage* 'casting off' – *coulage* 'pouring'
AV$_{ION}$	*généralisation* 'spread / generalization' – *manipulation* 'manipulation' – *dégradation* 'deterioration' – *simplification* 'simplification' – *stimulation* 'stimulation' – *contamination* 'contamination' – *dispersion* 'scattering' – *dénaturation* 'denaturation' – *transformation* 'transformation' – *récupération* 'recovery' – *utilisation* 'use' – *perception* 'perception' – *différenciation* 'differentiation' – *substitution* 'substitution' – *mutation* 'mutation / transfer' – *fixation* 'fastening' – *vérification* 'verification' – **persistance** 'persistence' – *prolifération* 'proliferation' – *assimilation* 'assimilation' – *altération* 'alteration' – *détermination* 'determination' – *réduction* 'reduction' – **saisie** 'seizure/input' – *dilution* 'dilution' – *conversion* 'conversion' – *activation* 'activation' – *compréhension* 'understanding' – *transmission* 'transmission' – *réutilisation* 'reuse' – *définition* 'definition' – **dégénérescence** 'degeneration' – **synthèse** 'overview / synthesis' – *redistribution* 'redistribution' – *modification* 'modification' – *multiplication* 'multiplication' – *régénération* 'regeneration' – *ponction* 'puncture' – **traitement** 'treatment' – *cristallisation* 'crystallization' – *décomposition* 'decomposition' – *détérioration* 'deterioration' – *stérilisation* 'sterilization' – *restriction* 'restriction' – *réaction* 'reaction' – *centralisation* 'centralization' – *dissociation* 'dissociation' – *dissémination* 'dissemination' – *standardisation* 'standardization' – *acceptation* 'acceptance'

| AV$_{\text{MENT}}$ | *déplacement* 'moving' – *étirement* 'stretching' – *durcissement* 'hardening' – *ajustement* 'adjustment' – *relâchement* 'slackening' – *traitement* 'treatment' – *adoucissement* 'softening' – **utilisation** 'use' – *échauffement* 'warm–up / heating' – *enfoncement* 'knocking in' – *décollement* 'detachment' – **rejet** 'rejection' – *rétrécissement* 'narrowing' – *affaiblissement* 'weakening' – *endommagement* 'damaging' – *fonctionnement* 'functioning' – *dépassement* 'passing' – *engorgement* 'congestion' – *allongement* 'lengthening' – *tassement* 'settling' – *encombrement* 'congestion' – *abaissement* 'lowering' – **usure** 'wear' – *débordement* 'overflowing' – *réajustement* 'readjustment' – *étalement* 'spread' – **remplissage** *'filling'* – *dégagement* 'release' – *écoulement* 'flow' – *éloignement* 'distancing' – **équilibrage** 'balancing' – *effritement* 'crumbling' – *isolement* 'isolation' – **inconfort** 'discomfort' – *fléchissement* 'bowing' – *gonflement* 'swelling' – *relèvement* 'raising' – *retournement* 'turnaround' – **déséquilibre** 'imbalance' – **colmatage** 'clogging' – *accroissement* 'increase' – *arrachement* 'abduction' – *basculement* 'toppling' – *accumulation* 'accumulation' – *balancement* 'swinging' – *amincissement* 'slimming' – **passage** 'passage' – *dessèchement* 'drying' – *dysfonctionnement* 'dysfunction' – **court–circuit** 'short circuit' – **activation** 'activation' |

In the following, we conduct the analysis of their 100 nearest neighbors. This number is chosen arbitrarily so as to provide as broad a view as possible, and yet be significant.

2.4. Results analysis

These results show that the first 100 nearest neighbors of each average vector present a strong homogeneity: 82% of the AV$_{\text{AGE}}$ neighbors are nouns in *-age*; 80% of the AV$_{\text{ION}}$ neighbors are suffixed in *-ion*; 73% of the AV$_{\text{MENT}}$ neighbors end with the *-ment* suffix. By comparison, the homogeneity is much lower for the neighbors of the average vectors of *-eur*, *-euse* and *-rice* agent nouns (Wauquier et al. 2018), with respectively 44%, 10% and 16% of their neighbors having the suffix of the class. In other words, AV$_{\text{AGE}}$, AV$_{\text{ION}}$ and AV$_{\text{MENT}}$ neighborhoods are located in distinct areas of the vector space, which reflects a clear distributional difference between the three classes of nominals.

    On a semantic and referential level, the three neighborhoods exemplified in Table 2 almost exclusively contain action nouns but display some semantic differences. The AV$_{\text{ION}}$ neighbors denote processes or phenomena related to sciences, such as *dilution* 'dilution' or *dénaturation* 'denaturation', which goes in line with Dubois' claim. There are also nouns describing psychological processes, such as *compréhension* 'understanding', *détermination* 'determination', or *perception* 'perception', and to a greater extent nouns denoting very broad concepts, such as *modification* 'modification' and *utilisation* 'use', characterized by a high degree of polysemy. The neighborhood of

the AV$_{AGE}$ vector contrasts strongly with that of AV$_{ION}$. It contains a large amount of nouns related to industrial skills or processes, such as *soudage* 'welding', *usinage* 'machining', and *brasage* 'brazing', and very few, if not any, generic or underspecified nouns. The most generic nouns like *stockage* 'storage', *nettoyage* 'cleaning' and *lavage* 'washing' denote actions which are intrinsically more technical than *utilisation* or *modification* and many other neighbors of the AV$_{ION}$ vector. The AV$_{MENT}$ neighborhood seems more mixed. Some neighbors designate relatively large concepts, such as *déplacement* 'moving' and *traitement* 'treatment', while others are more specific, such as *relèvement* 'raising' and *équilibrage* 'balancing'.

To summarize, the two semantic profiles that emerge for the nominals in *-age* and *-ion* are clearly distinct: the former displays a higher degree of technicality, while the latter has a higher degree of genericity. The semantic profile of the nominals in *-ment* is less marked. In what follows, we focus on this distinction and we propose a set of criteria that characterize technicality.

3. The concept of technicality

Technicality is a notion hardly ever discussed in the literature. The adjective 'technical' is sometimes used to refer to specific types of vocabulary or corpora, but technicality has not yet been properly defined, as Mudraya (2006) points out. However, we need such a definition in order to better characterize the semantic distinction we just highlighted. To the best of our knowledge, no work has been dedicated to this notion of technicality from a non-terminological point of view. Conversely, outside the domain of linguistics, the philosopher Simondon (1958) extensively examines the issue. First we review his description, then we propose our own definition of technicality for action nouns. Finally, we investigate its implementation by means of linguistic criteria that can be automatically measured. It should be pointed out that we do not take a terminological stand on technicality here. Our aim is to propose a definition that relies on lexical and referential criteria and is valid independently of any particular domain.

3.1. Definition of technicality

Simondon (1958) gives a definition of technicality along several dimensions. The first one is agentivity. According to the author, technicality is closely linked to human beings, as they are "among the machines that operate with them" ("*[ils sont] parmi les machines qui opèrent avec [eux]*") (1958: 12). Incidentally, the philosopher defines technicality with respect to machines and more largely to objects he describes as technical: "technicality manifesting itself through the use of technical objects" ("*la technicité*

se manifestant par l'emploi d'objets techniques") (1958: 156). The degree of technicality of an object depends on its level of sophistication and complexity, estimated according to whether the knowledge required to use it is innate or acquired. An innate, unreflective knowledge, related to an everyday life object, conveys a lesser degree of technicality of the action, whereas an action that involves "a reflective operation, a rational knowledge constructed by sciences" ("*une opération réfléchie, d'une connaissance rationnelle élaborée par les sciences*") (Simondon 1958: 85) is considered technical. The more the knowledge is learned and constructed, the higher the technicality of the denoted action. Ontologically, Simondon indicates that "technique concerns business, agriculture, industry" ("*la technique touche au commerce, à l'agriculture, à l'industrie*") (Simondon 1958: 97). Following Simondon (1958), we propose in (1) the definition of technical action nouns.

(1)   A technical action noun is a noun unfamiliar to non-experts, denoting a specific and complex action, whose achievement and understanding require an acquired skill and which is specific to a particular domain. Technical action nouns typically, but not exclusively, belong to domains such as industry, agriculture and arts and crafts.

As such, we consider as technical any noun that denotes an action performed intentionally by an agent and whose complexity requires from the agent some specific learned knowledge that might (but not necessarily) be at the core of a well-defined task. Technical actions cannot be successfully carried out by all comers in everyday life, as they involve in addition to specific knowledge a particular setup. This distinguishes action nouns such as *danse* 'dance' and *ébreuillage* 'fish evisceration': one can dance at any time even while not being a professional dancer, while fish evisceration requires particular skills and gears to be performed correctly. The complexity and specificity of the action result in the unfamiliarity of non-experts with the corresponding nouns. According to our approach, domains other than industry or agriculture, such as scientific disciplines, can also provide technical nouns that denote complex actions involving tools and devices, as well as specific knowledge (e.g. *inoculation* 'inoculation', *cassation* 'quashing'). However, being part of a specialized domain is neither a necessary nor a sufficient condition for a noun to be technical (see *aimance* 'child-specific affection', which belongs to a specialized domain, i.e. the psychology field, yet does not meet our definition of technical noun). Note that while definition in (1) suggests that there is a class of technical nouns, technicality is actually considered as a gradable property that is instantiated at various degrees.

3.2. Linguistic properties of technicality

As we just saw, technicality relies on world knowledge and referential criteria and needs to be translated into linguistic properties that can be empirically investigated. Three linguistics properties of technicality (henceforth *T1*, *T2* and *T3*) can be derived from the previous definition and implemented by a set of criteria automatically calculated on corpora and lexicons.

The first is **specialization** (*T1*): a technical action noun is more often used in specialized contexts than in general contexts because it falls within a particular domain and the action it denotes is specific. This denotational specificity gives rise to the second property, namely **obscurity** (*T2*): a technical action noun is more often described and explained to non-experts than a non-technical noun because it denotes a reality which is unfamiliar to them. The description can be a dictionary definition, an article in an encyclopedia, etc. The third property we derive from the definition in (1) is **univocity** (*T3*): a technical action noun tends to be unequivocal, in contrast with a generic noun. Monosemy can be used as an approximation of univocity.

3.3. Criteria to approximate technicality

At this stage, the previous three linguistic properties must be operationalized. Several measures can be used to assess the degree of specialization, obscurity and univocity of a given action noun. Our choice is influenced by the need for automatic annotation and is based on available linguistic resources. These criteria are exploratory and we are aware that this first attempt to estimate technicality will have to be refined with respect to the preliminary results we present in Section 4. Table 3 presents the criteria that we derived from the properties just described.

Tab. 3: Technicality criteria.

Property	Criteria
T1	Ratio of the relative frequencies in a technical corpus and in a reference corpus
	Number of lexicographical markers of domains in dictionaries or encyclopedia
	Presence or absence in transdisciplinary scientific lexicons
T2	Presence or absence of an article in an encyclopedia
T3	Number of synonyms
	Number of definitions in dictionaries
	Presence or absence in generic nouns lexicons

The degree of specialization of a word is assessed by corpora comparison (Lemay et al. 2005) and by means of lexicons that provide information on

domain membership and transdisciplinarity (Hatier 2016). As for the obscurity property, we use one criterion: the presence or absence of the noun as an entry in an encyclopedia. Different *T3* criteria allow for the approximation of the various aspects of equivocality, namely polysemy and underspecification. These criteria are computed from various resources, as shown in Table 4.

Tab. 4: Resources and lexicons.

Name	Size	Description
Wikipedia2018	600 million words	Encyclopedic corpus built from the 2018 French edition of *Wikipedia*
LM10	200 million words	French journalistic corpus made up of articles of *Le Monde* newspaper published between 1991 and 2000
DES	83,395 entries	Electronic dictionary of synonymes (Manguin et al 2004)
TLFi	54,280 entries	Electronic version of the *Trésor de la Langue Française* dictionary (Dendien and Pierrel 2003)
GLAWI	1,481,346 entries	Electronic dictionary built from the French *Wiktionary* (Hathout and Sajous 2016)
LexiTrans	1,611 entries	Transdisciplinary scientific lexicon (Drouin 2010)
LexNSS	305 entries	List of underspecified nouns extracted from Legallois and Gréa (2006)

In this study we choose to use several dictionaries in order to reduce the impact of specific lexicographic choices. Regarding corpora, the *LM10* corpus has been chosen as a reference corpus, as it is considered to be an example among others of non-technical discourse. The choice of the *Wikipedia2018* corpus as a technical corpus is supported by its encyclopedic nature, by the fact that it incorporates a large panel of technical domains, but also by the lack of a large and diversified technical corpus for French. We also use *Wikipedia2018* to test the presence or absence of an article describing the action denoted by the noun (PAGE_W18). Only the articles with a title strictly identical to the noun are taken into account. For example, we consider that *serrage* 'tightening' does not have an article in *Wikipedia2018* even if it contains articles entitled *collier de serrage* 'horse clamp' or *noix de serrage* 'clamp holder'. As for lexicons, we both use large general dictionaries for French (*Trésor de la Langue Française* and *Glawi*) and smaller specific lexicons that give access to synonymy, transdisciplinary vocabulary (as a clue for non-specialization) and shell nouns (as a clue for equivocality).

Our criteria along with their associated resources are given in Table 5.

Tab. 5: Implementation of the technicality criteria.

Property	Criteria	Name
T1	Ratio of the relative frequencies (per million words) in *Wikipedia2018* and *LM10*	RATIO_FREQR
	Number of category markers in *Wikipedia2018*	NB_CAT_W18
	Number of lexicographical markers of domains in *TLFi*	NB_DOM_T
	Number of lexicographical markers of domains in *GLAWI*	NB_DOM_G
	Presence or absence in *LexiTrans*	LST
T2	Presence or absence of an article in *Wikipedia2018*	PAGE_W18
T3	Number of synonyms in *DES*	NB_SYN
	Number of definitions in *TLFi*	NB_DEF_T
	Number of definitions in *GLAWI*	NB_DEF_G
	Presence or absence in *LexNSS*	NSS

As we can see in this table, we use very simple measures that check for the presence of the noun in the lexicons or count the number of lexical items (definitions, synonyms) that are found in the entry. The ratio of relative frequency is calculated by dividing for a given word its relative frequency in the technical corpus by its relative frequency in the reference corpus (Hatier 2016).

It should be noted that technicality is estimated by the combination of these criteria. They aim to highlight tendencies in the degrees of technicality of action nouns, and as such, there is no threshold value we could use as a clue for a binary characterization of a noun as technical or non-technical.

4. Statistical modeling of technicality

The criteria we just presented enable us to empirically test the hypothesis of a higher degree of technicality for *-age* nouns and of a lesser degree of technicality for *-ion* and *-ment* nouns. Following the definition of technicality given in Section 3.1., *-age* nominalizations are expected to have higher values than *-ion* and *-ment* nominalizations for only two criteria related to *T1* and *T2*, and lower values than *-ion* and *-ment* nominalizations for the others criteria: they will have a higher frequency ratio (RATIO_FREQR) and will be more likely to have an article in Wikipedia2018 (PAGE_W18), but they will be less present in the transdisciplinary lexicon (LST) and among the underspecified nouns (NSS), they will have fewer synonyms (NB_SYN), definitions (NB_DEF), and domain markers (NB_DOM).

4.1. Annotation

To test our predictions, we automatically annotated the 1828 -*age*, -*ion* and -*ment* action nouns selected in Section 2.1. for the criteria presented in Section 3.3. Table 6 presents the annotation of 4 nouns selected to illustrate the opposition between technical nouns (*alunissage* 'moon landing' and *cimentage* 'cementing') and non-technical nouns (*correction* 'correction' and *revendication* 'demand') according to our definition in (1). Our predictions are shown in the table with the (+) and (–) marks. First, we can see that the 4 examples are fairly well described by the criteria we implemented even if some criteria, when considered individually, do not systematically conform to our expectations. For example, the absence of a noun in the transdisciplinary (Lsт) lexicon does not ensure its technicality (*revendication*). Moreover, a high number of definitions (Nв_def_g and Nв_def_t) is not necessarily a good clue for the non-technicality of the noun (*cimentage*). Yet, the technical nouns *alunissage* and *cimentage*, denoting respectively a specific maneuver of a spatial engine and a process of the construction industry, have several values that are close to our expectations (higher ratio in ratio_freqr, and fewer synonyms, definitions and domains, respectively in Nв_syn, Nв_def and Nв_dom). Similarly, non-technicality seems to be globally captured, as shown by the nouns *correction* and *revendication*.

Tab. 6: Values of the technicality criteria for the action noms alunissage *'moon landing',* cimentage *'cementing',* correction *'correction' et* revendication *'demand'.*

	Technicality	*alunissage*	*cimentage*	*correction*	*revendication*
Ratio_freqr	+	3.06	2.18	1.14	0.23
Nв_cat_w18	–	3	0	0	0
Nв_dom_t	–	1	2	7	3
Nв_dom_g	–	1	0	4	1
Lst	–	No	No	No	No
Page_w18	+	Yes	No	Yes	No
Nв_syn	–	1	0	87	45
Nв_def_t	–	1	3	33	6
Nв_def_g	–	1	1	8	3
Nss	–	No	No	No	Yes

Table 6 highlights the role of our criteria as tendency indicators and not as class delimiters. We do not have technical and non-technical nouns *per say*, but nouns that have a higher degree of technicality than others. Among the nouns having the highest degree of technicality with respect to our criteria, i.e. the nouns that overall conform the most to our predictions, we find *hydroformage* 'hydroforming', *zingage* 'galvanizing', *cardage* 'carding' and *oxycoupage* 'oxy cutting', and among the lowest *association* 'combination', *division* 'division', *commencement* 'beginning' and *approbation* 'endorsement'.

We provide in Table 7 the average values of the technicality criteria for the 1.828 action nouns in *-age*, *-ion* and *-ment*. The presence in LST and NSS lexicons is presented as the percentage of nouns belonging to these lexicons. The existence of an article corresponding to the noun in the *Wikipedia2018* corpus is also given as a percentage.

Tab. 7: Average values of the technicality criteria with respect to each suffix.

	-age	*-ion*	*-ment*
Ratio_freqr	2.62	2.16	1.57
Nb_cat_w18	0.78	0.93	0.42
Nb_dom_t	1.2	2.65	1.27
Nb_dom_g	0.65	0.93	0.46
Lst (%)	0.1	8.13	2.23
Page_w18 (%)	48.97	69.87	34.08
Nb_syn	3.03	13.83	11.01
Nb_def_t	2.52	5.8	4.34
Nb_def_g	2.01	2.54	1.98
Nss (%)	0	2	1.11

The results shown in Table 7 corroborate the hypothesis of a higher degree of technicality of *-age* action nouns, and a lesser degree of technicality of *-ion* action nouns. We can see that *-age* nominalizations have on average fewer synonyms (Nb_syn), definitions (Nb_def) and domain markers (Nb_dom) than *-ion* nominalizations (p-value < 0.01 except for the number of category markers in *Wikipedia2018*, i.e. nb_cat_w18). Moreover, they are proportionally less present in the LST and NSS lexicons than *-ion* nominalizations (p-value < 0.01). It can be noticed that the differences of values displayed by the suffixes are higher for measures extracted from *TLFi* (Nb_def_t and Nb_dom_t) than for those from *GLAWI* (Nb_def_g and Nb_dom_g).

However, one criterion is not in line with our predictions, namely Page_w18 (presence or absence of a page in *Wikipedia2018*). We can see in Table 7 that the percentage of nouns having an article in the Wikipedia2018 corpus is lower for *-age* than for *-ion* (49% for the former vs 70% for the latter). Although the difference between each pair of suffixes is significant (p-value < 0.01), it fails to approximate the technicality of action nouns. Moreover, while the average ratio of relative frequency (ratio_freqr) is slightly higher for *-age* than for *-ion*, the difference is not significant. The inability of these criteria to discriminate between technical and non-technical action nouns is probably explained by the choice of the *Wikipedia2018* corpus. It is not technical enough and consequently not appropriate to evaluate the properties regarding specialization and obscurity.

Table 7 also shows that *-ment* action nouns are in an intermediate position with respect to our criteria: we notice that its average values lie between those of *-age* and *-ion* for criteria such as the number of synonyms (NB_SYN), the number of definitions in *GLAWI* (NB_DEF_G) or the presence in LST and NSS lexicons. Yet, its average values for other criteria such as the number of categories in *Wikipedia2018* (NB_CAT_W18) or the number of domains markers in both dictionaries (NB_DOM) are lower than those of *-age*. Furthermore, the difference between *-age* and *-ment* is not significant for NB_DEF_G and NB_DOM_T. In contrast, the average values for the suffix *-ment* are never higher than those of *-ion*. These observations suggest that *-ment* nominalizations are closer to *-age* than to *-ion* with respect to the technicality of the actions they denote.

4.2. Predictive capability of the criteria

In the last step of our analysis, we can now assess the extent to which our technicality criteria allow for the prediction of the suffix of an action noun. Their discriminative capability can be estimated by means of a decision tree which predicts the suffix of the noun from the values of the criteria. We used the *rpart* package in R to build a classification decision tree that classifies the 1828 *-age*, *-ion* and *-ment* action nouns based on their annotation. The resulting decision tree is presented in Figure 1. The values given under the leaves are the number of *-age*, *-ion* and *-ment* nominalizations respectively in each cluster. The letters "O" and "N" respectively stand for *yes* and *no*.
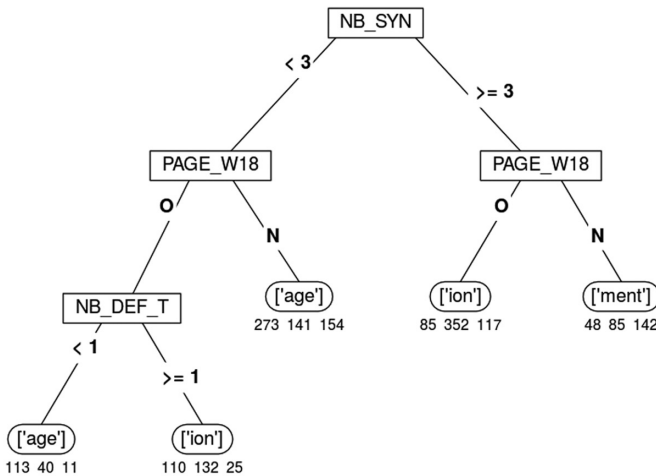

Fig. 1: Classification tree.

Figure 1 shows both the rules inferred from the data and the contribution of the criteria to the classification. We see that only 3 of the 10 initial criteria

are actually used to classify the nominals, namely the number of synonyms (Nb_syn), the number of definitions in *TLFi* (Nb_def_t) and the presence or absence of an article in Wikipedia2018 (Page_w18). This last rule confirms the observation made from Table 7 that this criterion does not satisfactorily operationalize the *T2* property and shows that it actually plays in the opposite way. The Page_w18 criterion selects the non-technical nominals, and the *-ion* suffix.

The overall accuracy of the model is 55.3%, meaning that hardly half of the action nouns are properly classified. Yet the three suffixes are not classified with the same precision. The performances of the model are given in the confusion matrix in Table 8. The number of items predicted as *-age*, *-ion* and *-ment* nominals (in the columns) is given with respect to the actual number of these nominals (in the rows). The correctly classified items are highlighted (in bold). Table 8 shows that 386 nominals in *-age* are identified as part of the *-age* class, 195 as *-ion* nouns and 48 as *-ment* nouns. It also shows that the model achieves similar performances for *-age* and *-ion* (respectively 61.4% and 64.5%) but a lower performance for *-ment* (31.6%).

Tab. 8: Confusion matrix.

		Predicted		
		-age	*-ion*	*-ment*
	-age	**386**	195	48
Observed	*-ion*	181	**484**	85
	-ment	165	142	**142**

A closer look at the misclassified nouns first shows that the criteria Nb_syn, Nb_def_t and Page_w18 do a good job in discriminating nouns in terms of technicality. Among the misclassified *-age* nominals (i.e. the ones that are predicted to be *-ion* or *-ment* nouns), we find nouns such as *papotage* 'chattering', *batifolage* 'frolic' and *babillage* 'babbling' which are not technical and are labeled as *-ment* by the classifier due to their high number of synonyms. Although they are misclassified at the morphological level, the prediction is correct on the semantic level, given their low degree of technicality. Other misclassified nouns such as *damage* 'tamping', labeled as *-ion* because it has a definition in *TLFi*, show that some rules are given too much importance. However, the error analysis brings to light that 11% of the misclassified *-ion* nouns (i.e. nominals in *-ion* which were assigned to another suffix) are suffixed in *-ification* or *-isation* (*panification* 'breadmaking'; *dessalinisation* 'desalination') and labeled as *-age* because of the inherent technicality of their base verbs.

In a second experiment, we further refined our assessment of the criteria. We excluded the PAGE_W18 criterion because it does not properly implement the *T2* property. The resulting decision tree is given in Figure 2.
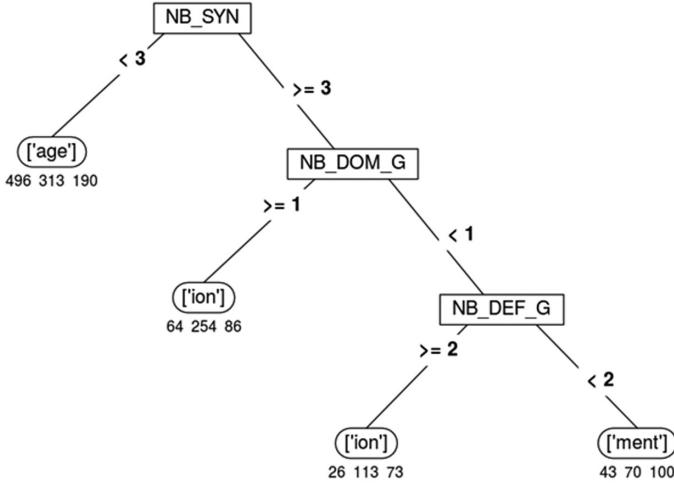
Fig. 2: Decision tree excluding the PAGE_W18 criterion.

In the new model the number of synonyms is still the first criterion and is now followed by the number of domains and the number of definitions in *GLAWI*. The rules inferred by this model are more in line with our hypothesis and predictions: *-age* nominals are discriminated from the others by their lower number of synonyms (< 3); action nouns are labeled as *-ion* nominals if they have more than one domain marker in *GLAWI*; *-ion* nominals are discriminated from the *-ment* nominals by their number of definitions (>=2). Compared to the previous model, this one has a lower overall accuracy (52.7%), but has an improved precision for *-age* nominals (78.9%) and a lower one for *-ion* (48.9%) and *-ment* (22.3%). The corresponding confusion matrix is given in Table 9.

Tab. 9: Confusion matrix.

		Predicted		
		-age	*-ion*	*-ment*
	-age	**496**	90	43
Observed	*-ion*	313	**367**	70
	-ment	190	159	**100**

A closer look at the *-age* misclassified nouns shows that they were already misclassified by the previous model. The labeling of *-ification* and *-isation*

nominalizations as technical is emphasized, since 53.2%, respectively 81%, of them are considered to be *-age*, vs 10.6%, respectively 7.1%, as *-ment*.

This experiment challenges the relevance of some of our criteria, in particular those linked to the corpora. Even though it was used in an opposite direction, Page_w18 actually improves the overall prediction of the first model, and in particular the classification of the *-ion* and *-ment* nouns. Regarding the ratio of relative frequencies, it may be that the corpus *Wikipedia2018* is not technical enough, and too diverse to actually favor *-age* nominalizations. Regarding the presence of an article in *Wikipedia2018*, we might want to revise its implementation and the hypothesis supporting this criterion. Taking into account all the articles whose title contains the noun might emphasize more the genericity than the technicality. While articles such as *noix de serrage* 'clamp holder' would be taken into account, allowing for *serrage* 'tightening' to validate the Page_w18 criterion, many generic nouns would also fulfill the Page_w18 criterion, such as *utilisation* 'use' which occurs in *utilisation frauduleuse des instruments de paiement* 'fraudulent use of payment.

5. Conclusion

In this study, we investigated the factor of technicality in rivalry between French action nouns in *-age*, *-ion* and *-ment*. We first proposed a definition of technical action nouns as being nouns unfamiliar to non-experts, which denote a specific, complex action that requires an acquired skill and is grounded in domains such as industry, agriculture and handicraft. We showed how linguistic properties and criteria could be extracted from this definition and computed from corpora and lexical resources in order to characterize the technicality of the nominals. Some of these criteria, such as the number of synonyms and definitions, have proven to be effective in discriminating *-age* nominals which are more technical, from *-ion* nominals which are less technical. Other criteria turned out to be less relevant, in particular the ones computed from corpora (e.g. the ratio of relative frequencies).

We also saw that *-ion* nominals are more heterogeneous than expected in terms of technicality, mainly because of the presence of inherently technical *-isation* and *-ification* action nouns. In future work, we intend to use a finer-grained dataset where these nominals will be part of the same class as *-age* nominals. We will also take into account other derivational processes, such as conversion (*baisse* 'lowering') or *-ure* suffixation (*rature* 'crossing out').

This work is a first attempt to characterize and approximate technicality through empirical criteria. It offers a new insight on the distinction between *-age*, *-ion* and *-ment* French nominals and highlights the limits of some exploratory criteria. In particular, we need to improve the criteria that estimate obscurity. We also intend to explore new criteria such as concreteness (Pierrejean 2020; Köper & Schulte im Walde 2016) and instrumentality (Missud 2019):

there is a certain amount of overlap between technicality and the simpler notion of concreteness because technical action nouns mostly denote concrete actions (*usinage* 'machining', *extrusion* 'extrusion'), as opposed to abstract ones (*compréhension* 'understanding', *perception* 'perception'). Finally, we will also evaluate the possible correlation between our criteria and the level of technicality perceived by the speakers. Ultimately, we aim to propose a single technicality score that aggregates the relevant criteria and annotations.

References

Dendien, Jacques & Jean-Marie Pierrel. 2003. Le Trésor de la Langue Française Informatisé: un exemple d'informatisation d'un dictionnaire de langue de référence. *Traitement Automatique des Langues* 44(2). 11–37.

Drouin, Patrick. 2010. Extracting a Bilingual Transdisciplinary Scientific Lexicon. In Sylviane Granger & Magali Paquot (eds.), *eLexicography in the 21st century: new challenges, new applications. Proceedings of eLex 2009*, 45–53. Louvain-la-Neuve: Presses Universitaires de Louvain.

Dubois, Jean. 1962. *Étude sur la dérivation suffixale en français moderne et contemporain: essais d'interprétation des mouvements observés dans le domaine de la morphologie des mots construits*. Paris: Larousse.

Firth, John R. 1957. A synopsis of linguistic theory 1930–1955. In John R. Firth (eds), *Studies in linguistic analysis*, 1–32. Oxford: Basil Blackwell.

Fleischman, Suzanne. 1990. *The French Suffix -age: its Genesis, Internal Growth, and Diffusion*. Ann Arbor: Univ. Microfilm Intern.

Fradin, Bernard. 2014. La variante et le double. In Florence Villoing, Sarah Leroy & Sophie David (eds.), *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 109–147. Paris: Presses universitaires de Paris Nanterre.

Harris, Zellig S. 1954. Distributional structure. *Word* 10(2-3). 146–162.

Hathout, Nabil & Franck Sajous. 2016. Wiktionnaire's Wikicode GLAWIfied: A Workable French Machine-Readable Dictionary. In *Proceedings of the tenth International Conference on Language Resources and Evaluation (LREC)*, 1369–1376. Portorož.

Hatier, Sylvain. 2016. *Identification et analyse linguistique du lexique scientifique transdisciplinaire. Approche outillée sur un corpus d'articles de recherche en SHS*. Grenoble: University Grenoble Alpes doctoral dissertation.

Kelling, Carmen. 2001. Agentivity and Suffix Selection. In *Proceedings of the Lexical-Functional Grammar '01 Conference (LFG'01)*, 147–162. Hong Kong.

Kintsch, Walter. 2001. Predication. *Cognitive science* 25. 173–202.

Köper, Maximilian & Sabine Schulte im Walde. 2016. Automatically Generated Norms of Abstractness, Arousal, Imageability and Valence for 350 000 German Lemmas. In *Proceedings of the 10th Conference on Language Resources and Evaluation (LREC)*, 2595–2598. Portorož.

Legallois, Dominique & Philippe Grea. 2006. L'objectif de cet article est de…Construction spécificationnelle et grammaire phraséologique. *Cahiers de praxématique* 46. 161–186.

Lemay, Chantal, L'Homme, Marie-Claude & Patrick Drouin. 2005. Two methods for extracting "specific" single-word terms from specialized corpora: Experimentation and evaluation. *International Journal of Corpus Linguistics* 10(2). 227–255.

Lenci, Alessandro. 2018. Distributional Models of Word Meaning. *Annual review of Linguistics* 4. 151–171.

Manguin, Jean-Luc, Jacques François, Rembert Eufe, Ludwig Fesenmeier, Corinne Ozouf & Morgane Senechal. 2004. Le dictionnaire électronique des synonymes du CRISCO: un mode d'emploi à trois niveaux. *Cahiers du CRISCO* 17.

Marelli, Marco & Marco Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological review* 122(3). 485–515.

Martin, Fabienne. 2010. The Semantics of Eventive Suffixes in French. In Artemis Alexiadou & Monika Rathert (eds.), *The Semantics of Nominalizations across Languages and Frameworks*, 109–141. Berlin & New York: de Gruyter.

Mikolov, Tomas, Kai Chen, Greg Corrado, & Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of International Conference on Learning Representations (ICLR)*. Scottsdale.

Missud, Alice. 2019. *Modélisation quantitative de la rivalité entre la suffixation en -age et la conversion de verbe à nom*. Paris: University Paris Nanterre master thesis.

Mudraya, Olga. 2006. Engineering English: A lexical frequency instructional model. *English for Specific Purposes* 25(2). 235–256.

Pierrejean, Bénédicte. 2020. *Qualitative Evaluation of Word Embeddings: Investigating the Instability in Neural-Based Models*. Toulouse: University Toulouse - Jean Jaurès doctoral dissertation.

Simondon, Gilbert. 1958. *Du mode d'existence des objets techniques*. Paris: Éditions Aubier-Montaigne.

Urieli, Assaf. 2013. *Robust French syntax analysis: reconciling statistical methods and linguistic knowledge in the Talismane toolkit*. Toulouse: University Toulouse - Jean Jaurès doctoral dissertation.

Uth, Melanie. 2010. The rivalry of French -ment and -age from a diachronic perspective. In Artemis Alexiadou & Monika Rathert (eds.), *The Semantics of Nominalizations across Languages and Frameworks*, 215–244. Berlin & New York: de Gruyter.

Wauquier, Marine, Cécile Fabre & Nabil Hathout. 2018. Différenciation sémantique de dérivés morphologiques à l'aide de critères distributionnels. In *Congrès Mondial de Linguistique Française (CMLF)*, Mons, Belgique.

Marine Wauquier, Nabil Hathout & Cécile Fabre
CLLE, University Toulouse Jean Jaurès & CNRS
5, allée Antonio Machado
31058 Toulouse cedex 9
France
{marine.wauquier, nabil.hathout, cécile.fabre}@univ-tlse2.fr

Gianina Iordăchioaia, Susanne Schweitzer, Yaryna
Svyryda & María Camila Buitrago Cabrera

Deverbal zero-nominalization and verb classes: Insights from a database

Abstract: We investigate deverbal zero-derived nominals in English (e.g., *to walk > a walk*) from the perspective of the lexical semantics of their base verbs and the interpretations they may receive (e.g., event, result state, product, agent). By acknowledging that, in the absence of an overt affix, the meaning of zero-nominals is highly dependent on that of the base, the ultimate goal of this study is to identify possible meaning regularities that these nominals may display in relation to the different semantic verb classes. We report on a newly created database of 1,000 zero-derived nominals, which have been collected for various semantic verb classes. We test previous generalizations made in the literature in comparison with suffix-based nominals and in relation to the ontological type of the base verb. While these generalizations may intuitively hold, we find intriguing challenges that bring zero-derived nominals closer to suffix-based nominals than previously claimed.

Keywords: morphology, lexical semantics, zero-derived nominals, manner and result verbs, argument structure

1. Introduction

Zero-derived nominals (ZNs) are deverbal nominalizations that do not show any overt marking compared to suffix-based nominalizations (SNs), which involve overt nominalizing suffixes, as illustrated in (1). ZNs are also known as conversion nouns.

(1) a. to walk – the walk-Ø (ZN) – the walk-**ing** (SN)
    b. to invite – the invite-Ø (ZN) – the invit-**ation** (SN)

English ZNs have not received much attention in recent generative literature, despite having sparked more interest in earlier studies such as Marchand (1969), Irmer (1972) or Cetnarowska (1993), which recognize their intriguing semantic and morphosyntactic properties (but see also the discussion on Borer 2013 in section 2.2.). Generative literature has neglected ZNs in favor of two competing formations: denominal zero-derived verbs and suffix-based nominalizations.

On the one hand, the study of conversion/zero-derivation has mostly dwelled on zero-derived verbs (ZVs), which are fully productive in English. As shown in Clark & Clark (1979) and Rimell (2012), in principle, any

noun can be converted into a verb, provided an appropriate context and essential common ground knowledge shared among the participants: see (2):

(2) a. He **wristed** the ball over the net.   (Clark & Clark 1979: 767)
    b. My sister **Houdini'd** her way out of the locked closet. (Clark & Clark 1979: 784)

By contrast, ZNs show lexical gaps: as Cetnarowska (1993: 19) notes, pairs of phonologically and morphologically similar verbs, as in (3), fail to systematically build ZNs.

(3) a. to permit – the permit   vs. to submit – *the submit (cf. the submission)
    b. to flow   – the flow   vs. to grow   – *the grow (cf. the growth)

On the other hand, studies on nominalization after Chomsky (1970) have mostly been devoted to SNs, which include Chomsky's 'mixed forms' with the suffix *-ing* and Borer's (2013) 'ATK-nominals' (i.e., *-ATion* and Kin), which involve Latinate suffixes (i.e., *-(at)ion*, *-ment*, *-al*, *-ance*). ZNs are usually mentioned for contrastive purposes, to argue, for instance, that they are morphosyntactically and semantically simpler than SNs (but see Alexiadou & Grimshaw 2008, Wechsler 2008, Harley 2009, Fábregas 2014 for additional observations). In her seminal work, Grimshaw (1990) claims that ZNs fail to inherit the event structure of their verbs, since they cannot realize argument structure (AS) on event readings such as in (4), from Borer (2013: 332), where allegedly only the SN is compatible with the internal argument and the purpose clause:

(4) the **importation/*import** *of goods* from China in order to bypass ecological regulations

This paper uses a substantial new resource of 1,000 English ZNs to investigate their semantic and morphosyntactic properties in relation to the lexical semantics of the verbs that they are derived from. While this study is part of ongoing research, this paper draws on some preliminary conclusions from the database to evaluate previous generalizations.

Section 2 provides a summary of previous observations on ZNs proposed in the generative literature. Section 3 describes the new database used in our investigation. The main discussion of our research questions comes in section 4, and we present our conclusions in section 5.

2. Previous generalizations on zero-derived nominals

2.1. Argument structure nominals vs. result and simple event nominals

Grimshaw (1990) distinguishes between two categories of deverbal nominals: i) what she calls 'complex event nominals' and we refer to as *argument*

structure nominals (ASNs), following Borer (2013), and ii) 'result nominals' (RNs) together with 'simple event nominals' (SENs), which are often grouped under the term 'referential nominals' from Borer (2013). The crucial difference between the two groups is that ASNs inherit the event structure of their base verbs, which is reflected in their obligatory realization of AS, while RNs and SENs lack these verbal properties and behave like lexical nouns.

Grimshaw (1990: 49–59) illustrates this contrast with examples as in (5). On its ASN reading, *examination* in (5a) combines with predicates of events like *took a long time*, realizes both internal and external arguments, and its compatibility with the agent-oriented adjective *intentional* confirms that *the instructor's* is the external argument, and not just a possessor. On its RN reading in (5b), the noun is synonymous with *exam*, disallows the internal argument *of the papers*, and combines with predicates of individuals like *was on the table* in (5b). The SEN reading in (5c) resembles the ASN in referring to events, yet, in the absence of the internal argument (which is hierarchically realized before the external one), it behaves like other event-denoting lexical (i.e., underived) nouns such as *trip*; the incompatibility with *intentional* indicates that SENs do not realize external arguments but just possessors.

(5) a. *The instructor's* (intentional) **examination** *of the papers* took a long time. (ASN)
     b. *The instructor's* **examination**/exam (*of the papers) was on the table. (RN)
     c. *The instructor's* (*intentional) **examination**/trip took a long time. (SEN)

The distinction between ASNs and RNs is usually straightforward due to the meaning difference: RNs usually refer to objects/individuals, while ASNs refer to events. The more difficult task is to differentiate between ASNs and SENs, which both denote events. Additional aspectual tests for event structure are usually necessary (see Roy & Soare 2013).

In syntactic models of word formation such as Distributed Morphology (DM; Marantz 1997, 2013, Harley and Noyer 2000, Alexiadou 2001) and Borer's (2013) Exo-Skeletal Model (XSM), the two types of nominalization have found a natural implementation along two patterns of word formation, illustrated in (6) for DM: i) RNs and SENs represent root-based derivations, as in (6a), while ii) ASNs instantiate word-based derivations, as in (6b).

(6) a. [$_{DP}$ D [$_{nP}$ n   [√ROOT]]]   (RN/SEN: no event structure)
     b. [$_{DP}$ D [$_{nP}$ n [**Ext-vP** [$_{vP}$ v   [√ROOT]]]]]   (ASN: with verbal event structure)

While in (6a) *n* – typically realized by a nominalizing suffix like *-ation* – assigns the noun category to the root, in (6b) *n* changes the category of a categorized word (the *v*P or an extended projection Ext-vP of it) into a noun. The two levels of derivation involve crucially different morphosyntactic and

semantic properties. Root-derivation predicts i) negotiation of idiosyncratic meanings of the root in the context of the first categorizing node, leading to polysemy (e.g., both RN and SEN readings are possible in (5b, c)); ii) limited productivity; iii) phonological changes on the root triggered by the suffix; iv) absence of argument structure, which is usually hosted by extended functional structure unavailable in (6a). In contrast, word-level attachment predicts i) compositional meaning derived from the functional structure of the base; ii) apparent (greater) productivity; iii) no phonological changes on the root; iv) realization of argument structure when the appropriate extended projections are available (see (5a)). This distinction is maintained in Borer's XSM to some extent, although she denies lexical categorizers like *n/v* and assumes that roots are indirectly categorized by corresponding functional material in whose nominal/verbal context they are realized.

2.2. Zero-derived nominals from the ASN vs. RN/SEN perspective

Grimshaw (1990: 67) briefly mentions ZNs as always realizing referential nominals (i.e., RNs or SENs) in contrast to *ing*-nominals, which she takes to typically realize ASNs, while ATK-nominals are usually ambiguous between ASN and referential RN/SEN readings in her view.

Borer (2013: ch. 7) is the first extended study of ZNs from the perspective of the ASN vs. RN/SEN dichotomy. She provides three main arguments in support of Grimshaw's thesis and a root-derivation comparable to that in (6a). First, she claims that ZNs do not form ASNs, as illustrated in (4), although she admits that some 'exceptionally' do so: (*ex*)*change*, *release*, *use*, *misuse*, *abuse*, *murder*, *rape* (Borer 2013: 331; cf. Wechsler 2008, Harley 2009, Newmeyer 2009, Lieber 2016).

Second, she shows that ZNs cannot be formed from verbs that involve overt verbalizing suffixes such as *-ize* and *-ify*, which realize the *v* head in (6b), indicating that they cannot instantiate such a structure, as overt suffixes in ATK-nominals usually do:

(7) a. to crystal(l)-**ize**   –   *the crystall**ize**   –   the crystall**iz**ation
    b. to acid-**ify**   –   *the acid**ify**   –   the acid**ifi**cation

Third, Borer highlights the ability of ZNs to exhibit stress shift (cf. Marchand 1969: 378, Kiparsky 1982, Cetnarowska 1993: 34–35, Hurrell 2001), which suggests that their formation triggers phonological changes on the root, as expected in root-derivation, as in (6a):

(8) a. to im**pórt**   –   the **ím**port
    b. to tor**mént**   –   the **tór**ment

In a footnote, Borer (2013: 331, fn. 13) claims that ZNs that exceptionally realize argument structure also block stress shift, which indicates that these may form ASNs and, if they do, they must involve some suprasegmental suffix phonologically strong enough to block stress shift. This can be observed by comparing ZNs of Latinate origin like *reléase* in (9) and *ímport* in (4): the former preserves the verbal stress pattern on the final syllable and can realize argument structure, while the latter shows stress shift and allegedly fails to realize arguments.

(9) the **reléase** *of prisoners of war* by Iraq

Following these arguments, Borer (2013) concludes that ZNs form only RNs or SENs and should receive an analysis as in (10a), where a root like √WALK is indirectly categorized as a noun by a nominal extended projection like D. When the same root appears in the context of a verbal extended projection like T, it becomes a verb, as in (10b). Both structures would correspond to the DM root-derivation in (6a).

(10) a. [DP D [√WALK]]]   b. [TP T [√WALK]]]

Our study casts doubts on this analysis, by showing that, depending on the verb class, quite a few ZNs may realize AS and, importantly, availability of stress shift does not prevent AS-realization, which challenges both Borer's analysis and the dichotomy in (6).

2.3. ZNs in the result-manner verb dichotomy

Another argument that could further support this kind of analysis of ZNs comes from research on verb meaning. Rappaport Hovav & Levin (1998 and subsequent work) argue for a split among verbs depending on the ontology of the root that they lexicalize. They distinguish between result and manner verbs: the roots of the former denote the result state of the event lexicalized by the verb, while the latter specify the manner in which the event is carried out. For transitive verbs we can test the two different types by denying their typical result state: with result verbs one cannot deny the result state that they lexicalize (see use of *dirty* with *to clean* in (11a)); however, the result state that one would typically assume for a manner verb can easily be denied, as in (11b) (cf. further tests in Beavers & Koontz-Garboden 2020).

(11) a. I **cleaned** the tub, #but it is still *dirty*.   (result verb)
      b. I **scrubbed/wiped** the tub, but it is still *dirty*.   (manner verb)

In their discussion of morphosyntactic tests to distinguish the two verb types, Levin and Rappaport Hovav argue that if a verb forms a ZN, and the ZN receives a result (i.e., RN) interpretation, this indicates that it is a

result verb, while if the ZN has an event (i.e., SEN) interpretation, the base is a manner verb (Levin 1993: 8, Levin & Rappaport Hovav 2013). Based on this evidence, they argue that *cut* and *break* are result verbs because their ZNs denote results, while *touch* and *hit* are manner verbs, as their ZNs denote (manners of) events.

Although Levin and Rappaport Hovav do not analyze ZNs, their reasoning would support a root-derivation, to the extent that ZNs are expected to be fully faithful to the ontology of the root: a result root will yield result ZNs, and a manner root will yield event ZNs.

Our investigation of ZNs based on manner and result verbs will show that Levin and Rappaport Hovav's intuition about the faithfulness of ZNs to the root holds for some but not all (sub)classes. We find event readings with result verbs, some of which also realize argument structure, which suggests that ZNs may also be more complex than root-derivations.

2.4. Research questions

Making use of our database of ZNs, as described in section 3, we aim to test the following generalizations made in the literature summarized above:

1) ZNs derived from manner verbs are SENs, and those formed from result verbs are RNs.
2) ZNs do not generally form ASNs; the exceptional AS-ZNs block stress shift.

If we go back to the two DM word-formation patterns in (6), we see exactly how these properties lead to a root-derivation of ZNs: root-derivations allow phonological changes on the root and cannot include argument structure. Co-occurrence of argument structure and stress shift is banned, since they associate with different patterns. Note, though, that the faithfulness of ZNs to the root ontology assumed by Levin and Rappaport Hovav is more restrictive than predicted by (6a), which in principle allows SEN-RN polysemy.

In testing these hypotheses, we also aim to identify meaning regularities among ZNs derived from the same verb class and to find out which classes are likely to build AS-ZNs.

3. The database of zero-derived nominals

The current database contains 1034 ZNs and was created in order to investigate how the meaning and the ASN-potential of ZNs depend on the semantic classes of their base verbs.

The directionality from V-to-N or N-to-V is an unsettled issue in the literature on zero derivation (see Balteiro 2007, Bram 2011 for criteria and literature overviews). First, it should be noted that on a root-based derivation of

ZNs (as in (6a)), the issue of directionality disappears, since the assumption is that both ZNs and ZVs are derived from an underspecified root. Second, for our database we collected nouns zero-related to verbs of particular semantic classes and so, to the extent that we find semantic regularities expected for V-to-N derivations, these support the derivational direction for that class. Thus, (instance of) event or agent readings are expected to originate in verbs rather than in nouns (Marchand 1963; cf. Kisselew et al. 2017). Our use of semantic verb classes potentially offers a better grounded explanation for Marchand's (1963) reasoning in distinguishing between noun groups corresponding to *father* and *bridge*, as bases for ZVs, and those like *cheat* and *look*, as ZNs, but this remains to be confirmed by a closer analysis. Third, the nouns in our database are first attested after, at the same time, or not long before their base verbs.[1] Historical attestation is not always a reliable criterion, but is useful as additional support (Plag 2003, Bram 2011).

We started with a list of ZNs documented by the Oxford English Dictionary (OED), which includes about 2,800 items, from which we extracted the ZNs first attested starting with 1860. This section of the database includes 573 ZNs and was created in order to find out which morphological patterns of ZNs have been formed in recent times. Almost half of these recent ZNs are derived from morphologically complex verbs (with particles or prefixes), which we describe in Section 4.1.1. Furthermore, we checked several semantic classes that would correspond to the result and manner root types on the basis of two sources. We started with Levin's (1993) verb classes by checking, for each verb, if it had a corresponding ZN, according to the OED. Given that many of Levin's classes are very numerous, with most verbs not forming ZNs, we then turned to Irmer's (1972) semantic lists of ZNs and mapped them onto Levin's relevant classes, some of which we present in Section 4.

For each ZN entry we recorded various information from the OED online: date of attestation of the noun/verb, (Latinate/Germanic/Anglo-Saxon etc.) etymology,[2] frequency of the noun/verb, whether it involves stress shift, and their possible readings. We further added the verb class recorded for the base verb in VerbNet and searched natural text corpora for possible examples of

---

1 About 2.6% of the database represent ZNs whose attestation date is more than two decades earlier than that of the verb and they were included because the OED or Irmer (1972) records them as ZNs.

2 Our database includes originally borrowed ZNs like *change* and *collapse*. If there is a corresponding verb belonging to a semantic verbs class for which English shows ZN formation, we consider such old borrowings to be synchronically analyzed as ZNs. This assumption would be worth testing experimentally, as in Darby (2015).

ZNs realizing verbal arguments. The corpora we searched are the Corpus of Contemporary American English (COCA), News on the Web (NOW), and Corpus of Global Web-based English (GloWbE), all available at www.english-corpora.org (Davies 2008–, 2013, 2016–). VerbNet is an extended version of Levin's (1993) verb classes with online access at https://verbs.colorado.edu/verb-index/vn3.3/ (Kipper Schuler 2005). It includes at least 108 verb classes, some of which may comprise a few hundred words or less than a dozen. Many classes have several subclasses, following Levin's model. Our database includes ZNs derived from 72 such verb classes, 21 of which have only one ZN, and 177 ZNs without a VerbNet class.[3] Many verbs are cross-classified and, in these cases, we focused on the verb class(es) that best represented the meaning of the ZN. Nevertheless, about 180 ZNs in our database relate to more than one verb class.[4]

There is no exhaustive description as to which of these verb classes lexicalize manner or result, and one may even need to distinguish between their subclasses in this respect (e.g., Levin's result *murder* verbs vs. manner *poison* verbs, both verbs of killing). Here we report on the verb classes in (12), which have been discussed in the literature as lexicalizing result and manner, although polysemy covering both types or a combination of both cannot be excluded (see Levin & Rappaport-Hovav 2013, Beavers & Koontz-Garboden 2020 for discussion). Individual examples and discussion will be offered in Section 4.

(12) a. **Result verb classes:** verbs of change of state, psych verbs, verbs of cutting, verbs of killing, morphologically complex verbs
      b. **Manner verb classes:** verbs of motion, verbs of communication, verbs of emission, verbs of contact

An important task was to categorize the different ZN senses reported in the OED in relation to the meaning of the verb and the typical interpretations of derived nominals. To define them, we follow Lieber (2016:18), with small adjustments. These readings belong to two categories: event-related and participant-related. The former denote the eventuality of the base verb, as in (13a), and the latter some participant in the eventuality, as in (13b).[5]

---

3   For the verbs we could not find in VerbNet we sometimes inserted the verb class of a close enough synonym. For morphologically complex verbs we checked the class of the main verb.

4   In future research we aim to draw statistically relevant implications from this database.

5   We also collected readings about measure (e.g., *melt* 'quantity of metal melted at one operation') and location (e.g., *walk* 'place or path for walking'), which are rare but would belong to the participant-related group.

(13) a. **Event-related ZN readings:**

   event (the action of V-ing): *fall, collapse, murder, run, walk*

   instance (an act of V-ing): *click, kiss, plunk, pat, pop*

   state (the state of V-ing/being V-ed): *daze, shock, sorrow, concern, dread*

   result state (the state brought about by having V-ed): *collapse, meltdown, decrease*

   b. **Participant-related ZN readings:**

   result entity/product (the thing that is produced by V-ing): *cut, chant, crack, bruise*

   agent (the person who V-s): *cook*, *guide*, *kick* (as one who kicks), *dispatch* (agency)

   cause (the thing that V-s): *wilt* (as disease), *surprise, wonder, trouble*

   instrument (a thing to V with): *nudge, drill, smell* (as the sense)

In Table 1 we offer a few simplified entries from our database, with their classified OED senses. The event reading best matches the meaning of the verb, and it is on this reading that ZNs should realize AS (with verbal event structure). The instance reading resembles SENs and may include more idiosyncratic readings such as the manner of the event (e.g., *His walk was cocky*: OED). The result entity/product reading best relates to RNs and also comprises several types of meanings exemplified in Table 1: e.g., a created product such as *cut*, a patient (which was not created but underwent a change of state via the event: see *change*, *melt*, *transport*), or even something abstract coming about via the event, such as *walk* 'procession'.

Tab. 1: Samples of ZNs with their readings from the database.

ZN	Verbnet class	Event	Instance	(Result) State	Product/ Result Entity	Agent/ Instr/ Cause
murder	Verbs of Killing	The action of killing	An act of killing	no	no	no
melt, n3	Verbs of Change of State	The action of melting	no	no	A substance which has melted	no
change	Verbs of Change of State	The action of substituting one thing for another	An instance of this	An alteration in the state or quality of something	Something that may be substituted for another thing	no

cut, n2	Verbs of Cutting	no	Act of cutting; a stroke or blow with a sharp-edged instrument	no	Product of cutting; shape cut off; a piece cut off	no
wonder	Psych verbs	no	no	The state of mind in which this emotion exists	no	Some-thing that causes aston-ish-ment; a won-derful thing
transport	Psych-Verbs; Verbs of Sending and Carrying	The action of carrying/ conveying a thing or person from one place to another	no	The state of being 'carried out of oneself'; ecstasy	A transported convict (rare)	A means of trans-porta-tion
row-over	Verbs of Sending and Carrying [row]	no	An instance of rowing over	no	no	no
walk	Verbs of Motion	The action of traveling or wandering	An act of traveling or wan-dering	no	A journey; a proces-sion	no

For the database, we did not apply distributional tests to systematically delimit these readings from one another, and in what follows we focus on broad semantic distinctions relevant for our questions: e.g., (instance of) *event* vs. *state* vs. *product/result entity* vs. *agent/instrument/cause*. In classifying these readings, we had to assess which senses from the OED best suited these groups. We did not classify all OED senses; we looked for those that came closest to our meaning classes and recorded the most common ones when several qualified for the same class. Sometimes the OED offers straightforward definitions: see the event, instance or even state readings in Table 1. However, this was not always the case and, working on the database,

we developed several conventions: if a ZN was defined by an *-ing* nominal (which usually refers to processes), we took this as a clear event reading; if it was translated by an ATK-nominal, we further checked whether it was an event or a state, or even both. For most senses, we used our intuitions, and critical cases were discussed among the four of us.[6] However, inaccuracies cannot be entirely excluded, as is common with human annotation. The data reported here have gone through an additional cycle of verification.

Coming back to the patterns in (6), we note that the event-related readings of ZNs may instantiate either root- or word-based derivation: the former correspond to SENs, the latter to ASNs. We assume that participant-related readings are all root-derived, since they do not represent a typical pattern of the zero suffix, as the description of our database will confirm.

4. Testing previous hypotheses

4.1. Semantic verb classes and ZN readings

Our first question concerns Levin's (1993) and Levin & Rappaport Hovav's (2013) hypothesis that result ZNs point to result verbs and event ZNs to manner verbs. Table 2 is a summary of our verb classes specifying the amount of ZNs for each class and the approximate percentages for the different readings. For each verb class, we marked in bold the values for the readings that appear with more than 50% of the corresponding ZNs, although one should be alert that some verb classes are richer/poorer than others (see last column with total numbers).[7] All verb classes (except for psych) exhibit a majority of event ZNs, but some also present interestingly high values for other readings that we discuss below. We start with a description and examples of each class and assess Levin and Rappaport Hovav's hypothesis at the end. In the interest of space, we list lexemes as examples with further clarification for the less straightforward ones. The OED offers plenty of contexts for the different senses.

---

6 We are not native speakers but judging the classification of these senses is not a matter of native speaker intuition but of linguistic awareness in determining the linguistically relevant class for each sense.

7 We do not aim to draw any statistical conclusions from these numbers at the moment, but the contrasts are clear enough for our discussion of previous hypotheses.

Tab. 2: Overview of ZN readings for result/manner verb classes.

Verb class	ZN reading	(Instance of) Event	(Result) State	Product/ Result entity	Agent/ Cause/ Instrument	Total of ZNs
R E S U L T	Change of state	**70%**	30%	**59%**	20%	115
	Psych verbs	**51%**	**66%**	25%	40%	73
	Verbs of cutting	**70%**	4%	**70%**	30%	27
	Verbs of killing	**100%**	10%	50%	20%	10
	Complex verbs	**80%**	13%	44%	26%	285
M A N N E R	Verbs of motion	**95%**	4%	18%	35%	75
	Communication	**88%**	7%	**77%**	18%	56
	Verbs of emission	**74%**	9%	**98%**	30%	87
	Verbs of contact	**100%**	16%	**75%**	44%	32

4.1.1. Result verbs

We considered four semantic verb classes for result verbs: change of state, psych verbs, verbs of cutting, and verbs of killing. We added morphologically complex verbs, as their prefixes and particles are standardly assumed to contribute a result component to the main verb (see McIntyre 2007 for an overview; Harley 2008). Importantly, this group includes manner verbs with particles (which were excluded from the manner verb classes), since we expected them to resemble result verbs more. Result verbs with particles were counted twice: both in their result verb class and as morphologically complex verbs.

Change of state verbs are the typical result verbs discussed by Levin and Rappaport Hovav. Their richest ZN group denotes events (e.g., *change*, *advance*, *(de)freeze*, *collapse*, *melt*, *fall*), but they also build a large number of result entity ZNs (e.g., *rot* 'rotten material', *broil* 'broiled meat', *roast* 'roasted meat', *split* 'narrow break', *shatter*, *blossom*, *crack*). Some realize a result state reading (e.g., *rot* 'state of being rotten', *collapse*, *meltdown*, *blossom*, *degrade*) and a few refer to instruments or causes (e.g., *drain*, *rot* and *wilt* as diseases, *soak/ scald* 'liquid used for V-ing', *compress* 'machine that V-s'). Iordăchioaia (2020) shows that some subclasses present sharper tendencies. On the one hand, verbs of breaking (*break, crack, rip, split*), bending (*bend, fold, crinkle, stretch*) and cooking (*bake, fry, roast, toast, boil*) display a large number of result ZNs, while their event-like ZNs lack event structure, which typically comes from light verbs (see *to **take a break*** vs. **the break **of the glass***). On the other hand, verbs of calibratable (*rise,*

fall, decrease, drop) and entity-specific (*melt, thaw, decay, rot*) change of state often display event ZNs, which may realize argument structure.

Psych verbs are not discussed in Levin and Rappaport Hovav's work on manner vs. result, but their roots usually express states. For our distinction between manner and result, their root meaning is closer to the latter. We considered all psych verbs, whether causative or non-causative; VerbNet mentions three subclasses called *amuse, admire* and *marvel*.[8] As expected, most psych verbs form stative ZNs (e.g., *disgust, puzzle, dread, delight, concern, torment, trouble*). Event readings are possible when the verbs allow them (e.g., *support, worship, sorrow, insult*); some relate to non-psych readings of the verbs (see *transport* in Table 1 but also *ruffle* 'slipping playing cards rapidly through the fingers', *exhaust* 'expulsion of combustion products from a combustion engine) or are ambiguous between psych and non-psych readings (*torment, stir, refresh*). Result entity readings are abstract and difficult to distinguish from result states (e.g., *grudge, wound, hurt, respect, esteem, support*). Many ZNs acquire such meanings on non-psych readings of the verbs (e.g., *lull* and *mourn* referring to sounds, *impress* as a mark, *insult* 'words produced by insulting', *revolt* 'rebellion', *muddle* 'mistake arisen from confusion', *exhaust* 'products expelled from an internal-combustion engine'). In addition, psych verbs form a substantial number of ZNs that refer to the stimulus/cause of the psych state: e.g., *sorrow, dread, insult, surprise, concern, torment, haunt* as 'things that V the experiencer'.

Verbs of cutting are not numerous, but they are typical result verbs. They mostly derive ZNs with event (e.g., *mow, cut, grind, squash, clip, hack, carve*) and result entity readings (e.g., *slit* and *cut* as incisions, *slice, bore* 'hole made by boring', *chop, chip, shred* as 'pieces resulting from V-ing'). The only ZN close to a result state reading is *squash* on the OED interpretation 'the fact of some soft substance being crushed' or in the expression *to go to squash* 'to become ruined'. A few ZNs denote instruments: e.g., *slice, bore, drill, chop, hack*.

Verbs of killing are only a few, but they all form event ZNs: *kill, murder, massacre, slaughter, rubout, dispatch, shoot, reshoot, overkill,* and *overshoot*. Only *massacre* resembles result states as 'great destruction'. Result entity ZNs are *shoot* 'result of game-shooting', *overshoot* 'result of overshooting',

---

8   Iordăchioaia (2019) argues that even nominals corresponding to causative psych verbs are root-derived, so we do not expect differences in ZN interpretation among the three subclasses, but this may be worth further study.

reshoot 'session of photography', and *dispatch* 'dispatched message', often without involving killing. For agent/instrument readings, *dispatch* may refer to an agency transmitting goods, and *overkill* to the overkilling capacity of nuclear weapons.

A great number of ZNs originate in **complex verbs with prefixes and particles,** most of which are attested over the last 150 years. The examples in (14) illustrate some prefixes and particles that yield larger numbers of ZNs.

(14) a.   *re-* (±70 ZNs): rewrite, re-read, re-mark, re-edit, redial, redo, rewind, rewire, retry

   b.   *de-* (±10 ZNs): declutter, decoke, defreeze, declaim, detox, dequeue, derequisition

   c.   *-out* (18 ZNs): bailout, buildout, close out, drop-out, opt-out, pinch-out, read-out

   d.   *-off* (13 ZNs): die-off, ring-off, lay-off, rake-off, rip-off, row-off, sign-off, walk-off

   e.   *over-* (±30 ZNs): overbid, overcall, overcross, overdrive, overfill, overkill, overfit

   f.   *under-* (7 ZNs): underbid, underbite, undershoot, underspend, understeer, understudy

Some of these already appear in the different result verb classes discussed above, but here we consider them separately. Their most frequent reading is eventive: e.g., *re-run*, *rewrite*, *redial*, *detox*, *makeover*, *trade-up*, *trade-in*, *pat-down*, *take-back*, *transport*, *overspill*, *overtake*. A few have result state readings: *meltdown*, *disconnect*, *die-off*, *distrust*, *comeback* 'return to success', *lay-off* 'spell of relaxation', *interlock* 'condition of being interlocked'. Another rich class is that of result entity ZNs: e.g., *premix*, *admix*, *frame-up*, *miscue*, *takeaway* 'key point to remember', *under-bid*, *download*, *upload*, *bailout*. Among the agent/instrument readings we find *detox* as a center, *rewrite* as a department at a newspaper, *screw-up*, *rip-off* for persons, or *redial*, *undelete* for functions of a phone/computer program.

4.1.2. *Manner verbs*

For manner verbs we considered verbs of motion, verbs of communication, emission, and contact. Almost all **verbs of motion** form event ZNs – their percentages are very high in Table 2 (e.g., *run*, *follow*, *rush*, *trot*, *float*, *race*, *gallop*, *hop*, *sprint*, *jump*). Exceptions are *enter* on the instrument reading of 'enter key' on the keyboard, *Flit* as the name of the insecticide, *skirr* (referring to a sound), and *steer* ('a lead, piece of advice'). Result state readings appear with *coil*, *turn* and *twist*, which all refer to a twisted condition and belong to the *rotate* subclass. Some ZNs have result entity readings such as *step* but most are remote to motion: see *scramble* as 'a mixed dish', *float* (as 'flood'), *twist* 'thread composed of fibres', *wind* 'a curved form', *skitter*, *shuffle*, *scuttle* as specific sounds of the different motions. For the agent/instrument reading we find

a considerable number of ZNs: see *steer*, *enter*, and *Flit* above, *enter* 'entrance', *trudge* 'a trudger', *trot* 'toddling child', *skip* 'one who defaults in payment', *swim* 'part of liquid that floats above the sediment', *creep* 'creeping fellow'.

ZNs derived from **verbs of communication** display a similar picture with respect to event, result state, and agent/instrument readings: the great majority refer to events (e.g., *blubber*, *call*, *twitter*, *wail*, *roar*, *moan*, *stutter*, *declaim*, *gurgle*), only a few to result states (i.e., *twitter* 'state of agitation', *wail* 'state of woe', *roar* 'state of extreme amusement', and *moan* 'state of lamentation'), and a few more refer to causes/instruments (e.g., *scream* 'cause of laughter', *yell* 'very amusing thing', *grouse* 'reason for grumbling', *teach* 'something that teaches'). Unlike ZNs derived from verbs of motion, however, those based on verbs of communication frequently yield result entity readings, which name the sounds or some substance that accompany what is communicated (e.g., *scream*, *blubber*, *yell*, *chatter*, *whimper*, *babble*, *bark*, *gabble*, *gurgle*, *murmur*, *mutter*). In this respect, they are similar to verbs of emission, and many of these verbs cross-classify between the two classes.

ZNs derived from **verbs of emission** are comparable to those formed on verbs of communication, but their result entity group is even richer than the event group. Almost all verbs of emission build result entity ZNs: next to the ones named under verbs of communication, we find those related to light emission (e.g., *shine*, *glitter*, *gleam*, *sparkle*, *flash*), substance emission (*gush*, *spew*, *puff*, *dribble*), and smell emission (*reek*, *stink*). The two ZNs that do not denote result entities are *ooze* and *smell*, which resemble more the agent/cause reading as 'something that Vs' than the result entity. Other ZNs that belong to the agent/instrument/cause group and would qualify as the 'emitter' for these verbs are: *flash*, *shine*, *flicker*, *thunder*, *whistle*, *ring* and *chime* 'sets of bells', *stink* 'a stinkard', *spout* 'object that discharges liquid'. Under event ZNs we find, among others, *bang*, *blast*, *crash*, *crackle*, *cry*, *flash*, *flicker*, *gleam*, *twinkle*, *dribble*, *ooze*, *puff*, and *spew*.

All the **verbs of contact** form event ZNs (e.g., *hit*, *strike*, *caress*, *rub*, *nudge*, *kiss*, *pat*, *touch*, *kick*). Only five ZNs acquire special meanings that resemble result states: *kick* 'feeling of marked enjoyment', *dash* 'discouragement', *rub* 'injury inflicted on the feelings of another', *pinch* 'crisis', and *belt* 'a thrill'. Some ZNs receive result entity interpretations (e.g., *crack*, *thump*, *knock* for sounds, *sting* 'wound', *pinch* 'a bend in the brim of a hat', *graze* 'superficial wound'), and a few are used for agents/instruments: e.g., *nudge* 'reminder', *pinch* 'mean person', *kayo* 'knock-out blow', *kick* 'kicker'.

4.1.3. Root ontology and ZNs

Our research question is whether the formation of event or result entity ZNs correlates with a manner and result verb, respectively, as first hypothesized

by Levin (1993). From our data we can say that change of state verbs and verbs of cutting are well-behaved result verbs in exhibiting a great proportion of result entity ZNs, while verbs of motion are well-behaved manner verbs in showing a high number of event ZNs and very few result entity ZNs. Yet, there is no explanation for the high frequency of event ZNs derived from result verbs (i.e., change of state verbs, verbs of cutting, and verbs of killing) or for the great amount of result entity ZNs with the manner verb classes involving communication, emission, and contact.

Many of the verbs exemplified above are polysemous between different classes, and this is reflected in the interpretations that their ZNs receive. For instance, the result state readings that occur with ZNs derived from manner verbs mostly originate in psych uses of the verbs and not some result state of the event, as is typical for change of state verbs. If we control for all these aspects, we should obtain a cleaner picture. Yet, the unexpected event and result entity ZNs represent such a great majority for the result and manner verb classes under discussion that further research should closely consider finer-grained subclasses of these verbs to see how the expected patterns may be confirmed and how the other readings could be explained. As Iordăchioaia (2020) shows, ZNs derived from some change of state subclasses present a higher proportion of result readings than others (see section 4.1.1.), indicating that Levin and Rappaport Hovav's hypothesis may indeed hold for such subclasses. Moreover, Melloni (2011: sec. 4.3.) groups verbs of emission under non-prototypical creation verbs, in that they yield product readings even for suffixed nominalizations. We could likewise argue that the root ontology of verbs of communication, emission and contact provides information about the product of the verb's action to explain the high number of result ZNs. It is unclear, though, which part of the root ontology this should be: Levin and Rappaport Hovav's manner vs. result ontology may have more general implications, while information about products in emission, communication and contact verbs may be more local.

In conclusion, our data collection does not entirely support Levin's (1993) and Levin & Rappaport Hovav's (2013) hypothesis. First, the high frequency of event ZNs with all non-psych result verb classes shows that an event reading of a ZN is no certain indication of a manner verb. Second, some manner verb classes exhibit a high amount of result entity ZNs, which indicates that a result entity ZN is not a guarantee for a result verb either. Future study should disentangle the different ontological aspects that interact in this process.

4.2. Argument structure in ZNs

The second generalization in the literature that we want to test is whether ZNs indeed fail to realize argument structure and whether the allegedly few

exceptional AS-ZNs block stress shift, as Borer (2013) claims. We focus on ZNs derived from morphologically complex verbs, as they are conclusive regarding both aspects. See Iordăchioaia (2020) for evidence on ZNs based on change of state verbs, which may realize ASN readings as in (6b), by contrast to ZNs derived from psych verbs, which are all argued to instantiate root-derivations as in (6a).

Morphologically complex verbs are interesting for our purposes, since they involve two sub-events, most likely realized by different roots: on the one hand, the verb root introduces a manner (possibly causative) event; on the other hand, the prefix/particle adds a result similar to the result state of result verbs. Following Rappaport Hovav & Levin's (1998) approach to argument realization, each of these sub-events license syntactic arguments, and we expect event ZNs that inherit the verb's event structure to also realize these arguments syntactically.

Our corpus investigation reveals that most of the event ZNs derived from complex verbs appear in contexts where they realize a semantic internal argument.[9] Given that we have not tested these ZNs for event structure via introspection, we cannot guarantee that all of them indeed represent ASNs. However, even in corpora, we find plenty of examples in which the ZN realizing the internal argument is also modified by an event-modifying adjective, which is typically taken to indicate the presence of verbal event structure. In (15) we illustrate such ZNs with prefixes, in (16) with prefixed particles, and in (17) with postposed particles (see also Iordăchioaia to appear).

(15) a. *The **ongoing rewrite** of the city's antiquated zoning code* will help […]. (NOW)
b. [I] witnessed *the **constant replay** of this atrocity* unfolding on television. (NOW)
c. [S]uch bills are useful for *the **ongoing declutter** of the legislative landscape*. (NOW)
d. It has been fitted to enable *the **continuous discharge** of treated sewage*. (NOW)
e. [T]he surgery will also stop *the **constant increase** of pain*. (GloWbE)

(16) a. A malware attack took place during *a **recent upload** of information* (NOW)
b. Businesses […] shouldn't encourage *the **continual download** of […] apps* (NOW)
c. [It] is effectively powerless to stop *a **potential override** of its decision*. (NOW)
d. Ejectives […] require *a **quick outrush** of air*. (NOW)
e. UK is heading for *a **persistent undershoot** of the 2 % inflation target*. (GloWbE)

---

9   That is, 65% of the event ZNs and 51% of all complex ZNs.

(17)   a.   *the **ongoing bailout** of the European banking system* (NOW)

         b.   *[a] story of **continuous rip-off** of the people [...] by the insurance industry* (NOW)

         c.   *the **ongoing meltdown** of Greenland and the Arctic* (NOW)

         d.   *speaking of **continuing takeover** of the world by a few not so good banks* (GloWbE)

         e.   *Syracuse's **recent cutback** of its Posse Leadership Scholarship programs* (NOW)

All three groups of complex ZNs exhibit ASN-readings, but those with postposed particles form a richer class, as they show a greater variety. Prefixes and prefixed particles are not as diverse, but some are very frequent such as *re-* and *over-*. It would be worth investigating the theoretical consequences of these phenomena.

We now come to the question of stress shift. Borer (2013) speculates that its presence should block argument realization, as also predicted by the two patterns of word formation in (6). Our data, however, provides clear evidence against this generalization. All the ZNs illustrated in (15) to (17) exhibit stress shift from a final verbal pattern to an initial nominal pattern with the exception of *declutter*.

This observation poses a serious challenge not only to Borer's claim but also to the DM reasoning promoted in Marantz (2013) on the parallel mapping of the locality domains of morphophonology with those of morphosyntax and semantics in the two word-formation cycles in (6). The mismatch between the phonological changes on the root (i.e., stress shift) and the availability of event structure in ZNs as in (15)–(17) leads us to conclude that one may not want to merge the locality cycles of morphosyntax and semantics with those of morpho-phonology/allomorphy, but rather keep them apart, in full accordance with the separationist position, generally embraced in DM by featuring null affixes (vs. Borer's XSM system).

A possible solution within DM, however, is the pruning rule that Embick (2010: 58–60) employs to account for root allomorphy triggered by Tense inflection, despite intervening *v*P and AspectP: e.g., *sing – sang*. Embick argues that allomorphy is possible because the intervening nodes are null: when an overt verbalizer is present, such as *-ize*, root allomorphy is excluded and regular inflection appears instead: e.g., *real-ize – realized*. This predicts that ZNs that include overt verbalizers should not undergo stress shift. Borer showed that ZNs cannot be formed from overtly verbalized bases (see (7)). However, some ZNs with overt *-en* and postposed particles seem to be possible, as in (18), and our consultants confirm that they bear the same stress pattern as their base verbs, validating Embick's proposal (see Iordăchioaia to appear).

(18)   New towels were provided every day, as well as *a straighten up of the bed*.[10]

Even if this solution can solve this puzzle, the distinction between the two word-formation patterns in (6) also predicts that we may find cases where derived nominals block allomorphy as ASNs, but allow it as RNs/SENs, a mismatch we are unaware of and find counterintuitive, since these readings are usually associated with the same morphophonology.

5. Conclusions

In this paper we have empirically investigated two aspects in the behavior of ZNs and their implications for theories of nominalization and word formation, namely, 1) how their interpretations relate to the base verb class and the manner-result ontology; and 2) to what extent they may realize argument structure and also conform to the theoretical expectation that argument realization should block stress shift.

First, the manner-result dichotomy among verb classes is not perfectly mapped onto the interpretation of the corresponding ZNs, and further study is needed in order to fine-grain the various interfering factors. Second, we have closely investigated complex ZNs and have shown that many display ASN uses in natural text corpora. Most of these also involve stress shift, casting doubts on the isomorphic mapping between morphophonology and morphosyntax/semantics in word formation, even though solutions may be available, especially in DM.

To conclude, we expect a model like DM to be able to account for this diversity of ZNs. If zero is a possible spell-out of the nominalizer *n* in (6), similarly to -*ing* and ATK-suffixes, we can implement all these properties (see Alexiadou & Grimshaw 2008). However, we do not see how Borer's (2013) model, in which ZNs are root categorizations in context in the absence of any suffix, could account for the similarity between ZNs and SNs in building ASNs, unless she posits a zero suffix, as suggested in a footnote (Borer 2013: 331).

Acknowledgements

---

10 Taken from the web (https://www.tripadvisor.co.uk/ShowUserReviews-g186338-d285524-r724451834-QueensHotel-LondonEngland.html) and found acceptable by our consultants.

References

Alexiadou, Artemis. 2001. *Functional structure in nominals: Nominalization and ergativity*. Amsterdam & Philadelphia: Benjamins.

Alexiadou, Artemis & Jane Grimshaw. 2008. Verbs, nouns and affixation. In Florian Schäfer (ed.), *Working papers of the SFB 732: 01*, 1–16. University of Stuttgart: OPUS.

Balteiro, Isabel. 2007. *The directionality of conversion in English. A dy-synchronic study*. Bern: Lang.

Beavers, John & Andrew Koontz-Garboden. 2020. *The roots of verbal meaning*. Oxford: Oxford University Press.

Borer, Hagit. 2013. *Taking form. Structuring sense*. Volume III. Oxford: Oxford University Press.

Bram, Barli. 2011. *Major total conversion in English*. Wellington, New Zealand: Victoria University of Wellington dissertation.

Cetnarowska, Bożena. 1993. *The syntax, semantics and derivation of bare nominalizations in English*. Catowice: Uniwersytet Slaşki.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick Jacobs & Peter Rosenbaum (eds.), *Readings in English Transformational Grammar*, 184–221. Waltham, MA: Ginn.

Clark, Eve V. & Herbert H. Clark. 1979. When nouns surface as verbs. *Language* 55(4). 767–811.

Darby, Jeannique. 2015. *The processing of conversion in English: Morphological complexity and underspecification*. Oxford, UK: University of Oxford dissertation.

Davies, Mark. 2016. *Corpus of News on the Web (NOW)*: 10 billion words from 20 countries, updated every day. https://www.english-corpora.org/now/

Davies, Mark. 2013. *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE)*. https://www.english-corpora.org/glowbe/

Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA): One billion million words*, 1990–2019. https://www.english-corpora.org/coca/

Embick, David. 2010. *Localism versus globalism in morphology and phonology*. Cambridge, MA: MIT Press.

Fábregas, Antonio. 2014. Argument structure and morphologically underived nouns in Spanish and English. *Lingua* 141: 97–120.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press.

Harley, Heidi. 2009. The morphology of nominalizations and the syntax of vP. In Anastasia Giannakidou & Monika Rathert (eds.), *Quantification, definiteness, and nominalization*, 321–343. Oxford: Oxford University Press.

Harley, Heidi. 2008. The bipartite structure of verbs cross-linguistically. In Thaïs Cristófaro Silva & Heliana Ribeiro de Mello (eds.), *Conferencias do V Congresso Internacional da Associacao Brasileira de Linguistica*, 45–84. Belo Horizonte, Brazil: ABRALIN and FALE/UFMG.

Harley, Heidi & Rolf Noyer. 2000. Formal versus encyclopedic properties of vocabulary: Evidence from nominalizations. In Bert Peters (ed.), *The lexicon-encyclopedia interface*. Amsterdam: Elsevier Press.

Hurrell, Esther. 2001. *The morphology and phonology of English noun-verb stress doublets: Base-driven lexical stratification, prefixes and nominalisation*. Edinburgh: University of Edinburgh dissertation.

Iordăchioaia, Gianina. To appear. Event structure and argument realization in English zero-derived nominals with particles. In Andreas Trotzke & George Walkden (eds.), CGSW 34. Proceedings of the 34th Comparative Germanic Syntax Workshop, special issue of Nordlyd 44.1.

Iordăchioaia, Gianina. 2020. Categorization and nominalization in zero nominals. Artemis Alexiadou & Hagit Borer (eds.), *Nominalization: 50 Years on from Chomsky's* Remarks. Oxford: Oxford University Press.

Iordăchioaia, Gianina. 2019. The root derivation of psych nominals: Implications for competing overt and zero nominalizers. *Bucharest Working Papers in Linguistics* 21(2). 57–79.

Iordăchioaia, Gianina & Martina Werner. 2019. Categorial shift via aspect and gender change in deverbal nouns. *Language Sciences* 73. 62–76.

Irmer, Roland. 1972. *Die mit Nullmorphem abgeleiteten deverbalen Substantive des heutigen Englisch*. Tübingen: University of Tübingen dissertation.

Kiparsky, Paul. 1982. From cyclic phonology to lexical phonology. In Harry van der Hulst & Norval Smith (eds.), *The structure of phonological representations*, 131–175. Dordrecht: Foris.

Kipper Schuler, Karin. 2005. *VerbNet: A Broad-coverage, Comprehensive Verb Lexicon*. University of Pennsylvania dissertation.

Kisselew, Max, Laura Rimell, Alexis Palmer & Sebastian Padó. 2016. Predicting the direction of derivation in English conversion. In *Proceedings of the ACL SIG-MORPHON workshop*, 93–98. Berlin.

Levin, Beth. 1993. *English verb classes and alternations*. Chicago: University of Chicago Press.

Levin, Beth & Malka Rappaport Hovav. 2013. Lexicalized meaning and manner/result complementarity. In Boban Arsenijević, Berit Gehrke & Rafael Marín (eds.), *Studies in the composition and decomposition of event predicates*, 49–70. Dordrecht: Springer.

Lieber, Rochelle. 2016. *English nouns. The ecology of nominalization*. Cambridge: Cambridge University Press.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *UPenn Working Papers in Linguistics* 4(2). 201–225.

Marantz, Alec. 2013. Locality domains for contextual allomorphy across the interfaces. In Ora Matushansky & Alec Marantz (eds.), *Distributed Morphology today: Morphemes for Morris Halle*, 95–116. Cambridge, MA: MIT Press.

Marchand, Hans. 1969. *The categories and types of present-day English word formation*. Munich: Beck.

Marchand, Hans. 1963. On a question of contrary analysis. *English Studies* 44(1). 176–187.

McIntyre, Andrew. 2007. Particle verbs and argument structure. *Language and Linguistics Compass* 1(4). 350–397.

Melloni, Chiara. 2011. *Event and result nominals. A morpho-semantic approach*. Bern: Lang.

Newmeyer, Frederick J. 2009. Current challenges to the lexicalist hypothesis. An overview and a critique. In William D. Lewis, Simin Karimi, Scott O. Farrar & Heidi Harley (eds.), *Time and again*, 91–117. Amsterdam: Benjamins.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Rappaport Hovav, Malka & Beth Levin. 1998. Building verb meanings. In M. Butt & W. Geuder (eds.), *The projection of arguments: Lexical and compositional factors*, 97–134. Stanford: CSLI Publications.

Rimell, Laura D. 2012. *Nominal roots as event predicates in English denominal conversion verbs*. New York, NY: University of New York dissertation.

Roy, Isabelle & Elena Soare. 2013. Event-related nominalizations. In Gianina Iordăchioaia, Isabelle Roy & Kaori Takamine (eds.), *Categorization and category change*, 123–152. Newcastle upon Tyne: Cambridge Scholars Publishing.

Wechsler, Stephen. 2015. *Word meaning and syntax: Approaches to the interface*. Oxford: Oxford University Press.

Wechsler, Stephen. 2008. A diachronic account of English deverbal nominals. In Charles B. Chang & Hannah J. Haynie (eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 498–506. Somerville: Cascadilla Proceedings Project.

Gianina Iordăchioaia
Universität Stuttgart
Institut für Linguistik: Anglistik
Keplerstraße 17
70174 Stuttgart
gianina@ifla.uni-stuttgart.de

Susanne Schweitzer
New York University
Graduate School of Arts and Science
70 Washington Sq S
10012 New York City
susanneschweitzer@gmx.net

Yaryna Svyryda
Universität Stuttgart
Institut für Linguistik: Anglistik
Keplerstraße 17
70174 Stuttgart
yaryna.svyryda@gmail.com

María Camila Buitrago Cabrera
Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung
Pfaffenwaldring 5b
70569 Stuttgart
buitrama@ims.uni-stuttgart.de

Maria Bloch-Trojnar

Simple Event nominals with Argument Structure? – Evidence from Irish deverbal nominalizations[1]

Abstract: Deverbal nominals in Irish support Grimshaw's (1990) tripartite division into complex event (CE-), simple event (SE-) and result nominals (R-nominals). Irish nominals are ambiguous only between the SE- and R-status. There are no CE-nominals containing the AspP layer in their structure. SE-nominals (also found in Light Verb Constructions) are number-neutral and incapable of pluralizing and are represented as [nP[vP[Root]]]. R-nominals are devoid of the vP layer and behave like ordinary nouns. The Irish data point to v as the layer introducing event implications and the vP or PPs as the functional heads introducing the internal argument (Alexiadou and Schäfer 2011). Event denoting nominals in Irish can license the internal argument but aspectual modification and external argument licensing are not possible (cf. synthetic compounds in Greek (Alexiadou 2017)), which means that, counter to Borer (2013), the licensing of Argument Structure need not follow from the presence of the AspP layer.

Keywords: Complex Event Nominals, Simple Event Nominals, Argument Structure, AspectP licensing, Irish, deverbal nominalization

1. Introduction

In this paper we shall assess the validity of competing proposals advanced within the theoretical confines of the Distributed Morphology framework (Halle and Marantz 1993, 1994; Marantz 1997, 2013; Harley and Noyer 1999, 2000; Embick and Noyer 2007; Embick 2010) with respect to the category which traditional Irish grammars refer to as the verbal noun (VN). It will be demonstrated that the morpho-syntactic properties of deverbal nominalizations in Irish provide support for a tripartite division into complex

event nominals, simple event nominals and result nominals, as put forward in Grimshaw (1990) and refined by Alexiadou (2017a). In Section 2 a theoretical question will be raised concerning internal argument licensing in view of the distinction into complex- and simple-event nominals. In section 3 the properties of Irish VNs in their nominal function will be scrutinized in different contexts to show that they are ambiguous only between the simple event and result status with the proviso that simple event nominals are equated with a subtype of AS-nominals discussed in Alexiadou (2017a). In section 4 we will summarize the theoretical ramifications of the proposed analysis. Namely, that counter to Borer (2003, 2013, 2014), we need to sever the licensing of Argument Structure from the presence of an aspectual reading of the event.

2. The typology of nominals

Grimshaw (1990) classified nominals into three types, i.e. result nominals (R-nomials), simple event nominals (SE-nominals), and complex event nominals (CE-nominals). Only CE-nominals, also referred to in other approaches as Argument Supporting nominals (AS-nominals) (Alexiadou 2009; Borer 2003, 2013, 2014), are analysable in terms of aspectual distinctions and have an associated argument structure like verbs. In contradistinction to verbs, the external argument in CE-nominalizations is optional and if it is present (either as a NP in the genitive case or a *by*-PP), the internal argument is obligatory. They can license agent oriented modifiers (e.g. *deliberate*, *intentional*) as well as aspectual modifiers (e.g. *constant*, *frequent*). CE-nominals behave like verbs since they license event-related PPs (*in an hour*, *for an hour*) and cannot be made plural.

(1)   (the enemy's) destruction of the city in three days
        the deliberate destruction of the city (by the enemy)
        the constant shooting of rabbits by Bill

SE-nominals, like CE-nominals, have event implications and are barred from contexts appropriate for concrete objects, which are reserved for R-nominals. On a 'simple event' reading, like CE-nominals, they denote an event, but are not associated with an event structure and hence lack argument structure because they do not license event-related PPs and admit plural formation. Non-eventive R-nominals lacking an associated argument structure show a variety of readings, though they typically denote the product or result of the event denoted by the base verb and behave like non-derived nouns.

    As far as morphological marking is concerned, it is claimed that the ability to take arguments is always coupled with the presence of an overt nominalizer (Borer 2003: 47; Alexiadou and Grimshaw 2008: 3) and

that nominals with Latinate suffixes can be ambiguous between CE- and R-nominal status. In Alexiadou and Grimshaw (2008: 2) and Alexiadou (2009) we find a slightly modified version of this approach, where nouns like *examination* are three-way ambiguous, i.e. they have a complex event reading, a simple event reading, and a result reading, as shown in (2) below:

(2) a. The examination of the patients took a long   (Complex Event nominal)
         time.
     b. The examination took a long time.   (Simple Event nominal)
     c. The examination/exam was on the table.   (Result nominal)

A comprehensive elaboration of the debate instigated by Grimshaw's (1990) monograph concerning the characteristics of particular types of nominals would go far beyond the bounds of this paper,[2] but here we will specifically focus on the internal-argument licensing potential of SE-nominals.[3] We will contrast two competing theoretical proposals: the first advocated by Borer (2003, 2012, 2014), who conflates the categories of R-nominals and SE-nominals and calls them R(eferential)-nominals, and the second proposed by (Alexiadou 2017a), who claims that Grimshaw's tripartite division, though modified, seems to be relevant for synthetic compounds in Greek.

2.1. Borer's classification

The structures in (3) below translate the above-mentioned three-way distinction into a bipartition proposed in the exo-skeletal model of Borer (2005, 2013, 2014). In the XS-Model, roots devoid of formal (non-phonological) properties are merged with Categorial and Semantic Functors. $C_{N[V]}$ is a C-functor which projects N and which defines its complement as equivalent to V. In English it can be phonologically realized as, among others, *-ation*, *-ance*, *-ment*, *-al*. AS-nominals, in contradistinction to R-nominals, contain verbal functional structure. The root is dominated by functional nodes which are part of the verbal extended projection {Ex[V]}. In (3a) the root is immediately dominated by an ExP segment that licenses a quantity object (Asp_Q in Borer 2005). R-nominals are devoid of aspectual structure (3b), which explains why they do not co-occur with the internal argument.

---

2   Other important characteristics of the three types of nominals are discussed in Alexiadou (2001: 10–12), Alexiadou and Grimshaw (2008: 3), Borer (2003: 45; 2014: 71–73).

3   The term SE-nominal is used in this paper merely to highlight the fact that Irish deverbal nominals display properties which situate them in the fuzzy area between canonical CE-/AS-nominals and R-nominals. We will see that adopting a dichotomy approach results in different classifications depending on the inclusion criteria.

(3) a.   government of the people (Borer 2013: 420)

$$C_{N[V]}$$

$C_{N[V]}$ E

[c=v $^\pi\sqrt{}$ GOVERN$^{Q^\exists}$] $C_{N[V]}$

/$_\pi$ment/   << [c=v $^\pi\sqrt{}$ GOVERN$^{Q^\exists}$]   >>   ASP$_Q$

<< [c=v $^\pi\sqrt{}$ GOVERN^Q]   >>   [c=v $^\pi\sqrt{}$ GOVERN]
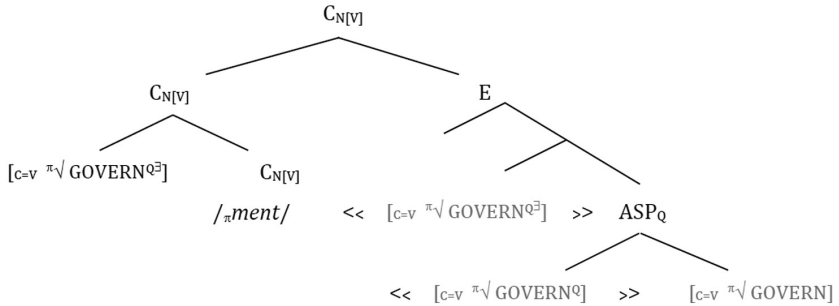
b. (The) government (is imperfect) (Borer 2013: 420)

$$C_{N[V]}$$

$C_{N[V]}$ [c=v $^\pi\sqrt{}$ GOVERN]

[c=v $^\pi\sqrt{}$ GOVERN] $C_{N[V]}$
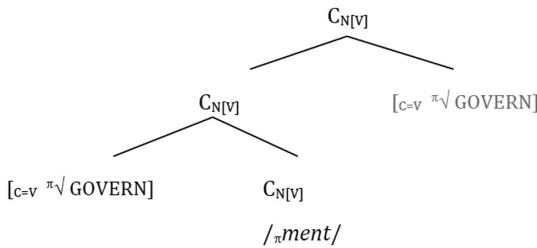
/$_\pi$ment/

The structure in (4) is a simplified representation of the structure of AS-nominals from Borer (2014: 83), in which both the internal and external argument are licensed by ExP-segments. As argued in Borer (2005), Asp_Q (Aspect of Quantity) introduces the internal argument, whereas Asp_{Ev} (Aspect of Event) licenses the event argument and an external argument.

(4) The government/governance (of the people by the people) (Borer 2014: 83)

$$C_{N[V]}$$

$C_{N[V]}$ X∈{Ex[V]}

[c=v $\sqrt{}$ GOVERN] $C_{N[V]}$ (ARG)

Y∈{Ex[V]}

ment

ance   (ARG)

[c=v $\sqrt{}$ GOVERN]
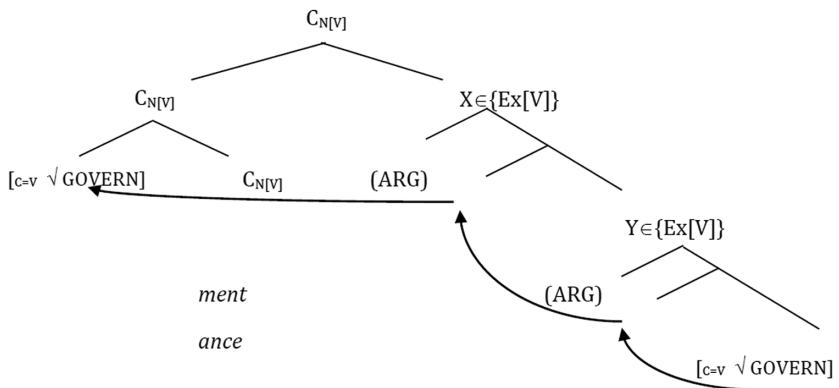
2.2. Alexiadou's classification

Following Kratzer (1996), Alexiadou et al. (2015) regard VoiceP as the locus of the external argument. VoiceP corresponds to the external-argument introducing vP shell in Distributed Morphology (Harley 2009: 325) and it is linked with the ability to take agent-oriented modifiers, *by*-phrases and instrumental phrases. The layer introducing the internal argument and event implications is associated with the categorizing head v. There is no 'RootP' and no argument can appear in the complement or specifier position of the root (Alexiadou and Schäfer 2011; Alexiadou 2014) and so "internal arguments are licensed via particles/prepositions/functional heads/small clauses" (Alexiadou et al. 2015: 13).[4] In Alexiadou's typology of nominals (see e.g. Alexiadou 2001, 2017a and Alexiadou et al. 2013) the structure in (5) corresponds to the structure of AS-nominals. It hosts an external argument in the specifier of VoiceP, the internal argument is licensed within vP and aspectual adverbials under AspP:

(5) [DP [nP [AspectP [VoiceP [vP [Root]]]]]] the training of the dog by John for 2 hours

---
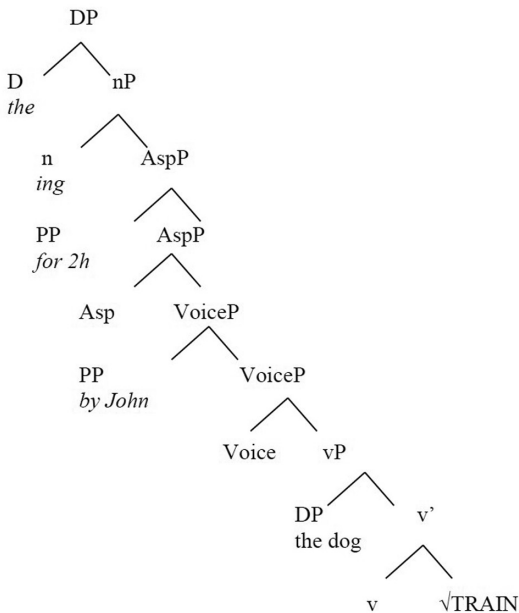
4   For a radically opposed view the reader is referred to Harley (2009, 2014), who extensively argues for the existence of the RootP.

R-nominals that lack event implications (as in (2c) above) will have the structure in (6), in which the root is directly merged with nominal functional layers. They are devoid of the vP functional projection which licenses the internal argument (Alexiadou 2009):

(6)   [DP [nP [Root]]]

In her analysis of deverbal compounds (DCs), Alexiadou (2017a) demonstrates that Greek synthetic DCs, just like AS-nominals, may have event implications, and allow an internal argument in their structure; unlike them, however, they lack an external argument, and do not tolerate agentive adverbials and aspectual modifiers. They are argued to contain v, the layer that introduces internal arguments. The examples in (7) below demonstrate that Greek DCs do not allow aspectual modifiers (7b) or *by*-phrases (7a), while they can refer to events (7c). Unlike English DCs, they can accept pluralization (7d), and disallow modifiers such as *frequent* (7e) (Alexiadou 2017a: 60):

(7)   a.   *kapnokaliergia   apo   agrotes
           tobacco-cultivation   by   farmers
      b.   *kapnokaliergia   ja   3 hronia
           tobacco-cultivation   for   3 years
      c.   I   kapno-kaliergia   sti Kavala   arhise   to   19o eona.
           the   tobacco-cultivation in Kavala   started   during   the 19th century
           'The tobacco cultivation in Kavala started during the 19th century.'
      d.   tosi   diafimisi   apo   kapnokaliergies
           so much   advertisement   from   tobacco-cultivations
           'So much advertising by tobacco-cultivations'
      e.   *i   sihni   kapnokalliergia   kurazi.
           the   frequent   tobacco cultivation   tires
           'Frequent tobacco cultivation is tiresome.'

Interestingly, the respective structures in English are compatible with *by*-phrases and agent-oriented modifiers (di Sciullo 1992; Iordăchioaia et al. 2017), which means that their structure additionally contains the VoiceP layer. Di Sciullo's (1992) example (30) shows that DCs in English can co-occur with the *by*-PP and agent oriented modifiers: [5]

---

5   For a detailed discussion of the differences in the internal structure of DCs in English and Greek and the ways in which the structure of a DC varies from the structure of an AS-nominal the reader is referred to Iordăchioaia et al. (2017) and Alexiadou (2017a).

(8) a. Taxi driving by John can be dangerous
    b. John's deliberate taxi driving did not please Harry.

Thus, Alexiadou's classification of deverbal nominals envisages the existence of nominals that lack implicit external arguments, but do have event implications, and internal arguments (9) as well as nominals which can additionally host the external argument but lack aspectual modifiers (10):

(9) [DP [nP [vP [Root]]]]

(10) [DP [nP [VoiceP [vP [Root]]]]]

In the analysis put forth in, e.g. Alexiadou (2001), (2009), Alexiadou et al. (2013), the absence of Aspect is not correlated with AS, whereas in Borer's (2013) system, where argument realization and aspect are tightly connected, the absence of aspectual structure implies the lack of AS. We will now examine the properties of nominals in Irish with a view to providing empirical evidence in favour of either of the two approaches.

3. The Irish data

3.1. Basic morphological and syntactic facts concerning VNs in Irish

The Irish data are interesting because there are no formal differences between nominalizations and non-finite forms (infinitives and present participles).[6] The occurrence of VNs, in different syntactic configurations can in no way be linked to distinct morphological markers.[7] The examples in Table 1 below illustrate VNs discharging the function of non-finite categories and nominalizations, where *socrú* is the suffixed VN of the verb *socr·aigh* 'settle'.[8]

---

6   The morphophonology of the Irish VN is analysed in detail in Bloch-Trojnar (2006). The formation of VNs in Irish involves over 20 morphophonological exponents, e.g.: *-(e)amh caith* 'spend' *caitheamh*, *-í cón·aigh* 'dwell' *cónaí*, *-ach ceann·aigh* 'buy' *ceannach*, *-t bain* 'cut' *baint*, *-chan beo·igh* 'animate' *beochan* etc. Regular default markers depend on the conjugation of the verb. First conjugation verbs take *-(e)adh* (e.g. *glan* 'clean' – *glanadh*), whereas second conjugation verbs are subject to the rule attaching *-ú* (e.g. *maslaigh* 'offend' – *maslú*).

7   Despite surface homonymy, it is possible to tease apart verbal from nominal categories with the aid of syntactic and morphological criteria. A detailed exposition of arguments for categorizing VNs in different contexts can be found in McCloskey (1983), Doyle (2002), Bloch-Trojnar (2006), Carnie (2011).

8   The citation form of the verb is the 2nd person singular imperative, because in many cases it coincides with the verbal root. In cases where it does not, it is necessary to leave out the inflectional marker *-(a)igh*.

Tab. 1: Different syntactic configurations featuring VNs in Irish.

Infinitive	Caithfidh mé fiacha a **shocrú**.[9] must I debt.PL PRT settle.VN 'I have to settle debts.'
Present Participle	Tá an aimsir ag **socrú**. is the weather PRT settle.VN 'The weather is becoming settled.'
AS-nominal (non-count)	**socrú** gaoithe, gleo settle.VN wind.GEN noise.GEN 'abatement of wind, noise'
Light Verb Construction (LVC)	**socrú** a dhéanamh le duine faoi rud settle.VN PRT do.VN with s.o. about sth 'to reach an agreement with s.o. about sth'
R-nominal (count)	**socruithe** sochraide settle.VN.PL funeral.GEN 'funeral arrangements

   In the vast majority of cases the formation of VNs involves roots devoid of overt verbalizing morphology.[9] However, it should be borne in mind that default affixation rules are conditioned by conjugation class membership and that the conjugations should be viewed as stem classes rather than inflectional classes (Ó Sé 1991). This means that stem types are diacritic features encoded on the v.[10] There are also cases where the verbalizer is overt

---

9 The letter *h* which immediately follows the initial consonant is the marker of its lenition, caused by the infinitival particle *a* (*socrú – a shocrú*).

10 Much like the prosodic templates of Semitic binyanim are encoded on the v (Arad 2005), there are two types of v, which receive a different realization depending on the structural and lexical properties of the root. The grammar also makes reference to diacritic features, arbitrary features that must simply be memorized as belonging to particular roots. Features relating to Conjugation count among them. Such features are relevant for morphological spell-out, but do not have any semantic interpretation. "This means that it is not the phonological form of the root which deterministically decides the conjugation of a verb; on the contrary, belonging to the second conjugation is a lexical property, which takes the form of a prosodic constraint on output forms" (Acquaviva 2014: 553). As rightly observed by the anonymous reviewer, the fact that the opposition is morphologically visible on verb stems only (hence 'conjugation') points to the presence of the [v] level. There is another piece of evidence in favour of the claim that conjugation is not a property of roots but the [Root-v] complex. In Irish, one and the same root can be merged with an aP, nP and vP categorizing head. A root like √DEARG depending on the categorizing head may be a predicate of a property (*dearg* 'red.ADJ'), may denote an abstract entity (*dearg* 'red.N, red colour') or an event. When the root is verbalized, it may be assigned to either

as in 1ˢᵗ conjugation verbs in *-ál*,[11] whose corresponding nominalization is marked overtly by means of a palatalizing autosegment, e.g. *buama* 'bomb' – *buamál-* 'to bomb' – *buamáil* 'bomb-VN' or *planda* 'plant' – *plandál-* 'plant' – *plandáil* 'to plant-VN'. Thus, there is morphological evidence supporting the presence of the v layer responsible for event implications.

3.2. The external syntax of nominals and their structure

Unlike Latinate nominalizations in English, which are ambiguous between the AS-, SE- and R-nominal status, nominals in Irish appear to show a different puzzling type of ambiguity. Namely, they appear to preserve AS but show no trace of Aktionsart/Aspect information.

Plural forms, as illustrated in (11) below, are a clear indication of R-nominal status. They are typically associated with resultative or concrete semantics[12] and they are never accompanied by NPs corresponding to participants involved in the event denoted by the base verb. Such nouns would be represented as [N[V[√]] or [Class[√]]] complexes in Borer's approach or as [DP [nP [Root]]] complexes in Alexiadou's approach.

(11)	Verb – VN	Nominalization (plural)
a.	tóg 'raise, build'– tógáil	Is breá na tógálacha iad. 'They are fine structures.'
b.	gear 'cut, levy' – gearradh	na gearrthacha a íoc 'to pay the rates'
c.	imigh 'go, go on'– imeacht	imeachtaí an lae 'events of the day'
d.	oir 'suit, fit, need' – oiriúint	oiriúintí oifige 'office accessories'
e.	éag 'die' – éag	éaga na bliana 'the year's mortality'
f.	leigheas 'cure, heal' – leigheas	leigheasanna 'cures'
g.	sáigh 'thrust'– sá	sáite ciseáin 'stakes of basket'

With reference to the syntactic properties of nominalizations, Doyle (2002: 100–101) points out that nominalizations are hardly ever accompanied by a complement and constructions with the corresponding finite verbs are usually preferred (see also Bloch-Trojnar 2016). He gives two examples which, in his opinion, border on ungrammaticality:

---

the 1ˢᵗ or the 2ⁿᵈ conjugation and will accordingly combine with a different nominalizer, i.e. *dearg* 'redden, I' – *deargadh*, *dearg·aigh* 'redden, II' – *deargú*.

11 The suffix *-ál*, is used to nativize English verbal roots, provided they do not exceed three syllables in length, as in *draibheál-* 'to drive', *traenál-* 'to train', *péinteál-* 'to paint' or *robál-* 'to rob' (Doyle 1992: 99). It is also a means of turning nouns, both native and of English provenance, into verbs, as in *buama* 'bomb' – *buamál-* 'to bomb' or *planda* 'plant' – *plandál-* 'to plant' (Wigger 1972: 207–210).

12 Lexicalized senses of deverbal nominals are extensively discussed in Bloch-Trojnar (2018).

(12)   a. */? Chuir   briseadh   na   gcloch   ionadh   orm.
           put.PST   break.VN   the   stones.GEN   surprise   on-me
           'The breaking of the stones surprised me.'
       b. */? Chuala   siad   briseadh   na gcloch.
           hear.PST   they   break.VN   the stones.GEN
           'They heard the breaking of the stones.'

He concludes that there may be some constraint in Irish "to the effect that nominalizations do not inherit the subcategorization frames of their verbal bases" (Doyle 2002: 101). However, the apparent ungrammaticality of his examples may stem from something else. In cases of logical polysemy, the superordinate predicate determines the availability of a given interpretation and the superordinate predicates in (12) above enforce non-actional readings. The predicate *hear* requires the result interpretation – you cannot hear the action, only its acoustic result (Puzynina 1969: 166). A predicate such as *surprise* triggers a factitive meaning (Melloni 2011), which is not available for Irish nominalizations. In Bloch-Trojnar (2018: 34), I demonstrate that searches for translations of 'the fact that' constructions invariably return sentences with dependent clauses, e.g.

(13)   'I was surprised at his late arrival.'
       a. Bhí   iontas   orm   go   raibh sé déanach ag   teacht.
           was   surprise   on-me   that   was   he late   PRT come.VN
       b. *Chuir   a   theacht   déanach   iontas   orm.
           put.PST   his   come.VN   late   surprise on-me

Nevertheless, structures where action nouns are accompanied by event participants are not impossible. We can find examples of VNs with complements in dictionaries (Ó Dónaill 1977; de Bhaldraithe 1959) and the New Corpus for Ireland (Nua-Chorpas na hÉireann).[13] Eventive interpretation is a prerequisite for the presence of AS. Hence, in what follows we will limit our investigation to nouns which refer to events and apply the standard diagnostic tests to establish the structural layers in their representation. We randomly select 3 verbs belonging to the traditionally recognized situation types (Vendler 1967; Smith 1997), i.e. states (*creid* 'believe' – *creidiúint*, *tuig* 'understand' – *tuiscint*, *amharc* 'see' – *amharc*), activities (*maslaigh* 'offend' – *maslú*, *iompair* 'carry' – *iompar*, *gearáin* 'complain' – *gearán*), accomplishments (*scríobh* 'write' – *scríobh*, *téigh* 'warm' – *téamh*, *crúigh* 'milk' – *crú*), achievements (*maraigh* 'kill' – *marú*, *athraigh* 'change' – *athrú*,

---

13 The New Corpus for Ireland (henceforth NCI) is a corpus of approximately 30 million words, which allows access to words in context and makes it possible to identify sentences produced by native speakers. Unless otherwise specified the example comes from the NCI.

díol 'sell' – *díol*) and semelfactives (*buail* 'hit' – *bualadh*, *léim* 'jump' – *léim*, *croith* 'shake' – *croitheadh*). We filter the results to obtain sentences produced by native speakers and examine their complementation and modification patterns. We also conduct the search of the entire corpus to find out if particular modifiers are attested with VNs in their nominal function. Corpus searches are supplemented with material from standard dictionaries, i.e. Ó Dónaill (1977) and de Bhaldraithe (1959).

3.2.1. *The presence of v*

VNs can make reference to a simple event, i.e. they can act as subjects of predicates such as 'last', 'occur', 'start', 'stop', 'go on', 'take place' etc.

(14) a. Mhothaigh sé an t-athrú   spioradálta a   bhí   ar siúl.
        feel.PST   he the change.VN spiritual   that be.PST on go-on.VN
        'He felt the spiritual change that was happening.' (NCI)

   b. … agus tharla   an marú céanna air   siúd.
        and happen.PST the kill.VN similar on-him yon
        '… and a similar death happened to him.' (NCI)

3.2.2. *The presence of vP*

The presence of the vP layer is associated with the licensing of the internal argument. If present, the NP complement corresponding to the direct object of a transitive verb or to the subject of an intransitive (unaccusative) verb is in the genitive case, as shown in (15a) and (15b) respectively. Also VNs can license PP complements selected by verbs underlying the nominal structure (15c).

(15) a. Níl   **iompar**   **scine**   ceadaithe.
        be.NEG   carry.VN   knife.GEN   permit.PPRT
        'It is not permitted to carry a knife.' (Ó Dónaill 1977: 722)

   b. Bhí   **titim**   **na**   **hoíche**   ann.
        be.PST   fall.VN   the   night.GEN   there
        'There was falling of the night. / The night was falling.' (Ó Cadhlaigh 1940: 74)

   c. faoi   **dhíol**   **gunnaí**   leis   na   hIndiaigh
        about   sell.VN   gun.GEN.PL   with   the   Indian.COM.PL
        'about selling guns to the Indians' (NCI)

VNs in Irish can also act as complements of light verbs such as, e.g. *tabhair* 'give', *lig* 'let', *faigh* 'get', *bain* 'take, extract' and *cuir* 'put' (Ó Siadhail 1989: 304–308; Wigger 2008).[14] The licensing of the participants in the event denoted by the VN in light verb constructions (henceforth LVCs) is far from settled.[15] The nominals corresponding to the direct object of the base verb mainly feature as complements of prepositions. Ó Siadhail (1989: 306) points to the spread of prepositional phrases with *ar* 'on' (16a).[16] If the light

verb is a prepositional verb (i.e. *tabhair do* 'give to', *bain as* 'extract from'), a specialized preposition will precede the object of the VN (16b–c). The genitive case marking of the object is rare but not impossible (16d).[17]

(16)

a. Do   dheineas riarú   **ar** an leasú. Cf. Riaraigh   mé an leasú.
   PRT   do.1SG.PST apportion.VN on the manure   apportion.PST I   the manure
   'I apportioned the manure.' (Ó Siadhail 1989: 306)   'I apportioned the manure.'

b. Bhain   sé   croitheadh **as**   an mbuidéal.
   extract.PST he   shake.VN   from   the bottle
   'He shook the bottle.' (de Bhaldraithe 1959: 651)

c. Tabhair   téamh   beag   eile   **do**n bhainne.
   give.2SG.IMPER warm.VN   small   another   to-the milk
   'Warm the milk a little more.' (Ó Dónaill 1977: 1217)

d. Ná   tabhair   **léim**   **na díge**   sin.
   NEG   give. 2SG.IMPER jump.VN the ditch.GEN that
   'Don't jump that ditch.' (Ó Dónaill 1977: 1187)

Another indicator of vP is manner modification. As we can see in (17), the VN accompanied by a complement expressed as a prepositional pronoun, i.e. a preposition+pronoun combination *air* 'on+him', is modified by an adjective:

(17) Agus ní marú   tapaidh imeoidh   air (…), ach   marú mall millteach!
     and not kill.VN quick   come.FUT on-him but   kill.VN slow dreadful
     'And it won't be a quick death that will happen to him either, but a slow and dreadful one' (NCI)

---

14 A detailed analysis of the semantics, aspectual properties and the nominal status of the VN in LVCs is available in Bloch-Trojnar (2009, 2010, 2014).

15 Bearing in mind the cross-linguistic prevalence and diversity of LVCs, linguists are far from unanimous as to the exact nature of this interaction. There are proposals, in which the verb is regarded merely as locus for agreement and tense morphology and has no influence on the number and type of arguments side by side analyses involving argument structure composition in which light verbs have partially specified argument structures which are shared, fused, superimposed on or merged with the argument structure of the complement. For a detailed discussion of these proposals see Butt (2003) and the references therein.

16 In the framework of Grimshaw (1990), the occurrence of prepositions introducing the direct object NPs/DPs in the (de)verbal nominal could be taken as a clue that these are not arguments but complements. However, in this paper I follow Alexiadou and Schäfer (2011), Alexiadou (2014), Alexiadou et al. (2015) in regarding vP or PPs as functional heads introducing the internal argument.

17 Cases where the modifying NP in the genitive case corresponds to the object of a transitive verb are extremely rare. Notwithstanding, they are not impossible, which would mean that genitive marking and acting as a PP complement are non-distinctive strategies of expressing the internal argument.

3.2.3. *The presence of VoiceP*

The standard tests for the presence of the VoiceP layer licensing the external argument are *by*-phrases, agent oriented modifiers and instrumental phrases. Notably, cases with both arguments realized are virtually unattested (especially in Traditional Modern Irish).[18] Examples where the external argument is realized as an NP dominated by the preposition *ag* are usually dismissed as being calques from English, which do not reflect the true structure of the language, i.e. they can be found in Non-Traditional Late Modern Irish, which is a variety influenced by English and L2 speakers of Irish. Examples in (18) are a case in point. (18a) comes from an Online Irish-English Dictionary,[19] whereas there is no information of the source of (18b) in NCI:

(18)

a. maidir le beostoc a chosaint **óna** **gciapadh ag madraí**
  as regards livestock PRT protect.VN from-their harass.VN at dog.COM.PL
  'for the protection of livestock from worrying by dogs (lit. from their worrying by dogs)'

b. Léiríonn seo **athrú** mór **stíle** **ag na bhFrancaigh.**
  explain.PRES this change.VN big style.GEN at the French
  'This explains the great change of style of the French.' (NCI: source unknown)

However, we do find examples produced by native speakers in which the subject of the verb, in a nominalized structure is realized by the *ag*-phrase, providing that the internal argument is introduced by *ar*:

(19) Caoinim an fáil **ar** bhás atá **ag** gach a maireann.
  cry.1SG.PRES the get.VN on death that-is at every that live.3SG.PRES
  'I lament the finding of death by everything that lives.' (Ó Cearúil 1999: 110)

The external argument can sometimes feature in the genitive case as in (20) below. Note that the internal argument is licensed by the preposition *de* 'of':

(20) Tá fágaint na mairnéalach **den** leac seo ar na
  is leave.VN the sailors.GEN.PL from-the rock this among the
  hiontaisí is mó i stair an Bhlascaoid.
  wonders COP greatest in-the history the Blasket.GEN
  'The departure of the sailors from this rock is one of the strangest things in the history of the Blasket.' (Ó Criomhthain 1997: 140)

In LVCs the subject of the light verb is semantically co-identified as the external argument of the deverbal nominal complement. It can in no way be treated as its syntactic argument.

---

18 For the distinction between Traditional Modern Irish and Non-Traditional Late Modern Irish see Ó Béarra (2007).

19 http://www.englishirishdictionary.com/dictionary (accessed January 2016).

Agentive modifers such as *toiliúil* 'willful, deliberate', *aireach*, *faichilleach*, *cúramach* 'careful' have been found with VNs, but not in the presence of complements, e.g. *dúnmharú **toiliúil*** 'willful murder', *pleanáil **chúramach*** 'careful planning'. No instrumental phrases with VNs in their nominal usage have been found in this pilot study.

To sum up: Irish nominals acting as subjects, objects or complements of prepositions[20] can be accompanied by the internal argument (in the genitive case or preceded by a preposition). The examples with the external argument licensed by the *ag*-phrase are extremely hard to come by and their acceptability is linked with the obligatory presence of the object licensed by a preposition. Thus, nominals in Irish can be regarded as Argument Supporting. Is this tantamount to saying that they are CE-nominals in the sense of Grimshaw? To claim so we have to provide positive evidence for the presence of the AspP layer in their structure.

3.2.4. *The presence of AspP?*

The presence of AspP can be probed with the aid of frame adverbials such as 'in an hour/for an hour' and the possibility of modification with 'frequent' and 'constant'. Neither *minic* 'frequent' nor *leanúnach* 'constant' can be found with VNs if they are accompanied by NP satellites. Consider some examples in (21) below. In (21a) *minic* 'frequent' is the complement of the copula. (21b) exemplifies *leanúnach* modifying a VN which is an SE-nominal acting as the object of the verb *tarlaigh* 'occur', all similar cases being confined to non-native contexts in the corpus.

(21)

a. Is   minic   marú   éisc   in   aibhneacha na   hÉireann na   laethanta seo
   cop often   kill.VN fish.GEN in   river.PL   the   Ireland   the   day.PL   this
   'The killing of fish in Ireland's rivers is a frequent occurrence these days.' (NCI)

b. tá   deireadh leis   an   marú   leanúnach a   tharla   lá   i ndiaidh lae
   is   end   with   the   kill.VN constant   that occur.PST day after   day
   'There is an end to the constant killing that has been going on day after day'
   (NCI: non-native)

The corpus data does not provide positive evidence for their compatibility with durative and frame adverbials, i.e. they do not occur with aspectual adverbials of the type 'in X time', 'for X time'. They are not attested with phrases such as *ar feadh an lae /bliana/míosa/an ama* 'for a day/month/year/time etc.' As regards durative expressions such as *le linn* 'during, in the course of' and *i rith* 'in the course of, throughout', it can be observed that

---

20 For more linguistic examples see, e.g. Bloch-Trojnar (2006, 2013, 2016).

they are followed by non-finite clauses, i.e. VNs preceded by the auxiliary *bí* (22a), or VNs with demonstratives (22b), a feature of SE nominals. What look like VNs with complements in (22c) do not provide conclusive evidence for the presence of AS:[21]

(22) a. | Le linn | **dó** | **bheith** | **ag** | **freastal** | ar | an | ollscoil |
during	his	be.VN	at	attend.VN	on	the	university
bhí	sé	faoi	gheasa	ag	ceol		
be.PST	he	under	spell	by	music		

'when he was attending university he was enamoured by music' (NCI)

   b. | Le linn | **na** | **cainte** | **sin** | bhí | a | shúile | go dlúth |
during	the	talk.VN	that	be.PST	his	eye.PL	intensely
ar	an	bhfuinneog	ag	Mr Morris			
on	the	window	at	Mr Morris			

'during this talk Mr Morris had his eyes directed intensely at the window' (NCI)

   c. | Le linn | **éirí** | **na** | **gréine** | tháinig | néal | dubh | codlata |
during	rise.VN	the	sun-gen	come.PST	cloud	black	sleep.GEN
i gceann	na	laoch					
in head	the	worrior.GEN.PL					

'During sun-rise a black cloud of sleep came over the heads of the warriors' (NCI)

The adjectives *céimseach* and *dréimreach* 'gradual' give a handful of hits each but none of them occurring with VNs. Ordinarily, complex adverbial expressions are used such as *diaidh ar ndiaidh*, *céim ar chéim*, *de réir a chéile* 'gradually' which co-occur with clauses and not NPs. We will interpret these facts as the absence of the AspP layer in the nominal structure. Neither Aktionsart nor Aspect seem to be encoded.

As far as LVCs are concerned, Ó Siadhail (1989: 307) and Wigger (2008) underline that these structures are used to achieve a partitive or singulative

---

21 The scarce examples such as those in (22c) above could be interpreted as syntactic phrases but more plausibly as left-headed NN$_{\text{GEN}}$ compound structures on account of the generic reference of the second N (Doyle 1996). We have seen that compounds in Greek and in English can be viewed as having argument structure. However, the presence of internal arguments is not always indicative of AS. Borer (2014: 78 n10) points out that simple event nominals may have an internal argument, without a verbal base introducing it (e.g. *my impulse to be daring*, *my lab's assistant culture of new forms of bacteria*). In Irish, R-nominals can also take a complement as indicated by example (11c), (11d), (11e) and (11g) above. Their R-nominal status is confirmed by plural number marking. The VN *éirí* in (22c) is non-count and since AS-nominals must be singular, whereas R-nominals may pluralize, it could be either. This issue definitely merits further investigation on a more extensive body of data.

effect. The telicizing effect, which is cross-linguistically attributed to LVCs (Butt 2003), is achieved in Irish by the ability of complement VNs to appear with cardinal numbers, enumerative determiners (e.g. *amháin* 'one', *eile* 'another', *chéad* 'first', *iomaí* 'many', *cúpla* 'a few'), and adverbials (*cúpla uair* 'a few times', *arís* 'again'), which bring about a bounded interpretation of the entire predicate (Bloch-Trojnar 2014). Interestingly, even though VNs can co-occur with counting adverbials in these structures, they are morphologically non-count and do not feature in the plural. The plural never appears on the complement in LVCs, even though some VNs have the plural form available. Substituting the plural for the singular always results in ungrammaticality. Consider the example in (23) below.

(23) *gearán*, gen.sg. & nom.pl. -*áin*, pl. ~ 'complain-VN'

Bhí	sé	ina	shuí	ag	déanamh gearáin /		*gearán	leis	fein.
be.PST	he	in-his	sit.VN	PRT	do.VN	complain.VN. GEN.SG	complain.VN. GEN.PL	with-him	self

'He sat there feeling sorry for himself.' (Ó Dónaill 1977: 622)

An analysis of LVCs using the tools of DM and Minimalist Syntax is offered in Alexiadou (2017b), who builds on the structure proposed for compositional resultatives in Hale and Keyser (2002) and depicted in (24) below, in which v takes as its complement a complex phrase xP or root.[22]

(24)

```
        vP
       /  \
      v    ResultP<s>/=√
```
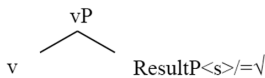
In Greek, English, and Kurdish examples discussed in Alexiadou (2017b) the complement of v is nP embedding a verbal structure:

(25)

```
        vP
       /  \
      v    nP
          /  \
         n    vP
```
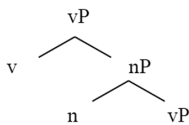
This embedded verbal structure, in turn, introduces the internal argument, i.e. [nP[vP theme [Root]]]. Light verbs in this analysis are treated as the overt realization of the v head.

---

22 Eventive v<e> in the context of a root with a phonological matrix (WHITE) will give rise to *whiten*. If it does not find a root with a phonological matrix it will be spelt out as a copula or light verb (*become/go/ turn white*) because an xP cannot move into v.

3.3. The internal structure of nominals in Irish

The results of the application of the standard tests for the presence of particular layers of verbal structure prompt an analysis on which VNs in Irish in their nominal function are either R-Nominals or SE-nominals. In the former case the root is merged with nominal functional layers and behaves like an ordinary noun and may therefore pluralize. In the latter case we are dealing with nominals that are event denoting and argument licensing but lack the AspP layer in their structure. They are neutral with respect to the inflectional category of number and incapable of pluralizing. They are [nP[vP [Root]]] or at most [nP[VoiceP[vP [Root]]]] outside LVCs. Crucially in both cases the structures are devoid of the AspP layer.

The use of the term SE-nominal may seem controversial since it does not converge with the widely accepted definition proposed by Grimshaw (1990) – that of an (underived) event denoting noun which lacks both event structure and argument structure. However, the properties of Irish deverbal nominals make them pretty hard to tag. The task is made more difficult by the fact that most linguists work with a two-way distinction into AS-/CE- nominals and R-nominlas. My main contention is that there are no prototypical AS-/CE-nominals in Irish, in which the licensing of argument structure is coupled with the presence of the internal aspectual structure. If we apply Borer's criteria, Irish nominals do not quite match the class of R-nominals or 'non-AS de-verbal nominals' (Borer 2014: n1) on account of the fact that they structurally license arguments and they always embed a real, attested verb (have a licit v spellout). Artemis Alexiadou and her collaborators, in turn, (e.g. Alexiadou et al. 2010, Alexiadou and Schäfer 2010, Alexiadou 2017a) place emphasis on the contrast between deverbal nominals with argument structure (Argument Supporting Nominals, i.e. ASNs) and those without argument structure (Referential Nominals, i.e. RNs). In this approach there is a split between grammatical aspect (AspectP) and argument realization. Event structure is implied by argument realization and represents lexical aspect. Therefore, this approach is preferable with respect to the Irish data.

4. Conclusion

Our analysis bears out the classification of nominals proposed by Alexiadou (2017a) and the hypothesis advanced in Alexiadou and Schäfer (2010), according to which it is necessary to sever the licensing of AS from the presence of an aspectual reading of the event. Event denoting nominals in Irish can license the internal argument but aspectual modification is not possible (pretty much like in synthetic compounds discussed in Alexiadou 2017a). Such nominals are represented as the [nP [vP [Root]]] complex (also found in light verb constructions). The expression of the external arguments

outside LVCs is not impossible, but highly restricted. R-nominals are devoid of the vP layer – [nP [Root]] – and behave like ordinary nouns. Our data seem to indicate, counter to Borer (2003, 2005, 2013) and in line with Alexiadou (2017a), that internal argument licensing does not automatically imply the presence of the Asp layer in the structure. According to Acquaviva (2014: 548) Irish verbal inflection lacks an aspect morpheme, which "makes it easier to understand why this category finds a periphrastic expression." If the AspP is located above the functional projection licensing the external argument, we can hypothesize that what is a participle in the progressive or other aspect related constructions or the infinitive in modal constructions is in fact a structure involving the root and argument structure, leaving the aspectual information to be expressed by the light/copula/auxiliary verb in the matrix clause. This would explain the widespread homonymy presented in Table 1.

References

Acquaviva, Paolo. 2014. The categories of Modern Irish verbal inflection. *Journal of Linguistics* 50. 537–586.

Alexiadou, Artemis & Florian Schäfer. 2010. On the syntax of episodic vs. dispositional -er nominals. In Artemis Alexiadou & Monika Rathert (eds.), *The syntax of nominalizations across languages and frameworks*, 9–38. Berlin: de Gruyter.

Alexiadou, Artemis & Florian Schäfer. 2011. There-insertion: An unaccusative mismatch at the syntax-semantics interface. *Online Proceedings of WCCFL* 28. https://sites.google.com/site/wccfl28pro/alexiadou-schaefer

Alexiadou, Artemis & Jane Grimshaw. 2008. Verbs, nouns and affixation. In *SIN-SPEC (Working Papers of the SFB 732)* 1. 1–16.

Alexiadou, Artemis, Elena Anagnostopoulou & Florian Schäfer. 2009. PP licensing in nominalizations. *Proceedings of NELS* 38. 39–52.

Alexiadou, Artemis, Elena Anagnostopoulou & Florian Schäfer. 2015. *External arguments in transitivity alternations: A layering approach*. Oxford: Oxford University Press.

Alexiadou, Artemis, Gianina Iordăchioaia & Elena Soare. 2010. Number/Aspect interactions in the syntax of nominalizations: A Distributed Morphology approach. *Journal of Linguistics* 46. 537–574.

Alexiadou, Artemis, Gianina Iordăchioaia & Florian Schäfer. 2011. Scaling the variation in Romance and Germanic nominalizations. In Petra Sleeman & Harry Peridon (eds.), *The noun phrase in Romance and Germanic*, 25–40. Amsterdam: Benjamins.

Alexiadou, Artemis, Gianina Iordăchioaia, Maria Angeles Cano, Fabienne Martin & Florian Schäfer. 2013. The realization of external arguments in nominalizations. *Journal of Comparative Germanic Linguistics* 16. 73–95.

Alexiadou, Artemis. 2001. *Functional structure in nominals: Nominalization and ergativity*. (Linguistik Aktuell 42). Amsterdam: Benjamins.

Alexiadou, Artemis. 2009. On the role of syntactic locality in morphological processes: The case of (Greek) nominals. In Anastasia Giannakidou & Monika Rathert (eds.), *Quantification, definiteness and nominalization*, 253–280. Oxford: Oxford University Press.

Alexiadou, Artemis. 2014. Roots don't take complements. *Theoretical Linguistics* 40(3/4). 287–297.

Alexiadou, Artemis. 2017a. On the complex relationship between deverbal compounds and argument supporting nominals. In Maria Bloch-Trojnar & Anna Malicka-Kleparska (eds.), *Aspect and valency in nominals*, 53–82. Berlin: de Gruyter.

Alexiadou, Artemis. 2017b. Building verbs in language mixing varieties. *Zeitschrift für Sprachwissenschaft* 36(1). 165–192.

Arad, Maya. 2005. *Roots and patterns: Hebrew morpho-syntax*. Dordrecht: Springer.

Bloch-Trojnar, Maria. 2006. *Polyfunctionality in morphology. A study of verbal nouns in Modern Irish*. Lublin: Wydawnictwo KUL.

Bloch-Trojnar, Maria. 2009. Aspectual characteristics of light verbs constructions in Modern Irish. In Maria Bloch-Trojnar (ed.), *Perspectives on Celtic languages*, 63–85. Lublin: Wydawnictwo KUL.

Bloch-Trojnar, Maria. 2010. Semantic constraints on light verb constructions in Modern Irish. In Piotr Stalmaszczyk & Maxim Fomin (eds.), *Dimensions and categories of celticity. Studies in language*. (Studia Celto-Slavica 4), 67–88. Łódź: Łódź University Press.

Bloch-Trojnar, Maria. 2013. *The mechanics of transposition. A study of action nominalisations in English, Irish and Polish*. Lublin: Wydawnictwo KUL.

Bloch-Trojnar, Maria. 2014. Morphosyntactic and aspectual characteristics of predicative verbal nouns in SVCS in Irish. In Fryni Kakoyianni-Doa (ed.), *Penser le lexique-grammaire: Perspectives actuelles*, 61–74. Paris: Honoré Champion.

Bloch-Trojnar, Maria. 2016. Linguistic contact and the complementation of regular action nominals in Irish. *Academic Journal of Modern Philology* 5. 17–30.

Bloch-Trojnar, Maria. 2018. Semantic (ir)regularities in action nouns in Irish', *Studia Celto-Slavica: Journal of the Learned Association Societas Celto-Slavica* 9. 31–61.

Borer, Hagit. 2003. Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. In John Moore & Maria Polinsky (eds.), *The nature of explanation in linguistic theory*, 31–67. Stanford, CA: CSLI Publications.

Borer, Hagit. 2005. *The normal course of events: Structuring sense*. Oxford: Oxford University Press.

Borer, Hagit. 2012. In the event of a nominal. In Martin Everaert, Marijana Marelj & Tal Siloni (eds.), *The theta system: Argument structure at the interface*, 103–150. Oxford: Oxford University Press.

Borer, Hagit. 2013. *Taking form*. Oxford: Oxford University Press.

Borer, Hagit. 2014. Derived nominals and the domain of content. *Lingua* 141.71–96.

Butt, Miriam. 2003. *The light verb jungle*. www.ai.mit.edu/people/jimmylin/papers/Butt03.pdf

Carnie, Andrew. 2011. Mixed categories in Irish. *Lingua* 121. 1207–1227.

de Bhaldraithe, Thomas. 1959. *English-Irish Dictionary*. Dublin: An Gúm.

Di Sciullo, Anna Maria. 1992. Deverbal compounds and the external argument. In Iggy Roca (ed.), *Thematic structure: Its role in grammar*, 65–78. Dordrecht: Foris.

Doyle, Aidan. 1992. *Noun derivation in modern Irish. Selected categories, rules and suffixes*. Lublin: Wydawnictwo KUL.

Doyle, Aidan. 1996. Compounds and syntactic phrases in Modern Irish. *Studia Anglica Posnaniensia* 30. 83–96.

Doyle, Aidan. 2002. *Covert and overt pronominals in Irish*. Lublin: Folium.

Embick, David & Rolf Noyer. 2009. Distributed Morphology and the syntax-morphology interface. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford handbook of linguistic interfaces*, 289–324, Oxford: Oxford University Press.

Embick, David. 2010. *Localism vs. globalism in morphology and phonology*. Cambridge, MA: MIT Press.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press.

Hale, Ken & Samuel Jay Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.

Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20: Essays in honour of Sylvain Bromberger*, 111–176. Cambridge, MA: MIT Press.

Halle, Morris & Alec Marantz. 1994. Some key features of Distributed Morphology. In Andrew Carnie & Heidi Harley (eds.), *Papers on phonology and morphology (MIT Working Papers in Linguistics 21)*, 275–288. Cambridge, MA: The MIT Press.

Harley, Heidi & Rolf Noyer. 1999. State-of-the-article: Distributed Morphology. *GLOT International* 4(4). 3–9.

Harley, Heidi & Rolf Noyer. 2000. Licensing in the non-lexicalist lexicon. In Bert Peeters (ed.), *The lexicon/encyclopaedia interface*, 349–374. Amsterdam: Elsevier Press.

Harley, Heidi. 2009. The morphology of nominalizations and the syntax of vP. In Anastasia Giannakidou & Monika Rathert (eds.), *Quantification, definiteness and nominalization*, 321–343. Oxford: Oxford University Press.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40. 225–276.

Iordăchioaia, Gianina, Artemis Alexiadou & Andreas Pairamidis. 2017. Morpho-syntactic sources for synthetic compounds in English and in Greek. *Journal of Word Formation* 1. 45–70.

Krazter, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck & Laurie Zaring (eds.), *Phrase structure and the lexicon*, 109–137. Dordrecht & Boston: Kluwer.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4. 201–225.

Marantz, Alec. 2013. Locality domains for contextual allomorphy across the interfaces. In Ora Matushansky & Alec Marantz (eds.), *Distributed Morphology today: Morphemes for Morris Halle*. Cambridge: MIT Press.

McCloskey, James. 1983. A VP in a VSO language? In Gerald Gazdar, Ewan Klein & Geoffrey K. Pullum (eds.), *Order, concord and constituency*, 9–55. Dordrecht: Foris.

Melloni, Chiara. 2011. *Event and result nominals. A morphosemantic approach*. Bern: Lang.

Ó Béarra, Feargal. 2007. Late Modern Irish and the dynamics of language change and language death. In Hildegard L. C. Tristram (ed.), *The Celtic languages in contact*, 260–269. Potsdam: Potsdam University Press.

Ó Cadhlaigh, Cormac. 1940. *Gnás na Gaedhilge [The usage of Irish]*. Baile Átha Cliath: Oifig an tSoláthair.

Ó Cearúil, Micheál (ed.) (1999). *An Aimsir Óg*. Baile Átha Cliath: Coiscéim.

Ó Criomhthain, Tomás. 1997. *Bloghanna ón mBlascaod*. Baile Átha Cliath: Coiscéim.

Ó Dónaill, Niall. 1977. *Foclóir Gaeilge-Béarla [Irish-English dictionary]*. Baile Átha Cliath: An Gúm.

Ó Sé, Diarmuid. 1991. Verbal inflection in Modern Irish. *Ériu* 42. 61–81.

Ó Siadhail, Mícheál. 1989. *Modern Irish. Grammatical structure and dialectal variation*. Cambridge: Cambridge University Press.

Puzynina, Jadwiga. 1969. *Nazwy czynności we współczesnym języku polskim [Action nouns in Modern Polish]*. Warszawa: PWN.

Smith, Carlota. 1997. *The parameter of aspect*. Dordrecht: Kluwer.

Vendler, Zeno.1967. *Linguistics in philosophy*. New York: Cornell University Press.

Wigger, Arndt. 1972. *Preliminaries to a generative morphology of the Modern Irish verb*. Ériu 23. 162–213.

Wigger, Arndt. 2008. Advances in the lexicography of Modern Irish verbs. In Anna Bloch-Rozmej (ed.), *Issues in Celtic linguistics*, 233–250. Lublin: Wydawnictwo KUL.

Maria Bloch-Trojnar
The John Paul II Catholic University of Lublin
Department of Celtic Studies
Al. Racławickie 14, 20-950
Lublin
bloch@kul.pl

Marie Laurence Knittel & Florence Villoing

Instrument and Means interpretation of deverbal nominals: The role of ambiguous stative verbs in French VN compounding[1]

Abstract: This article examines French Verb-Noun compounds with Means value (*couvre-pied* 'blanket', lit. cover-feet), derived from stative bases. It shows that they are generally ambiguous between Means and Instrument reading. The regularity of this double value discards an analysis relying on verbal homonymy, in favor of Rothmayr's (2009) hypothesis of bi-eventive verbs. We assume that the presence of an agentive as well as a stative component in the verbal bases accounts for the double Means/Instrument value of the VNs studied here. We also examine "pure" Instrument VNs, available with similar verbal bases. We show that the distribution of the Instrument vs. Means/Instrument values relies on the state of the referent of the noun involved in the compound after the event described by the verbal base occurred. A permanent state entails a "pure" Instrument reading, whereas Means/Instrument reading obtains if the state of N is reversible (Fábregas & Marín 2012).

Keywords: French, VN compounds, participant nominals, stativity, reversible states, Means reading, Instrument reading

1. Introduction

This article deals with French Verb-Noun compounds (henceforth VN) with a Means interpretation, as illustrated in (1-3).

(1)   a. *couvre-pieds* 'feet blanket' (lit. cover-feet)
       b. *Le couvre-pieds rose lui couvre les pieds.*
          'The pink blanket covers his feet.'

(2)   a. *protège-cahier* 'notebook cover' (lit. protect-notebook)
       b. *Le protège-cahier vert protège son cahier de poésie.*
          'The green notebook cover protects his poetry notebook.'

(3)   a. *pince-nez* 'nose clip' (lit. pinch-nose)
       b. *Son pince-nez de natation lui pince trop fort le nez.*
          'His swimming nose clip pinches his nose too hard.'

Based on the study of a list of 1473 French VN compounds from both ordinary and specialized language found in large French dictionaries[2] as well as

---

1 We are grateful to the participants of JENOM 8 and to the reviewers of ZWJW for their valuable comments.

2 Specifically, *Trésor de la Langue Française* (https://www.cnrtl.fr/), *Grand Robert de la langue française* (Dictionnaire alphabétique et analogique de la langue française,

randomly collected neologisms, we show that VN compounds have a Means interpretation, that is, they describe inert causes of stative eventualities. This interpretation relies on the stativity of their base verbal lexeme, as shown in examples (1b, 2b, 3b), and is distinct from the Instrument reading (Fradin 2012, Fradin & Winterstein 2012, Villoing 2018). In this way, Means VNs behave as other participant nominals. Morphological operations such as nominalization operate indeed on the semantic participants of verbal bases, disregarding their syntactic status (Fradin & Kerleroux 2003, 2009). Since Means are semantic participants in the event denoted by the verb[3], Means VNs, similarly to Agents / Instruments / Patients VNs, are deverbal participant nominals; their interpretation relies heavily on the aspectual, argumental and semantic properties of their base verb. As we will show, however, most Means VNs also display an Instrument reading, as exemplified in (4).

(4) a.   *J'ai couvert mes pieds avec le **couvre-pied** rose.*
          'I covered my feet with the pink blanket.'
     b.   *J'ai protégé mon cahier de poésie au moyen du **protège-cahier** vert.*
          'I protected my poetry notebook by the means of the green notebook cover.'
     c.   *Cette fois-ci, j'ai bouché mon nez non pas avec les doigts mais avec un **pince-nez** avant de plonger du plongeoir de 5 mètres.*
          'This time, I blocked my nose not with my fingers but with a nose clip before jumping from the 5 meters diving board.'

In this article, we hypothesize that the Means/Instrument polysemy in VNs originates in the properties of the verbal lexemes selected by VN compounding. We examine these properties, and show that the verbal bases involved comprise a stative as well as a dynamic subcomponent, both selected by the compounding rule. This questions the analysis according to which the verbal lexemes used as bases in Means/Instrument VN compounds are homonyms. In our view, their relationship rather pertains to polysemy (Apresjan 1974).

Our work is organized as follows. In section 2, we introduce the morphological formation schema of VN compounding and its treatment in the framework of Lexematic Morphology (Matthews 1991, Anderson 1992,

---

(1951–1966 [1ère édition], 9 vol.), Paris: P. Robert, Dictionnaires LE ROBERT), *Grand Larousse Universel de la langue française* (7 vol., 1971–1986), *Dictionnaire Général de la Langue Française*, du commencement du XVIIè siècle à nos jours (A. Hatzfeld & A. Darmesteter, 1890–1900. Paris: Librairie Ch. Delagrave), *Dictionnaire de la langue française* (Littré, 1863–1872. Paris: Hachette).

3   Since our study does not relate to syntax, we will not discuss the argumental status of Means. Observe however that their possible use as subjects (cf. (1–3)) would advocate for treating them as arguments.

Aronoff 1994, Fradin 2003, Booij 2010), and we review their semantic values. Section 3 characterizes Means (Melloni 2007, Fradin 2012) as opposed to Instrument readings, and describes VNs with a Means interpretation and their regular shift to an Instrument value. In section 4, we show that the double Means/Instrument value of VNs relies on the properties of their base verbs, whose structure comprises a dynamic as well as a stative subcomponent (Kratzer 2000, Rothmayr 2009), as is illustrated in (5).

(5) $[_{vP}\ v_{DO}\ [_{vP}\ V_{CAUSE}\ [_{vP}\ V]]]$

We further hypothesize that it is the selection of both subcomponents by the compounding rule that produces the double value in question. In section 5, we offer some clues about the mechanism responsible for the appearance of the Means/Instrument value, as opposed to "pure" Instrument (i.e. deprived of any other value), and we suggest in section 6 that it relies on the (non) reversibility of the state characterizing the referent designated by N after the occurrence of the eventuality described by V. In section 7, we study the properties of "pure" Means VNs and argue that they result from a specific formation pattern, in which the verbs do not receive a literal meaning.

2. French VN compounds

VN compounding is a common morphological schema of word formation consisting in the formation of a nominal lexeme by the combination of a verbal lexeme and a nominal lexeme (i.e. uninflected verbal and nominal bases), as schematized in (6):

(6) $Lx1_V + Lx2_N = Lx3_N$

Since VN compounding constructs nominal lexemes on lexeme bases and that they share the variety of interpretations of nominalizations built by suffixation or conversion, it is considered as a morphological schema of word formation (see Corbin 2005, Villoing 2009, 2012, Fradin 2009), common to all Romance languages (Gather 2001, Ricca 2015) (7).

(7) Verb-Noun compound 'corkscrew' (lit. pull-cork)
     tire-bouchon (French); *cavatappi* (Italian); *llevataps* (Catalan); *sacacorchos* (Spanish); *saca-rolhas* (Portuguese)

VNs are exocentric compounds (see the classification of Bisetto & Scalise 2005, and Scalise & Bisetto 2009), where the verbal and nominal bases are prototypically in a transitive predicate-patient relation. VN compounding mostly forms nominals denoting participants in the eventuality described by

their verbal base[4] (see Villoing 2009 and Villoing 2012 for more details). This is illustrated in (8). Event denotation, while available, as in (9), is infrequent.

(8)   a. Agent: *brise-image* 'iconoclastic' (lit. break-picture)
      b. Cause: *tue-mouche* 'fly agaric' (lit. kill-flies) (mushroom)
      c. Instrument: *tire-bouchon* 'corkscrew' (lit. pull-cork)
      d. Instrument-Causer: *chauffe-eau* 'water heater' (lit. heat-water)
      e. Location: *rince-doigts* 'finger bowl' (lit.rince-fingers)
      f. Patient: *broute-biquet* 'honeysuckle' (lit. graze-goat)
      g. Experiencer: *souffre-douleur* 'whipboy' (lit. suffer-pain)

(9) *baisemain* 'handkissing' (lit. kiss-hand)

Deverbal nouns built by derivation or compounding do not necessarily describe arguments; VN compounding, like *-eur* or *-oir* suffixation, for example, can nominalize syntactic arguments and produce Agent nominals (10b), but also participants such as Instruments (10c), which are implied by the verb meaning but lack an argumental status (see, among other, Namer & Villoing 2008, Villoing 2009, Ferret & Villoing 2015, Huyghe & Tribout 2015).

(10)   a. *gratter* 'to scratch'
       b. Agent: *gratte-papier* 'paper pusher' (lit. scratch-paper); *gratteur* 'scratcher'
       c. Instrument: *gratte-dos* 'brush for the back' (lit. scratch-back); *grattoir* 'scraper'

As we show in the following section, among the 1473 VNs of our corpus, 536 (36%) display a Means interpretation.

3. Means VNs

In this section, we describe the Means interpretation, and show that Means nominals are built on stative verbal bases.

3.1. Means interpretation

The identification of a Means interpretation for deverbal nominals is due to Bierwisch (1991), Melloni (2007) and Fradin (2012), who observed it as an additional semantic value of (Event/)Result nominals, in the terminology of

---

4   As a matter of fact, it is crucial to distinguish between the intrinsic semantic value of the nominal and the role the NP it heads plays in the sentence in which it occurs. The former is determined by the properties of its verbal base, while the latter relies on the verb on which it syntactically depends. For example, while *rince-doigt* 'finger bowl' intrinsically denotes a location (cf. *se rincer les doigts dans un rince-doigt* 'to rinse one's fingers in a finger bowl'), due the properties of its V and N components, it can head a NP receiving a Patient role in a sentence such as *Le rince-doigt est tombé.* 'The finger bowl fell'.

Grimshaw (1990). Fradin (2012) defines Means as non-dynamic entities participating into stative eventualities in which they act as inert causes. In other words, Means can be analyzed as stative performers. As such, they occur in the subject position of stative sentences, and are built on stative Vs (11).

(11)   *protection* 'protection/guard' from V *protéger* 'to protect'
   a.   *La protection des genoux lors de la pratique du ski est indispensable.* [Event N]
        'The protection of knees while skiing is essential.'
   b.   *Le skieur porte des protections aux genoux.* [Means N]
        'The skier wears knee guards.'
   c.   *Ses nouvelles protections lui ont protégé (efficacement / *lentement) les genoux et lui ont évité une fracture.* [Means N]
        'Her new knee guards protected her knees (efficiently /*slowly) and prevented fracture.'

The identification of VNs with Means value, exemplified in (12), is due to Villoing (2018), who noticed that VN compounds can be built on stative bases, contrary to previous observations. Note from example (12d) that Means can refer to humans.

(12)   a. *pince-nez* 'nose clip'[5] (lit. pinch-nose) from *pincer* 'to pinch'
        Le pince-nez que j'ai emprunté à Elise pince vraiment trop fort le nez.
        'The nose clip that I borrowed from Elise pinches the nose really hard.'
   b. *couvre-pieds* 'blanket' (lit. cover-feet) from V *couvrir* 'to cover'
        Le couvre-pieds rose me couvrait aussi bien les pieds que les jambes et me tenait bien chaud.
        'The rose blanket covered my feet as well as my legs and kept me warm.'
   c. *protège-cahier* 'notebook cover' (lit. protect-notebook) from V *protéger* 'to protect'
        Le protège-cahier vert a bien résisté et a protégé son cahier de poésie toute l'année.
        'The green notebook cover withstood well and protected his poetry notebook for the whole year.'
   d. *porte-flingue* 'gunman (lit. carry-gun) from V *porter* 'to carry'
        A cause des flingues qu'ils portaient, les porte-flingues ont été repérés aux portiques de sécurité de l'aéroport.
        'Because of the gun they carried, the gunmen were detected at the security checkpoint of the airport.'

---

5   A methodological note about the use of syntactic tests to reveal the semantic properties of the deverbal morphological constructs under consideration is in order here. In our examples, the same verbs are used as sentential verbs and as verbal bases in VNs. The Means role appears when the main verb is stative, while the Instrument role is activated when the main verb is dynamic. This will be discussed in more detail in section 4.

In addition, Villoing (2018) observed that the base verbal lexemes represent two homonyms – a stative and a dynamic one, as we will see in 4.2. As she noted, this explains why VNs with stative bases had not been previously identified.

3.2. The stativity of verbal bases in VN compounding

The identification of Means VNs relies on the stativity of their verbal bases. The verbs used in such VNs reject both the progressive *en train de*, and dynamic adverbs (De Miguel, 1999, Márin & McNally 2011, Fábregas & Marín 2012), while the sentences in which they appear do not admit eventive anaphoric reference with *cela s'est passé* 'this happened' (Maienborn 2005: 285–286). This is shown in (13–15).

(13) a. *{Mon pince-nez / mes lunettes} {est / sont} en train de pincer mon nez.*
          'My {nose clip / glasses} {is / are} pinching my nose.'
      b. *{Le couvre-pieds / l'édredon} est en train de couvrir ses pieds.*
          [comp. (12b)]
          'The {blanket / quilt} is covering his feet.'
      c. *Le {protège-cahier / papier kraft} est en train de protéger le cahier.* [comp. (12c)]
          'The {notebook cover / kraft paper} is protecting the notebook.'

(14) a. *{Le pince-nez / les lunettes} pince(nt) lentement le nez.*
          'The {nose clip / glasses} pinch(es) my nose slowly.'
      b. *{Le couvre-pieds / l'édredon} a vite couvert ses pieds.*
          'The {blanket / quilt} quickly covered his feet.'
      c. *Le {protège-cahier / papier kraft} protègera rapidement le cahier.*
          'The {notebook cover / kraft paper} will rapidly protect the notebook'

(15) a. *{Le pince-nez / les lunettes} {a/ont} pincé mon nez. *Cela s'est passé pendant la soirée.*
          'The {nose clip / glasses} pinched my nose. This happened during the evening.'
      b. *{Le couvre-pieds / l'édredon} m'a couvert les pieds. *Cela s'est passé pendant la nuit.*
          'The {blanket / quilt} covered my feet. *This happened during the night.'
      c. *Le {protège-cahier / papier kraft} a protégé le cahier. *Cela s'est passé pendant le cours.*
          'The {notebook cover / kraft paper} protected the notebook. *This happened during the lesson.'

The behavior of *pincer* 'to pinch', *couvrir* 'to cover' and *protéger* 'to protect' above qualifies them as stative[6]. Consequently, we can analyze *pince-nez* 'nose clip', *couvre-pieds* 'blanket' and *protège-cahier* 'notebook cover' as Means, similarly to nouns that are not VN compounds such as *lunettes* 'glasses', *édredon* 'quilt' and *papier kraft* 'kraft paper'.

It appears, however, that the stative verbal bases used in the above VNs also have dynamic homonymous lexemes, which cooccur with agentive subjects (Villoing 2018), and combine with the progressive and dynamic adverbs, as in (16). Due to their different argument structure and selectional properties, Villoing analyzes them as homonyms, following Fradin & Kerleroux's (2009) and Fradin's (2012) theoretical perspective. In this view, each semantic value corresponds to a different lexeme, although both share the same inflectional properties (Fradin & Kerleroux 2003; Bonami & Crysmann 2018).

(16)   a.   *Le peintre a peint la harpiste*$_{Agent}$ *en train de pincer les cordes de son instrument.*
            'The artist$_{Agent}$ painted the harpist plucking the strings of her instrument.'
        b.   *L'infirmier*$_{Agent}$ *a rapidement couvert les jambes du malade pour qu'il ait chaud.*
            'The nurse$_{Agent}$ quickly covered the patient's legs to keep him warm.'
        c.   *Le vendeur*$_{Agent}$ *est en train de protéger le vase avec du papier bulle.*
            'The seller$_{Agent}$ is protecting the vase with bubble wrap.'

Villoing (2018) concludes that VN compounding, while preferably selecting the homonymous lexeme with a stative value (cf. *ouvre-bouche*$_{Means/Instr}$ 'mouth opener' lit. open-mouth, *serre-bouchon*$_{Means/Instr}$ 'cork tightener', lit. tighten-cork), may also select the dynamic homonymous verb to form instruments (e.g. *ouvre-boite*$_{Instr}$ 'can opener' lit. open-can, *serre-écrou*$_{Instr}$ 'nut tightener' lit. tighten-nut) or agents (e.g. *ouvre-route*$_{Agent}$ 'leader cyclist' lit. open-road, *serre-frein*$_{Agent}$ 'brakeman' lit. apply-brakes). The agentive and instrumental VNs are built on the dynamic lexemes (such as i.e. *couvrir* in (16b)) whereas Means VN are built on a homonymous lexeme *couvrir* that heads a stative construction as in (1b).

However, as we are going to show, most VNs with Means interpretation also qualify as Instruments, a property that Villoing (2018) had not considered. This leads to a revision of the theoretical position previously adopted about such homonymous lexemes.

---

6   According to Rothmayr (2005), as well as Villoing (2018), the verbs *couvrir* 'to cover' and *protéger* 'to protect' qualify as K(imian)-states in Maienborn's (2005) terminology.

4. The Instrument value of Means nominals

In this section, we show that the above Means VNs also qualify as Instruments.

4.1. Characterizing Instruments

The Instrument role of deverbal nominals has been studied by various authors (see, for French Namer & Villoing 2008, Fradin 2012, Fradin & Winterstein 2012, Ferret & Villoing 2015, Huyghes & Tribout 2015), who consider them as inert entities —typically artefacts— used by agents to perform some action. According to Alexiadou & Schäfer (2006), instruments need to be constantly manipulated by agents, and cannot act on their own[7]. Due to this requirement of constant manipulation, Grimm (2007) considers Instruments as mobile entities. These characteristics distinguish them from Means, which are defined as inert by Fradin (2012) (see 3.1.), and as inherently non-dynamic by Melloni (2007: 104). Since Instruments rely on agentivity, which is a property of dynamic verbs, they necessarily occur in dynamic sentences, while Means occur in stative sentences. Thus, when deverbal, Means and Instruments are distinguished by the [± dynamic] value of their base verbal lexeme. Finally, Instruments are not affected (i.e. modified) by the events into which they participate (see Fradin & Winterstein 2012), a property that Grimm (2007) labels *persistence*. As a result, unlike Means, Instruments cannot denote substances[8]. Thus, in (17) *de la colle forte* 'strong glue', although introduced by *avec* 'with', cannot be characterized as an Instrument since it is not qualitatively persistent throughout the event.

(17) *Sarah a collé le ruban avec de la colle forte.*
      'Sarah bonded the ribbon with strong glue.'

Now, some of the tests used to discriminate Instruments produce similar results when applied to Means. Consequently, the above properties have to be kept in mind when distinguishing them. Let us now describe the tests used to characterize Instruments. First, they occur in PPs introduced by the prepositions *avec* 'with' and *au moyen de* 'by means of' (Namer & Villoing 2008). Second, they occur as objects of *utiliser* 'to use' in the structure *utiliser .. pour V* 'to use .. to V' (Namer & Villoing 2008, Huyghe & Tribout 2015). Finally, as noted by Alexiadou & Schäfer (2006), contra Fillmore (1968) and subsequent literature, the use of Instruments is restricted in the subject position of S-level

---

7   On this basis, Kamp & Rossdeutscher (1994) distinguish instruments strictly speaking from instrument-causers, such as *medicine*, that can act independently from agents once applied or set in motion. Contrary to instruments, instrument-causers need not be persistent. See Section 6.

8   We thank the reviewer who draw our attention to this property.

sentences⁹, while they appear unrestrictedly as subjects in I-level sentences. As for VNs, their use as subjects in S-level sentences, while not being ungrammatical, is frequently odd, unless the main verb appears as an infinitive introduced by *permettre* 'to enable', 'to make something possible'.

The following examples illustrate these tests, showing that the compound *presse-agrumes* 'citrus juicer' (lit. squeeze-citrus) behaves as an Instrument.

(18) a. *Je suis en train de presser des oranges avec ton presse-agrumes en métal.*
           'I am squeezing oranges with your metal citrus juicer.'
      b. *J'ai utilisé un presse-agrumes pour presser les oranges.*
           'I used a citrus juicer to squeeze the oranges.'
      c. *Un presse-agrumes presse les oranges comme les citrons* [I-level / *S-level]
           'A citrus juicer squeezes oranges as well as lemons.'
      d. *Ce presse-agrumes m'a permis de presser les oranges pour ton petit déjeuner ce matin.* [S-level]
           'The citrus juicer enabled me to squeeze oranges for your breakfast this morning.'

4.2. Means nominals with an Instrument value

As shown in section 3.2., VNs such as *couvre-pied* 'blanket', *pince-nez* 'nose clip' and *protège-cahier* 'notebook cover' behave as Means, due to the stative value of their base verbs. Recall, however, that their verbal bases also display a dynamic value (16). Yet, as was shown in (13–15), the use of these VNs prevents a dynamic interpretation of the main verbs in S-level sentences.

However, the examples in (19–21) show that *couvre-pied* 'blanket', *pince-nez* 'nose clip' and *protège-cahier* 'notebook cover' also react positively to the instrument tests provided above and display the properties described in section 4.1.

(19) a. *Pendant qu'elle dormait, j'ai doucement couvert ses pieds {avec le couvre-pieds en laine / au moyen du couvre-pieds en laine}.*
           'While she was sleeping, I gently covered her feet {with / by the means of} the woolen blanket.'
      b. *J'ai utilisé un couvre-pied en laine pour couvrir mes pieds cette nuit.*
           'I used a woolen blanket to cover my feet last night.'
      c. *Quand je l'ai posé correctement, le couvre-pieds m'a bien couvert les pieds.*
           'When I positioned it correctly, the blanket covered my feet well.'

---

9 While S-level predicates are located in time, I-level predicates describe permanent properties. Although originally introduced by Carlson (1980) for predicates, this notion has been extended to sentences by Diesing (1992).

(20) a. *Pour éviter que l'eau n'entre par leurs narines, les nageuses de natation synchronisée se sont bouché le nez {avec / au moyen d'} un pince-nez.*
'To prevent water from entering their nose, the synchronized swimmers blocked their nose {with / by the means of} nose clips.'

 b. *Les nageurs olympiques ont tous utilisé un pince-nez en plastique flexible pour se boucher le nez.*
'The olympic swimmers all used flexible plastic nose clips to block their nose.'

 c. *Ces nouveaux pince-nez très efficaces ont permis aux nageurs olympiques de se boucher le nez pendant la compétition.*
'These very efficient new nose clips enabled the olympic swimmers to block their nose during the competition.'

(21) a. *L'institutrice est en train de protéger tous les cahiers des élèves {avec / au moyen de} protège-cahiers transparents.*
'The teacher is protecting all the pupils' notebooks {with / by the means of} transparent notebook covers.'

 b. *Il a utilisé un protège-cahier vert pour protéger mon cahier de poésie.*
'He used a green notebook cover to protect my poetry notebook.'

 c. *Une fois mis sur le cahier de poésie, le protège-cahier l'a bien protégé.*
'Once set on the poetry notebook, the notebook cover protected it well.'

As can be seen from (19a, 20a, 21a), the VNs previously analyzed as Means can be introduced by the prepositions *avec* 'with' and *au moyen de* 'by the means of', used for Instruments; (19b, 20b, 21b) show that they also occur as subjects of *utiliser* 'to use'. We can see in (20c) that the compounds can be used as subjects of *permettre* 'to enable', 'to make possible' in S-level sentences, similarly to Instruments (16d). In the same manner, (19c, 20c) indicate that they occur as subjects if dynamicity is explicitly specified. Dynamicity markers such as dynamic adverbs and the progressive in (19, 20) exclude a stative interpretation.

To summarize, we have seen that, provided the sentential verbs they depend on are dynamic, *couvre-pied, pince-nez*, and *protège-cahier* pass the Instrument tests, besides the Means tests. This double Means/Instrument value is quite systematic for the VNs we examined. Consequently, this interpretation does not rely on the context, but is lexically constrained; it depends indeed on (i) the stativity/dynamicity of the base verb and (ii) the semantic relation between the base verbal and nominal lexemes. Our analysis supports Melloni's (2007) and Jezek & Melloni's (2009) analysis of the event/result polysemy of nominalizations. These authors indeed argue that this phenomenon relies on the semantic structure of the base verb and the semantic properties of the morphological schema of nominalization, rather than on a

semantic shift from the event sense due to the context.[10] This is why we will refer to these nominals as 'Means/Instrument VNs'.

These observations crucially challenge Villoing's (2018) claim that the stative and dynamic readings are conveyed by homonymous verbs. Rather, they motivate an analysis according to which the two readings pertain to the same verb, as we are going to show in section 5.

5. The verbal origin of Means/Instrument interpretation

5.1. The properties of the verbal bases

The fact that the double Means/Instrument reading of the VNs under consideration is quite regular leads us to discard an analysis based on verbal homonymy for two reasons. First, homonymy is relatively fortuitous, contrary to polysemy, which is more systematic. Second, such an analysis would amount to neglecting the semantic proximity between the stative and the dynamic values, which describe events implying the same participants, and can be seen as successive: the stative event follows the dynamic one, as its consequence. Thus, instead of considering two different lexemes, one stative and one dynamic, as Villoing (2018) proposed, we rather analyze both values as pertaining to a unique verbal lexeme, due to the semantic relationship they stand in. To capture both their formal (i.e. flexional and phonological) identity and semantic proximity, we adopt a decompositional analysis of these base verbs (Hale & Keyser 1993), and claim that the aspectual properties and argument structures proper to each value depend on how many and which subevents they include.

More precisely, we consider that the verbs discussed here comprise both a stative and a dynamic subcomponent (Kratzer 2000, Rothmayr 2009), as illustrated in (22b), adapted from the full argument structure in (22a), from Rothmayr (2009: 48–49).

(22)   a. $[[_{vP} DP_{Agent} v_{DO} [_{VP} DP_{Causer} V_{CAUSE} [_{VP} DP_{Patient} V]]]$
       b. $[_{vP} v_{DO} [_{VP} V_{CAUSE} [_{VP} V]]]$

Let us first describe the internal structure of the verbs, common to (22a) and (22b). The V labelled CAUSE represents a causative relation between an event and a state and is internal to the verb meaning (Wunderlich 1997). The lower VP introduces the stative situation. Finally, the v labelled DO introduces agentivity. The structure is exemplified in (23), from Rothmayr

---

10 This seems to be also the case for most of the Result nominals with Means value presented in Fradin (2012), cf. *protection* in (11), *revêtement* 'coating', *déguisement* 'disguise', etc.

(2009: 49). Observe that the stative causer *the truck* corresponds to what we label Means.

(23)   a.   Irmi obstructs the street with her truck.
       b.   [$_{vP}$ [Irmi] v$_{DO}$ [$_{VP}$ [with her truck] V$_{CAUSE}$ [$_{VP}$ [the street] obstruct$_V$]]]

We can now describe the argument structure in (22a), in relation with example (23). In (23b) the Agent *Irmi* is introduced in the specifier of the agentive head DO, and the cause of the obstruction, i.e. [the truck], merges in the specifier of the causative verb[11]. The patient *the street* appears as a complement in the projection of the stative lexical V *obstruct*. Such a structure describes a transition to a state in which the street is obstructed, caused by the Agent *Irmi* by means of her truck.

Now, the structure in (23b) is syntactic; it aims to describe the sentence in (23a) and comprises a verb and its arguments. In this work, we are not concerned with the structure of sentences, but with that of compounds, which are morphological units. However, we consider that the semantic components and relations described in (22–23) also characterize morphological units, including compounds. Let us now go back to (22b) and show how it can describe the verbal component in VN compounds.

Recall that Means are inert causes of states; as such they rely on the presence of CAUSE and of the lower stative V. Since Instruments rely on agentivity (see 4.1.), we consider that the presence of the agentive DO$_v$ accounts for the Instrument value of these VNs. If we take Means/Instrument VNs to comprise the whole structure in (22), their double value is expected, due to the presence of both CAUSE and DO in the structure of their base verbs.

Further observations support such a view. First, as noted by Villoing (2018), no strictly stative verbs are used as Means VN bases. The most frequent verbal bases occurring in their formation can be found in the classes of support and holding verbs (cf. *porter* 'to carry' > *porte-bébé* 'baby carrier', lit. carry-baby), as well as obstruct verbs (cf. *cacher* 'to hide' > *cache-pot* 'planter', lit. hide-pot; *boucher* 'to block up' > *bouche-four* 'oven cover', lit. cover-oven), which are analyzed as verbs comprising a stative and a dynamic subevent by Kratzer (2000) and correspond to the structure in (22–23) in Rothmayr (2009). Similarly, Melloni (2007: 104) noted that the verbal bases producing Result nominals with Means interpretation are systematically ambiguous between an accomplishment and a stative value.

This consequently confirms that the structure in (22b) is responsible for the regular combination of Means and Instrument values in derived

---

11 Interestingly, Kratzer (2000) considers that the PPs introduced by *with*, that we analyze as Means, are Instruments independent from Agents.

nominals. Their Means value relies on the presence of a stative component in the base verb, while their Instrument value is made available by its dynamic component.

5.2. The derivation of "pure" Instruments

Some of the verbal bases used in Means/Instrument derivation also produce VNs which can only be construed as Instruments. This is the case for *presse-agrumes* 'juicer' (lit. 'squeeze citrus') illustrated in (18), and for *pince-pâte* 'pastry clip', (lit. 'crimp pastry') in (24).

(24) a. *Un pince-pâte, ça pince la pâte à tarte et ça permet d'en régulariser les bords.*
          'A pastry clip crimps pastry and is used to level the sides.'   [I-level / *S-level]

   b. *?? Le pince-pâte a pincé la pâte ce matin à 10h.*
          'The pastry clip crimped pastry at 10 this morning.'

   c. *Mon nouveau pince-pâte m'a permis de pincer et de régulariser rapidement le bord de la tarte aux pommes que j'ai faite hier.*
          'My new pastry clip enabled me to crimp pastry and level rapidly the sides of the apple pie I made yesterday.'

The examples (24a,b) show that, unless the verb *permettre* 'to make possible' is introduced in (24c), *pince-pâte* 'pastry clip' can, conversely to Means (11), only occur as subject of I-level sentences. Conversely, and as was illustrated in (13a), (14a) and (15a), the verb *pincer* 'to pinch' is the base of the Means/Instrument *pince-nez* 'noseclip'. As for *presser* 'to squeeze', in addition to the instrument *presse-agrumes* 'juicer' (lit. squeeze citrus) (18), it produces *presse-raquette* 'racket press' (lit. press racket), that has a Means value. The possible use of *presse-raquette* in subject position of a stative sentence (25a), and the fact that it can be introduced by *utiliser* 'to use' (25b) guarantee its Means/Instruments value.

(25) a. *Les presse-raquettes$_i$ qui pressaient nos vieilles raquettes en bois ont évité leur déformation pendant les mois humides.*
          'The racket press that pressed our old wooden tennis rackets prevented their warping during wet seasons.'

   b. *Tous les hivers, mon grand-père utilisait son presse-raquette pour protéger ses raquettes de tennis des déformations dues à l'humidité.*
          'Every winter, my grandfather used his racket press to protect his tennis rackets from deformations due to humidity.'

We conclude that the verbal bases used in Means/Instrument compounding produce either "pure" Instruments, or Means/Instruments. By contrast, the VNs built on strictly dynamic verbs only behave as Instruments, as expected. This is the case for *épluche-légumes* 'vegetable peeler' (lit. peel vegetables),

from *éplucher* 'to peel' and *casse-noix* 'nutcracker' (lit. crack-nuts), from *casser* 'to crack'.

(26) a. *Un épluche-légumes épluche les légumes et les fruits.* [I-level / *S-level]
          'A vegetable peeler peels vegetables and fruits.'

     b. **L'épluche-légumes épluchera les légumes pour la soupe.*
          'The vegetable peeler will peel the vegetables for the soup.'

     c. *Ce vieil épluche-légumes a permis d'éplucher facilement 3 kg de légumes ce matin.*
          Lit. 'This old vegetable peeler made possible to peel 3 kg of vegetables easily this morning.' (i.e. it made the peeling of the vegetables possible)

(27) a. *Un casse-noix casse aussi bien les noix que les noisettes.* [I-level / *S-level]
          'A nutcracker cracks nuts as well as hazelnuts.'

     b. **Ce casse-noix a cassé les noix pour le gâteau hier soir.*
          'This nutcracker cracked nuts for the cake yesterday evening.'

     c. *Ce nouveau casse-noix a permis de casser les noix pour le gâteau hier soir.*
          Lit. 'This new nutcracker made possible to crack nuts for the cake yesterday evening.' (i.e. it made the cracking of the nuts possible)

At this point, two observations are in order. First, all the VNs examined so far behave either as Means/Instrument or as "pure" Instruments. A careful study of our data has indeed shown that only a very small proportion of VNs behave strictly as Means[12]. This is the case for *coupe-vent* 'wind breaker' (lit. cut-wind) or *abat-jour* 'lampshade' (lit. 'fell day'). These VNs will be examined in section 7. Second, we need to account for the fact that the same verbal bases can derive both Means/Instruments and "pure" Instruments. In other words, the question that needs to be addressed is why we unambiguously identify a Means/Instrument value in *pince-nez* 'nose clip' (lit. pinch-nose) but only an Instrument value in *pince-pâte* 'pastry clip' (lit. 'crimp pastry'). The derivation of "pure" Instruments is indeed expected from strictly dynamic bases, that do not comprise a stative subcomponent, but unexpected in the case of verbs displaying the structure in (22b). Recall from this representation that the dynamic component stands above the stative one. Consequently, the selection by the compounding rule of the dynamic component without the stative one is impossible.

---

12 When they do not also behave as Instruments, most Means convey a locative value, cf. *garde-meuble* 'furniture storage', lit. keep-furniture, *repose-plat* 'table mat', lit. rest-plate, *porte-parapluies* 'umbrella holder', lit. hold umbrella. When occurring in PPs, Locative Means are introduced by *dans* 'in' or *sur* 'on' instead of *avec* 'with', cf. *garder ses meubles {dans / *avec} un garde-meuble* 'to keep one's furniture {in / *with} a furniture storage'.

In fact, the only clear difference between Means/Instruments and "pure" Instruments derived from such verbal bases lies in the nominals selected by VN compounding. This leads us to suggest that these are crucial in distinguishing between these VNs' semantic values. In section 6, we examine the distinctive properties of these nouns.

6. The role of N in distinguishing Means/Instrument from Instruments

In this section, we provide an explanation as to why Means/Instrument as well as "pure" Instruments can be derived from the same verbal bases, and we claim that the distinction between these values depends on the permanent vs reversible state of the entity described by N after the eventuality described by V occurred. To do so, we introduce additional tests distinguishing between those types of states. But first, since we aim at distinguishing between Means and Instrument readings, a reminder of their definitions and properties is in order. Instruments are auxiliary objects used by agents to perform actions. When deverbal, Instrument nominals denote artefacts typically used to perform the action denoted by their base verb. Contrary to Instrument-Causers (Kamp & Rossdeutscher 1994), "pure" Instruments do not act on their own. In other words, they participate in an event as long as they are manipulated by the agent. Once the event performed, the role of the agent, and consequently that of the instrument is over, and the latter can be discarded.

Thus, sentences such as (28), ending by *puis j'ai rangé* +Instr 'then I stored + Instr', in which Instruments are discarded once the actions of the predicates are over, are perfectly fine.

(28) a. *J'ai épluché les légumes avec l'épluche-légumes$_i$, puis je l'$_i$ai rangé.*
'I peeled the vegetables with the vegetable peeler$_i$, then I stored it$_i$.'
   b. *J'ai ouvert la boîte avec l'ouvre-boîte$_i$ puis je l'$_i$ai remis dans le tiroir.*
'I opened the can with the can opener$_i$, then I put it$_i$ back in the drawer.'

What is crucial here is that the actions described by the verbs entail a stable change of state of the referent of the nouns; once peeled / opened, the vegetables and the can remain so, even after the storage of the vegetable peeler or the can opener.

Now consider the case of Means. Means are involved in stative eventualities, in which they act as inert causes. Their inertness contrasts with the mobility of Instruments pointed out by Grimm (2007), see section 4.1. As causers of states, Means crucially participate in stative eventualities as long as these eventualities hold. As atelic eventualities, states have no telos, or "natural ending", and if they end, it is due to external circumstances (Smith 1991). The removal of the object denoted by a Means from a state results in

such a circumstance, since it entails its interruption. Consider the examples in (29), forged on the same pattern as (28):

(29) a. *Durant la compétition de natation, j'ai bouché mon nez en le pinçant avec mon pince-nez$_i$, puis je l$_i$'ai rangé dans mon sac.*
          'During the swimming competition, I blocked my nose by pinching it with my nose clip$_i$, then I stored it$_i$ in my bag.'
      b. *Cette nuit, j'ai couvert mes pieds avec mon nouveau couvre-pieds$_i$, puis je l$_i$'ai rangé.*
          'I covered my feet with my new blanket last night, then I stored it.'

Contrary to (28), the examples in (29) imply that, after the storage of the nose clip and the blanket, the nose is no longer blocked and the feet are no longer covered. In these cases, the removal of the cause results in the interruption of the state.

   These contrasting properties provide us with an effective test to distinguish Instruments from Means: while a sentence containing a Means cannot be followed by *puis j'ai rangé* + Means' 'then I stored + Means' without implying the interruption of the state (29), such a continuation in a sentence with an Instrument does not entail a reverse of the Patient's state (28).

   These considerations lead to another observation concerning the reversibility of the Patient's state, that is, the property for the Patient to recover the state in which it was before the event occurred. As observed by Fábregas & Marín (2012), some change-of-state verbs admit the measuring of the duration of the state resulting from the event by 'for x time'. According to these authors, this is possible only if the state is reversible[13]. If it is not the case, the temporal extension, if possible, measures the duration of the event leading to the state. This is exemplified by the contrast in (30), from Fábregas & Marín (2012: 47–48):

(30) a. *The storm broke the communications down for two hours.*
      b. *The army destroyed the city for two months.*

In (30a), *two hours* is construed as measuring the duration of the break up, that is, of the state of the communications resulting from the breaking event. Put differently, it measures the duration of the broken state in which the communications are due to this event. This reading is made possible by the reversible character of the state in question. Conversely in (30b), *two months* can only measure the duration of the change of state (i.e. of the event), because the destruction of a city is perceived as non-reversible. The reading under which it would measure the duration of the destroyed state of the city is unavailable. Consider now the contrast between (31) and (32):

---

13 We define reversibility as the possible return to the state holding prior to the event.

(31) a. *Ce couvre-pied m'a couvert les pieds pendant deux heures (*lentement)*[14].
        'The blanket has covered my feet for two hours (*slowly).'
        ⇒ entails: 'My feet stayed covered for two hours.'

    b. *En me pinçant le nez (*rapidement), mon pince-nez m'a bouché le nez pendant 30 minutes*
        'By pinching my nose (*quickly), my nose clip blocked my nose for 30 minutes.'
        ⇒ entails: 'My nose stayed blocked for 30 minutes.'

    c. *Le protège-cahier vert a (*vite) protégé le cahier de poésie pendant 6 mois.*
        'The green notebook cover protected the poetry notebook for 6 months (*quickly).'
        ⇒ entails: 'The poetry notebook stayed protected for 6 months.'

(32) a. *Mon vieux presse-agrume m'a quand même permis de presser des oranges pendant deux heures ce matin!*
        'My old citrus squeezer still enabled me to squeeze oranges for two hours this morning!'
        ⇒ does not entail: 'The oranges stayed squeezed for two hours.'

    b. *Ce vieil épluche-légumes a permis d'éplucher ces légumes pendant 15 mn avant de se casser.*
        Lit. 'This old vegetable peeler has made possible to peel these vegetables for 15 mn before breaking apart.' (i.e. it made the peeling of the vegetables possible)
        ⇒ does not entail: 'The vegetables stayed peeled for 15 mn.'

    c. *Ce nouveau pince-pâte m'a permis de froncer le bord de ma tarte {en à peine 1 minute / *pendant 1 minute} pour la mettre au four sans tarder.*
        'This new pastry clip enabled me to crimp the sides of my pie (in one minute / *for one minute) and put it in the oven immediately.'
        ⇒ does not entail: 'The pie's sides stayed crimped in 1 minute.'

In (31) *pendant x temps* 'for x time' measures the duration of the state in which the Patient (i.e. *pieds* in *couvre-pieds, nez* in *pince-nez, cahier* in *protège-cahier*) is after the event occurred, thus patterning like (30a). We conclude that the nouns used in Means compounding describe entities whose states after the events described by the verbs occurred are reversible.

By contrast, *pendant x temps* in the examples (32) either measures the duration of the squeezing and peeling events (32a, 32b), similarly to (30b), or is ill-formed (32c). We assume that the impossibility for the temporal expressions to pertain to the states of the Patients (*agrumes* in *presse-agrumes,*

---

14 The unacceptability of the dynamic adverbs *lentement* 'slowly' *rapidement* 'rapidly' and *vite* 'quickly' in the examples (29) guarantees that the events described are not construed as dynamic, since the verbs *couvrir* 'to cover', pincer 'to pinch' and *protéger* 'to protect' have a dynamic as well as a stative value, see (19-21).

légumes in *épluche-légumes* and *pâte* in *pince-pâte* 'pastry clip' (lit. 'crimp pastry') after the events occurred is due to their irreversibility.

Now, the question of the internal structure of "pure" instrumental compounds built on the same verbal bases as Means/Instrument, such as *pince-pâte* 'pastry clip', has to be raised. Recall indeed from (22) that the agentive part of the verbal structure, which legitimates Instruments, dominates the causative structure. In other words, "pure" Instrument and Means/Instrument derivations from similar verbal structure are unexpected. However, as we have just seen, the "pure" Instrument vs Means/Instrument reading of VNs relies on the permanence or reversibility of N state after the occurrence of the event. In other words, it stems from the relationship between V and N. Consequently, we suggest that this difference is encoded in some manner at the lower VP level, where both V and N merge. For example, one could consider that V can come with a [±R] (for Reversible) feature depending on its object. Thus, *pâte* in *pince-pâte* 'pastry clip' would be [-R], but [+R] in *pince-nez* 'nose clip'[15].

7. "Pure" Means?

The above discussion mostly focused on VNs with a Means/Instrument value, and, to a lesser extent, on purely instrumental VNs. However, as we mentioned in section 5.1., a few cases of VN strictly behaving as Means ("pure" Means) have been found in the data under study. This is the case for *coupe-vent* 'wind breaker' (lit. cut-wind) and *abat-jour* 'lampshade' (lit. fell day). Observe that both *jour* 'day' and *vent* 'wind' describe natural forces, and that *couper* 'to cut' and *abattre* 'to fell', 'to kill' are destruction verbs. Since verbs of this class are dynamic, this would lead us to expect them to produce Agent, Instrument or Instrument-Causer nominals instead of "pure" Means. This pattern is however regular, as shown by *brise-lame* 'breakwater' (lit. break-wave), *brise-bise* 'curtain' (lit. break-wind), *tue-vent* 'shelter' (lit. kill-wind) *coupe-feu* 'firewall' (lit. cut-fire). Semantically, these VNs are causes of states: *coupe-vent* 'wind breaker' causes protection from the wind, and *abat-jour* 'lampshade' prevents the diffusion of light. The objects denoted by these nouns cause the states of being in the shadow (for *abat-jour*) and unaffected by the wind (for *coupe-vent*). As before, the removal of the objects entails the interruption of the states, which are reversible, a typical property of VNs with Means value (see section 6). Moreover, all these VNs fail the Instrument tests, as illustrated in (33) with *abat-jour*.

---

15 According to Fábregas & Marín (2012), non-reversible states are not encoded in the semantic structure of verbs and are only pragmatically inferred. We consider by contrast that the verbal bases of VNs always include a stative component, which is the result of the causing event.

(33) a.   *J'ai abattu le jour {avec / au moyen d'} un abat-jour.
           Lit. 'I fell the day {with / by the means of} a lampshade.'
      b.   ??J'ai utilisé un abat-jour pour abattre le jour.
           Lit. 'I used a lampshade to fell the day.'

Now, since the VNs in question here are inert causes, we expect them to pass the test for Means, that is, they should be able to occur as subjects of stative verbs. Consider, however, the examples in (34):

(34) a.   ?? L'abat-jour a abattu le jour pendant la soirée.
           Lit. 'The lampshade felled the day during the whole evening.'
      b.   ?? Le brise-bise a brisé la bise pendant la tempête.
           Lit. 'The curtain broke the wind during the storm.'
      c.   ?? Le coupe-feu a coupé le feu pendant 2 heures.
           Lit. 'The firewall cut the fire for two hours.'

The unacceptability of the above examples contrasts with the well-formedness of Means/Instrument VN (cf. *protège-cahier* 'notebook cover' *couvre-pieds* 'blanket', etc.) in the subject position of stative sentences (13, 14, 15). The examples (33) are ill-formed due to the fact that the verbs *briser* 'to break', *abattre* 'to kill' and *couper* 'to cut' do not select nouns such as *jour* 'day', *bise* 'wind' and *feu* 'fire' when used in sentences. In the VNs under discussion, these verbs are not used with their usual meaning[16]. Due to these interpretive properties, these verbs behave as blocking verbs[17] in this context. They are construed as *arrêter* 'to stop', whose use as a main verb would restore the acceptability of the examples in (34). We can consequently conclude that we are facing here rare (10 to 15 cases) and marginal construction patterns of lexeme coining, relying on phenomena such as metaphors, meaning narrowing, peculiar sociolects, and responding to onomasiological needs of the social community[18]. This shows moreover that the structure in

---

16 The fact that the relationship between the V and the N lexemes does not always reflect the syntactic relationship between the verb and its object provides an additional argument in favor of a morphological (vs syntactic) formation of French VN compounds. This is also the case for compounds such as *broute-biquet* 'honeysuckle' (lit. graze-goat) (8g), in which the N *biquet* 'goat' corresponds to the subject of the V *brouter* 'to graze', *réveille-matin* 'alarm clock' (lit. wake-morning) ou *cuit-vapeur* 'steam cooker' (lit. steam-cook), in which the Ns refer to a temporal span and a cause respectively. See Villoing (2009, 2012), Desmets & Villoing (2009), Fradin (2009).

17 We thank the reviewer who brought this property to our attention.

18 We leave for future research the question whether these data deserve a conceptual or cognitive analysis.

(35), corresponding to the non-agentive value of the verbs described in (22), is not a verbal base available in normal cases of VN compounding.

(35) $[_{VP}\ V_{Cause}\ [_{VP}\ V]]$

8. Conclusion

In this article, we have shown that French exhibits VN compounds with a Means value. However, "pure" Means VNs are infrequent, and rely on a non-literal use of the base verb. Means VNs in fact regularly display an additional Instrument reading. The regularity of this double Means/Instrument value has led us to discard the analysis that formally similar lexemes occurring in several constructions are homonyms, which has been put forward by Fradin & Kerleroux (2003, 2009) in the framework of Lexematic Morphology. Even though we do not question the validity of this proposal when applied to other cases, we have seen that it cannot handle the data that we examined here. The notion of polysemous lexeme has proved a more appropriate manner to account for both the relationship between Means and Instrument readings of the VNs, and the stative / dynamic values of the corresponding base verbs. That is why we adopted Kratzer's (2000) and Rothmayr's (2009) analysis, which is fully consistent with such an approach. This analysis is recalled in (36=22a):

(36) $[_{vP}\ DP_{Agent}\ v_{DO}\ [_{VP}\ DP\ V_{CAUSE}\ [_{VP}\ DP\ V]]]$

By postulating the presence of both an agentive and a stative/causative component in the verbal bases, (36) accurately accounts for the double Means/Instrument value of the VNs.

As for the selection of the "pure" Instrument reading, also available for VNs built on the same verbal bases, we argued that it relies on the permanent state of the Patient described by the Ns in the compounds after the occurrence of the event described by the verbal base. Conversely, the state of the Patient is reversible when VNs have a Means/Instrument reading. To distinguish between these two types of states, we provided two tests. First, after Fábregas & Marín (2012), we showed that reversible states can only be measured by *pendant x temps* 'for x time'. Second, we put forward a new test, based on the use of *puis j'ai rangé* +Instr 'then I stored +Instr', which entails the termination of the state of N, and consequently its reversibility, when the VNs describe Means, but not if they describe "pure" Instruments.

These results raise further questions for future research, such as the availability of the Means value for instrumental nominals in *-eur* (cf. *aérateur* 'aérator') or *-oir* (*bouchoir* 'oven door').

References

Alexiadou Artemis & Florian Schäfer. 2006. Instrument subjects are Agents or Causers. In Donald Baumer, David Montero & Michael Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*, 40–48. Somerville (MA): Cascadilla Proceedings Project.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge (MA): Cambridge University Press

Apresjan, Juri. 1974. Regular Polysemy. *Linguistics* 142. 5–32.

Aronoff, Mark. 1994. *Morphology by itself*. Cambridge (MA): MIT Press.

Bierwisch, Manfred. 1991. Event nominalizations: Proposals and problems. *Acta Linguistica Hungarica* 40(1–2). 19–84.

Bisetto Antonietta & Sergio Scalise. 2005. Classification of compounds. *Lingue e linguaggio* 2. 319–330.

Bonami, Olivier, & Berthold Crysmann. 2018. Lexeme and flexeme in a formal theory of grammar. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology. Empirically Oriented Theoretical Morphology and Syntax*, vol. 4, 203–235. Berlin: Language Science Press.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Carlson, Gregory. 1980. *Reference to Kinds in English*. New York: Garland.

Corbin, Danielle. 2005. French (Indo European: Romance) In Geert Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas (eds.), *Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An International Handbook on Inflection and Word-Formation*, vol. 2, [art. 121], 1285–1299. Berlin & New York: de Gruyter.

De Miguel, Elena. 1999. El aspecto léxico. In Ignacio Bosque, & Violeta Demonte (eds.), *Gramática descriptiva de la lengua Española*, 2971–3060. Madrid: Espasa Calpe.

Desmets Marianne & Florence Villoing. 2009. French VN lexemes: morphological compounding in HPSG. In Stefan Müller (ed.), *The Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar*, Göttingen, Germany, July 16–18, 2009, Standford, CA, USA: CSLI Publications: 89–109.

Diesing, Molly. 1992. *Indefinites*. Cambridge (MA): MIT Press.

Fábregas, Antonio & Rafael Márin. 2012. The role of Aktionsart in deverbal nouns: State nominalizations across languages. *Journal of Linguistics* 48. 35–70.

Fillmore, Charles J. 1968. The Case for Case. In Emmon Bach & Robert T. Harms (eds), *Universals in Linguistic Theory*, 1–88. New York: Holt, Rinehart & Winston.

Ferret Karen & Villoing Florence. 2015. French derived instrumentals N-*age*: semantic properties of the base verb. In Ingo Plag & Sabine Arndt-Lappe (eds), *Morphology* 25, 473–496. Netherlands: Springer.

Fradin Bernard. 2003. *Nouvelles approches en morphologie*. Paris: PUF.

Fradin, Bernard. 2009. Compounding in French. In Rochelle Lieber & Pavel Štekauer (eds.), *Oxford Handbook on Compounding*, 417–435. Oxford: Oxford University Press.

Fradin Bernard. 2012. Les nominalisations et la lecture 'moyen'. *Lexique* 20. 125–152.

Fradin Bernard & Françoise Kerleroux. 2003. Troubles with lexemes. In Geert Booij, Janet De Cesaris, Sergio Scalise & Angela Ralli (eds.), *Topics in Morphology. Selected papers from the Third Mediterranean Morphology Meeting (Barcelona, September 20–22, 2001)*, 177–196. Barcelona: IULA-Universitat Pompeu Fabra. Barcelona.

Fradin, Bernard, & Françoise Kerleroux. 2009. L'identité lexémique. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (eds), *Aperçus de morphologie du français*, 85–104. Saint-Denis: Presses Universitaires de Vincennes.

Fradin Bernard & Gregoire Winterstein. 2012. Tuning agentivity and instrumentality: deverbal nouns in -*oir* revisited. Paper presented at *Décembrettes 8, International Conference on Morphology*, Bordeaux, 6–7 décembre 2012.

Gather, Andreas. 2001. *Romanische Verb-Nomen Komposita: Wortbildung zwischen Lexikon, Morphologie und Syntax*. Tübingen: Narr.

Grimm, Scott. 2007. The bounds of subjecthood: evidence from instruments. *Berkley Linguistics Society* 33(1). 178–190.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press.

Hale, Kenneth & Samuel Jay Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building* 20, 53–109. Cambridge (MA): MIT Press.

Huyghe, Richard & Delphine Tribout. 2015. Noms d'agents et noms d'instruments: le cas des déverbaux en -eur. *Langue Française* 185(1). 99–112.

Jezek Elisabetta & Chiara Melloni. 2009. Complex Types in the (morphologically) Complex Lexicon. In Nicoletta Calzolari & Anna Rumshisky (eds.), *Proceedings of the Fifth International Workshop on Generative Approaches to the Lexicon (GL2009) (September 17–19, 2009,)* Pisa, Italy, 59–67. http://www.cs.brandeis.edu/~jamesp/gl2009/GL2009Proceedings.pdf

Kamp, Hans & Anja Rossdeutscher. 1994. Remarks on lexical structure and DRS construction. *Theoretical Linguistics* 20. 97–164.

Kratzer, Angelika. 2000. Building Statives. In Lisa Conathan, Jeff Good, Darya Kavitskaya, Alyssa Wulf & Alan Yu (eds.), *Proceedings of the Twenty-Sixth Annual Meeting of the Berkeley Linguistics Society*, 385–399. Berkeley: Berkeley Linguistics Society.

Maienborn, Claudia. 2005. On the limits of the Davidsonian approach: The case of copula sentences. *Theoretical linguistics* 31(3). 275–316.

Márin, Rafael & Louise McNally. 2011. Inchoativity, change of state, and telicity. Evidence from Spanish reflexive psychological predicates. *Natural Language and Linguistic Theory* 48(1). 35–70.

Matthews, Peter 1991. *Morphology*. Cambridge: Cambridge University Press.

Melloni, Chiara 2007. *Polysemy in Word Formation: The Case of Deverbal Nominals*. Verona: University of Verona PhD. Dissertation.

Namer Fiammetta & Florence Villoing. 2008. Interpréter les noms déverbaux: quelle relation avec la structure argumentale du verbe de base? le cas des noms en -OIR du français. In Franck Neveu, Valelia Muni Toke, Thomas Klingler, Jacques Durand, Lorenz Mondada & Sophie Prévost (eds.), *Actes du 1er Colloque Mondial de Linguistique Française*, 1539–1557. Paris: ILF.

Ricca, Davide. 2015. Verb-noun compounds in Romance. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, 688–707. Berlin: de Gruyter.

Rothmayr, Antonia. 2009. *The structure of stative verbs*. Amsterdam: Benjamins.

Scalise, Sergio & Antonietta Bisetto. 2009. The classification of compounds. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 34–53. Oxford: Oxford University Press.

Smith, Carlota S. 1991.*The parameter of Aspect*. Dordrecht: Kluwer Academic.

Villoing, Florence. 2009. Les mots composés VN. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (eds.), *Aperçus de morphologie du français*, 175–198. Saint-Denis: Presses Universitaires de Vincennes.

Villoing Florence. 2012. French compounds. *Probus, International Journal of Latin and Romance Linguistics* 24(1). 29–60.

Villoing, Florence. 2018. Stative verbs and French Verb-Noun compounds: a discreet preference. In Dany Amiot, Georgette Dal & Delphine Tribout (eds.), *Tendances actuelles en morphologie* (*Lexique* 23), 90–130.

Wunderlich, Dieter. 1997. Cause and the structure of verbs. *Linguistic Inquiry* 28(1). 27–68.

Marie Laurence Knittel
Université de Lorraine & ATILF
CLSH
23 Boulevard Albert 1er
BP 60446
54001 Nancy Cedex
marie-laurence.knittel@univ-lorraine.fr

Florence Villoing
UMR 7114 MoDyCo, CNRS & Université Paris Nanterre
200 Avenue de la République
92000 Nanterre
villoing@parisnanterre.fr

Delphine Tribout

Nominalization, verbalization or both? Insights from the directionality of noun-verb conversion in French

Abstract: Nominalization in French can be done by means of conversion, which is characterized by the identity between the base and the derived lexeme. Since both noun→verb and verb→noun conversions exist, this property raises directionality issues, and sometimes leads to contradictory analyses of the same examples. The paper presents two approaches of conversion: derivational and non-derivational ones. Then it discusses various criteria used in derivational approaches to determine the direction of conversion: diachronic ones, such as dates of first attestation or etymology; and synchronic ones, such as semantic relations, noun gender or verb inflection. All criteria are evaluated on a corpus of 3,241 French noun~verb pairs. It is shown that none of them enables to identify the direction of conversion in French. Finally, the consequences for the theory of morphology are discussed.

Keywords: conversion, directionality, historical criteria, synchronic criteria, French, paradigmatic morphology

1. Introduction

In French, as in many languages, verbs can be nominalized by means of different suffixes. The most frequent ones are *–ion* (e.g. DIVISION 'division' from DIVISER 'to divide'), *–age* (e.g. LAVAGE 'washing' from LAVER 'to wash') and *–ment* (e.g. CHANGEMENT 'change' from CHANGER 'to change'). But the nominalization of a verb can also be done with a zero suffix as can be seen with the examples in (1).

(1) a. MARCHER 'to walk' → MARCHE 'a walk'
      b. ATTAQUER 'to attack' → ATTAQUE 'an attack'
      c. OUBLIER 'to forget' → OUBLI 'forgetfulness'

Cases such as in (1) have long been called *dérivation régressive* 'regressive derivation' because of the deletion of the verb's *–(e)r* ending (see Nyrop 1936). However, this ending is a mere inflectional marking on the infinitive form of the verb and plays no role in the derivation (it is also deleted before suffixes such as *–ion, –age* and *–ment*). In more recent literature, nouns in (1) are referred to as *zero derivation/zero suffixation* (Dubois 1962) or *conversion* (Corbin 1987; Kerleroux 1996, 1999; Fradin 2003). The present study will use the latter term and focus on verb→noun conversion like the examples in (1) compared to noun→verb conversion as illustrated by the examples in (2).

(2) a. COLLE 'glue' → COLLER 'to glue'
     b. POIVRE 'pepper' → POIVRER 'to pepper'
     c. SINGE 'ape' → SINGER 'to ape'

Conversion is a widely discussed phenomenon that is usually defined as the change of category of one lexeme without any change in its form. Because there is no change in the form, conversion raises specific issues regarding the direction of the derivation, as it has already been noticed by many authors (Marchand 1963, 1964; Kerleroux 1996; Balteiro 2007; Rodrigues Soares 2009, among others). Indeed, since the lexemes involved in conversion are identical, one cannot formally determine which one is the base and which one is the converted lexeme. Moreover, this sometimes leads to contradictory analyses of the same pairs. For instance, the noun~verb[1] pairs in (3) are analysed by Adouani (1989) as cases of verb→noun conversion, whereas Labelle (1992) considers them to be noun→verb conversions.

(3) a. CHASSE 'hunting' ~ CHASSER 'to hunt'
     b. DANSE 'dance' ~ DANSER 'to dance'
     c. GUIDE 'guide' ~ GUIDER 'to guide'

The aim of the present study is to discuss this directionality issue and quantify the different problems it can raise on the basis of a corpus of French noun~verb conversion pairs. Next section will outline different approaches to conversion with respect to directionality. Sections 3 and 4 will discuss various criteria to determine the direction of conversion. Finally, section 5 will draw out theoretical implications and conclusions will be presented in section 6.

2. Different approaches to conversion

There are two different theoretical ways to deal with the directionality problem: the first one is to postulate no derivational relation between the noun and the verb, so that neither derives from the other. The second one is to assume a derivational relation between lexemes and to identify criteria in order to determine the direction of the derivation. Both approaches are found in the literature.

Studies postulating a non-derivational relation are found in diverse frameworks but most of them share the common assumption of underspecified categories. For example, the analysis by Farell (2001) is carried out within the frameworks of Cognitive Grammar (Langaker 1987,

---

1 Throughout the article, conversion pairs will be presented as "noun~verb pairs" when the direction of the conversion is either not relevant or unknown.

1991) and Construction Grammar (Goldberg 1995) and relies on the under-specification of categories. According to Farrell, the meanings of words are schematic concepts compatible with both nouns and verbs and the nominal or verbal aspects of their meaning is given by the morphosyntactic context in which they appear. *Bag* and *kiss*, for instance, are neither nouns nor verbs but have a conceptual structure compatible with the meaning of both an event and a thing. Thus, according to Farrell, since words do not have categories, there is no rule deriving a noun from a verb or a verb from a noun. Barner and Bale (2002) is another study arguing for category underspecification, which relies on earlier assumptions in the Distributed Morphology framework (Halle and Marantz 1993, Marantz 1997; cf. Borer 2013). According to Barner and Bale, words derive from lexical roots that are underspecified with respect to categories. Roots are inserted into syntax under functional heads that give the nominal or verbal status to words. Thus, according to the authors, no noun is derived from verb or vice-versa, and there is no need for conversion rules.

To a certain extent, these analyses relying on category underspecification solve the problem of directionality: because there is no derivation, there is no direction. However, underspecified categories have been criticized by different studies. For instance, according to Croft (1991), words have an inherent category that is defined by two factors: their semantic class and their pragmatic function. Prototypically, nouns are words that denote an object and their pragmatic function is the reference, adjectives denote properties and their pragmatic function is to modify, while verbs denote an action and allow for predication. While the pragmatic function could be an effect of the syntactic context, the semantic class of a word does not, in Croft's view, depend on the context. According to the author, words bear a category outside any syntactic context. Don (2004) has also argued against the category underspecification in Dutch. In this language, nouns and verbs have phonological properties that distinguish them very clearly. The syllabic structure of nouns is more complex than that of verbs, which allows speakers to identify the category of a word, even outside any syntactic context. Moreover, Don ran an experiment on nonsense words that confirm speakers' ability to classify words as nouns or verbs according to their syllabic structure. In a more recent study on English, Lohmann (2017) also argues for a lexical specification of categories. The author tested various phonological properties of unambiguous nouns and verbs taken from the Celex database, such as the word length, syllabic complexity, word onset complexity, vowel height and backness, types of consonants. Results reveal that nouns and verbs differ in many phonological dimensions: of fifteen variables that have been tested, thirteen significantly allow for a distinction between nouns and verbs. All these studies question the validity of underspecified categories,

and consequently, the validity of a non-derivational relation between nouns and verbs. They rather argue for fully specified lexemes with respect to categories and thus, for a derivation from one lexeme to another.

Just like non-derivational approaches, derivational approaches to conversion are found in various frameworks: in structuralist works such as Jespersen (1942) or Bally (1944) for French, in transformational analyses such as Marchand (1963, 1964), Dubois (1962), or in Distributed Morphology (Harley 2005), but also in cognitive analyses such as those by Dirven (1988, 1999) or Labelle (1992, 2000), and in Lexemic Morphology such as Aronoff (2007), Kerleroux (1996, 1999). In these analyses, the assumption of a derivation arises from the comparison with affixation. Indeed, many studies claim that the meanings of converted words are similar to those of affixed ones (see for example Marchand 1963; Dubois 1962; Corbin 1976; Plag 1999; Don 2005) and because affixation overtly marks the derivation, the same principle is applied to conversion. In derivational analyses, the identification of the basic lexeme and the derived one is therefore a crucial issue. Approaches differ according to the authors. Hale and Keyser (1993) or Harley (2005), for instance, within the framework of Distributed Morphology, postulate that verbs always derive from nouns. This assumption, however, cannot hold in situations where the verb already derives from another lexeme. For example, in French, in the pair ALLONGE 'extension'~ALLONGER 'to lengthen', the verb already derives from the adjective LONG 'long' with the meaning 'to make longer', so that it cannot derive from the noun. Conversely, the noun cannot derive from the adjective because the prefix *a–* cannot form nouns in French. This shows that the directionality problem cannot be solved by postulating a unique direction for all cases. Analyses that assume a derivation from one lexeme to the other must set criteria to determine the direction of the derivation.

Various criteria have been proposed in the literature in order to determine the direction of conversion. They are of two types: either historical or synchronic. The next sections will present both types of criteria and evaluate them on a corpus of 3,241 noun~verb conversion pairs in French. These noun~verb pairs have been gathered from two French dictionaries: *Trésor de la Langue Française informatisé* (hereafter *TLFi*) and *Petit Robert Électronique*.

3. Historical criteria

Historical criteria are dates of first attestation and etymology. From a theoretical point of view, the use of such criteria has often been criticised. In French, Corbin (1976), for instance, disapproves the use of historical criteria because, according to her, it blurs the distinction between synchronic and

diachronic analyses. Disregarding the theoretical aspect of the question, this section aims at assessing the reliability of both criteria on empirical grounds and at quantifying cases where they can be useful or must be ruled out. In order to evaluate the criteria, a random sample of 15% of the whole dataset has been extracted. This sample gathers 483 noun~verb pairs. The date of first attestation and the etymology of each noun and verb in the sample have been collected from the reference dictionary of etymology for French: the *Dictionnaire Historique de la Langue Française* (hereafter *DHLF*) and from the historical section of the *TLFi* when needed.

3.1. Dates of first attestation

Dates of first attestation give a direction to the derivation for 463 pairs, that is, for 96% of the data. The lexeme that is attested first is considered to be the base of the conversion. According to dates, 331 pairs are noun→verb conversions and 132 are verb→noun conversions. Only 20 pairs cannot be determined by using dates. Examples are given in Table 1. In these particular cases, the direction cannot be decided because either both lexemes are attested during the same year, or the dates are not accurate enough. For instance, COMPTE and COMPTER are both attested during the year 1100. As for BAGUE and BAGUER, the datation of the verb is not accurate enough to know if it appeared before or after the noun.

　　This kind of problems has already been noticed by Marchand (1963). As they concern only a few pairs (4% of the data) one may think that dates make a good criterion. However, using dates is problematic because of their lack of reliability. Indeed, Corbin has pointed out in different works (particularly in Corbin (1976) and Corbin (1987)) that they highly depend on chance. Moreover, as it has already been noticed by Tournier (1980) and Balteiro (2007), the minimum number of years between the two dates for the interval to be reliable is debatable.

Tab. 1: Examples of indeterminacy due to first attestation.

Noun	1ˢᵗ attestation of noun	Verb	1ˢᵗ attestation of verb
BAGUE 'ring'	1416	BAGUER 'to ring'	15th century
COMPTE 'count'	1100	COMPTER 'to count'	1100
CONJECTURE 'conjecture'	1246	CONJECTURER 'to conjecture'	13th century
PAGNOT 'bed'	end of 19th century	PAGNOTER 'to go to bed'	1859
PROFIL 'profile'	1621	PROFILER 'to profile'	1621
TRACE 'mark'	1120	TRACER 'to draw'	1120

More importantly, dates of attestation raise a problem never discussed so far: they are sometimes inconsistent with the morphological construction of lexemes. For example, LOUANGER 'to commend' is attested in 1155 and LOUANGE 'praise' in 1160. According to these dates, the pair should thus be a verb→noun conversion. However, the noun LOUANGE already derives from the verb LOUER 'to praise' by means of the old suffix –*ange*, just like VIDANGE 'emptying' derives from VIDER 'to empty' or MÉLANGE 'mixing, mixture' from MÊLER 'to mix'. Moreover, there is no suffix –*anger* that could form a verb out of a verb. Therefore, the morphological analysis of the pair indicates a noun→verb conversion (LOUER →)LOUANGE→LOUANGER, in contradiction with the chronological analysis. The problem is similar with CHARROI 'convoy, carting' and CHARROYER 'to carry along'. Since the noun is first attested in 1150 and the verb in 1225, the pair should be a noun→verb conversion. Yet, the verb morphologically derives from the noun CHAR 'cart' by means of the suffix –*oyer*, like FESTOYER 'to feast' from FÊTE 'party', GUERROYER 'to wage war' from GUERRE 'war', FOUDROYER 'to strike' from FOUDRE 'lightning' etc. The morphological analysis thus indicates a verb→noun conversion (CHAR→)CHARROYER→CHARROI, in contradiction, once again, with the chronological analysis. This inconsistency between the morphological analysis and the attestation of lexemes is not specific to conversion and can be observed with suffixation too. For instance BIFFURE 'crossing-out', which derives from the verb BIFFER 'to cross out' with the suffix –*ure*, is attested in 1580 while the base verb is attested in 1584. In the case of suffixation, we can doubt that we would put more trust in dates than in the morphological construction. Therefore there is no reason to do it for conversion.

To conclude, even if dates of first attestation give a direction to the conversion in most cases, they raise significant problems. They are not reliable: neither theoretically, because they depend on chance; nor empirically because we cannot assess what would be a reliable interval between two dates and because they sometimes contradict morphology.

3.2. Etymology

According to Balteiro (2007), the etymology provided by dictionaries is the best criterion to assess the directionality of conversion. When applied to the sample under study, the reference dictionary of etymology for French gives a direction to 387 pairs, *i.e.* 80% of the data. According to etymology, 278 pairs are noun→verb conversions and 109 are verb→noun. Unlike dates of attestation, etymology is always consistent with the morphological analysis. For each pair, the *DHLF* gives the etymology of the lexeme that is supposedly the base of conversion, and indicates this base as the etymology of the supposedly

converted lexeme, as shown in Tables 2 and 3. It seems that, at least for the supposed derived lexeme, the difference between a morphological and an etymological analysis is blurred, so that, from a theoretical and methodological point of view, we can question the boundary between the two domains.

Tab. 2: Noun→verb conversions according to etymology.

Noun	Etymology of noun	Verb	Etymology of verb
CENTRE 'center'	Latin *centrum*	CENTRER 'to center'	derived from *centre*
COTON 'cotton'	Arabic *qutun*	COTONNER 'to cover with cotton'	derived from *coton*
ENTHOUSIASME 'enthusiasm'	Greek *enthousiasmos*	ENTHOUSIASMER 'to fill with enthusiasm'	derived from *enthousiasme*
GAZON 'turf'	Francique °*waso*	GAZONNER 'to grass over'	derived from *gazon*
TAG 'tag'	English borrowing	TAGUER 'to tag'	derived from *tag*

Tab. 3: Verb→noun conversions according to etymology.

Verb	Etymology of verb	Noun	Etymology of noun
AMBLER 'to amble'	Old Provençal *amblar*	AMBLE 'amble'	deverbal of *ambler*
ANNONCER 'to announce'	Latin *adnunciare*	ANNONCE 'announcement'	deverbal of *annoncer*
EMBARRASSER 'to embarrass'	Spanish *embarazar*	EMBARRAS 'embarrassment'	deverbal of *embarrasser*
FLIPPER 'to freak out'	from English *to flip*	FLIP 'anguish'	from *flipper*
LAYER 'to cut a path'	Francique *lakan*	LAIE 'path'	derived from *layer*

In her study, Balteiro (2007) only took into account data that are similar to situations illustrated in Tables 2 and 3. That explains why she considers etymology to be the most reliable criterion. However, as already noticed by Marchand (1963), besides these situations there are cases where etymologies do not allow to determine the direction of the conversion. These cases represent 96 pairs, that is 20% of the sample dataset. Unlike cases illustrated in Tables 2 and 3, for these pairs the *DHLF* provides an etymology for both lexemes, so that none seems to derive from the other. Examples of such cases are given in Table 4.

Tab. 4: Unknown directionality according to etymology.

Noun	Etymology of noun	Verb	Etymology of verb
ARGUMENT 'argument'	Latin *argumentum*	ARGUMENTER 'to argue'	Latin *argumentari*
BAN 'exile'	Francique *ban*	BANNIR 'to banish'	Francique ₒ*bannjan*
COACH 'coach'	English word	COACHER 'to coach'	from English *to coach*
FORGE 'forge'	Latin *fabrica*	FORGER 'to forge'	Latin *fabricare*
SOLDE 'balance'	Borrowed from It. *saldo*	SOLDER 'to settle'	Borrowed from It. *saldare*

This kind of situation, where both the noun and the verb were inherited from Latin, Greek or old French, or borrowed from another language, seems to be rather frequent according to the information collected on the sample (20% of the data). Yet, even if lexemes are inherited or borrowed by pairs, there can still be a morphological relation between them that is perceptible for speakers. For example, ARGUMENTER 'to argue' and ARGUMENTATION 'arguing, reasoning' have both been inherited from Latin (the verb comes from the latin verb *argumentari* and the noun from the latin noun *argumentatio)*, but they display in contemporary French the same kind of morphological relation as between NATIONALISER 'to nationalise' and NATIONALISATION 'nationalisation' that were both coined in French: a relation between a verb and its deverbal action noun suffixed with *–ation*. This shows that having an etymology does not mean that a morphological analysis is not possible. More generally, the example of ARGUMENTER and ARGUMENTATION reveals that etymology and morphology do not have the same goals: etymology studies the history, the genealogy of words, while morphology analyses the morphological relations between lexemes in a given state of a language. Since they do not have the same purpose, etymology may not be a good tool for a morphological analysis, including the case of conversion.

3.3. Conclusion on historical criteria

As it has been argued, even when it can provide a direction, historical information is not reliable to decide on the directionality of conversion. On the one hand, dates of first attestation are not always accurate enough and often contradict the morphological analysis of lexemes. On the other hand, etymology is of no help when both lexemes were borrowed or inherited together.

Another more important, yet never discussed, problem arises when these two historical criteria are compared. Indeed, the analyses provided by dates and etymology are sometimes contradictory. For example, from a chronological point of view, the pair AIDE 'help'~AIDER 'to help' can be considered to be a noun→verb conversion because the noun is attested before the verb, as can be seen in Table 5. However, according to the etymologies provided in Table 6, it is a verb→noun conversion because the verb comes from the Latin *adjutare* while the noun comes from the verb. The problem is similar with RIME 'rhyme'~RIMER 'to rhyme': according to the attestation dates it is a verb→noun conversion because the verb is attested first, but it is a noun→verb conversion if we rely on etymology because the noun comes from the Latin *glosa* while the verb derives from the noun. Table 7 provides examples of such inconsistencies between the two types of analysis.

Tab. 5: Dates of first attestation of certain lexemes.

Noun	1st attestation of noun	Verb	1st attestation of verb
AIDE 'help'	842	AIDER 'to help'	10th century
CHARROI 'convoy, carting'	1150	CHARROYER 'to carry along'	1225
GLOSE 'commentary'	1175	GLOSER 'to annotate'	1130
LOUANGE 'praise'	1160	LOUANGER 'to commend'	1155
REGARD 'look'	980	REGARDER 'to look'	1080
RIME 'rhyme'	1160	RIMER 'to rhyme'	1119

Tab. 6: Etymology of lexemes in Table 5.

Noun	Etymology of noun	Verb	Etymology of verb
AIDE	derived from *aider*	AIDER	Latin *adjutare*
CHARROI	derived from *charroyer*	CHARROYER	derived from *char*
GLOSE	Latin *glosa*	GLOSER	derived from *glose*
LOUANGE	derived from *louer*	LOUANGER	derived from *louange*
REGARD	deverbal from *regarder*	REGARDER	from *garder* with prefix *re-*
RIME	lat. *rhythmus*	RIMER	derived from *rime*

Tab. 7: Contradictory analyses according to dates and etymology.

Noun	Verb	Direction according to dates	Direction according to etymology
AIDE	AIDER	noun→verb	verb→noun
CHARROI	CHARROYER	noun→verb	verb→noun
GLOSE	GLOSER	verb→noun	noun→verb
LOUANGE	LOUANGER	verb→noun	noun→verb
REGARD	REGARDER	noun→verb	verb→noun
RIME	RIMER	verb→noun	noun→verb

These examples of inconsistency between dates and etymology raise a tricky issue: if we want to use historical information, we will have to make a choice between the two criteria in cases of contradictory analyses. Yet, there seems to be no obvious reasons to favour one criterion over the other.

4. Synchronic criteria

Marchand (1963, 1964) is the first to propose systematic synchronic criteria in order to determine the direction of conversion in English. These criteria are the following: comparison with affixation, semantic dependency of one lexeme on the other, frequency and semantic range of lexemes, semantic patterns between lexemes, phonetic shape, morphological types of lexemes and stress patterns. These criteria have been discussed by many studies on conversion, and some of them have been ruled out as not being operative. See for instance Sanders (1988) for a criticism of the comparison with affixation, Ljung (1977) for the rejection of the semantic dependency and the semantic range. Stress patterns have been described as a reliable criterion by Kiparsky (1997) in English and by Rodrigues Soares (2009) in Portuguese, but it is of no help in French because there is no word stress difference between nouns and verbs. Semantic patterns between lexemes is the most widely used criterion and will be discussed in the following section. Other criteria have been proposed in diverse studies on different languages. For instance Don (2004) has claimed that noun gender and verb inflection can determine the directionality in Dutch. Similarly, Rodrigues Soares (2009) argues that certain thematic vowels on verbs indicate the direction of conversion in Portuguese. Building on these studies, the following subsections will evaluate these criteria on French.

4.1. A reference database of 626 directional noun~verb pairs

In order to be sure that these criteria can help to determine the direction of conversion, they must be evaluated on conversion pairs the directionality of

which is certain. To obtain these pairs, a morphological criterion was first applied, following Rodrigues Soares (2009): if one lexeme already derives from another by other means than conversion, then it is the base of conversion. For example, for the pair PARLEMENT 'parliament'~PARLEMENTER 'to negotiate', the morphological analysis gives the result in (4a): since the noun PARLEMENT already derives from the verb PARLER 'to talk' by means of the suffix *–ment* and with the meaning 'group of persons who talk' (just like GOUVERNEMENT 'government' is the 'group of persons who govern'), then it is the base of the conversion and the verb is derived. In the pair RAPPEL 'recall'~RAPPELER 'to call back', the morphological analysis gives the result in (4b): the verb RAPPELER already derives from the verb APPELER 'to call' by means of the prefix *r(e)–*, which forms verbs out of verbs, so that RAPPELER is the base and RAPPEL is the converted lexeme.

(4)   a.   PARLER 'to talk' → PARLEMENT 'parliament' (lit. 'group of persons who talk') → PARLEMENTER 'to negotiate' (lit. 'to act like the parliament')
      b.   APPELER 'to call' → RAPPELER 'to call back' → RAPPEL 'recall'

This morphological analysis has been applied to each lexeme of the corpus. According to it, the directionality of conversion can be decided in 626 pairs: 460 are noun→verb conversions and 166 are verb→noun. Examples are given in Table 8.

Tab. 8: Examples of directional pairs.

N→V conversion	BARRICADE 'barricade' > BARRICADER 'to barricade'
	GRILLAGE 'fence' > GRILLAGER 'to fence'
	RATURE 'crossing-out' > RATURER 'cross out'
	RÈGLEMENT 'rules' > RÈGLEMENTER 'to regulate'
V→N conversion	DÉCHARGER 'to unload' > DÉCHARGE 'dump'
	DÉGELER 'to thaw' > DÉGEL 'thaw'
	ÉPURER 'to refine' > ÉPURE 'sketch'
	RÉEXAMINER 'to reexamine' > RÉEXAMEN 'reexamination'

In the remainder of this section, three synchronic criteria will be evaluated on the basis of these 626 pairs: semantic patterns in 4.2., noun gender in 4.3. and verb inflection in 4.4.

4.2. Semantic patterns

Marchand (1964) suggests that semantic patterns between nouns and verbs, such as "act of V-ing", "to use N" etc., can determine the direction of conversion. This synchronic criterion is most often used by linguists when they decide

on the direction of conversion. In French for instance, it is used by Corbin (1987) and Kerleroux (1996). Corbin (1987) analyses the pair VOL 'flight'~VOLER 'to fly' as verb→ noun conversion because VOL can be defined as "action de voler" 'act of flying', which instantiates the "act of V-ing" pattern. Although this criterion is often used, its practical application has never been measured. It can be done by listing all patterns in both directions and comparing them.

Thus, in order to evaluate the usefulness of this criterion, the semantic relation between noun and verb has been analysed for the 460 noun→verb and the 166 verb→noun pairs of the directional database. The classification of the semantic patterns for converted verbs builds on that provided by Plag (1999), to which 2 patterns have been added: a causative one and a pattern for verbs of feeling. Overall, ten different semantic patterns have been observed. They are provided in Table 9 with examples. As for converted nouns, the list of semantic patterns comes from Plénat (2005). Six different patterns have been observed, as indicated in Table 10. Both classifications have been applied to all directional pairs with the help of the definitions found in the two dictionaries: *TLFi* and *Petit Robert Électronique*.

Tab. 9: Semantic patterns for noun→verb conversions.

Pattern	Name	Example
put in/into N	locative	REMISER = to put into the REMISE ('shed')
put N in/on	ornative	CARTONNER = to put CARTON ('cardboard') on something
remove N	privative	ŒILLETONNER = to remove the ŒILLETON ('buds')
do/perform N	performative	AUBADER = to do a AUBADE ('dawn serenade')
produce N	resultative	RATURER = to produce a RATURE ('crossing-out')
act/be like N	similative	PARLEMENTER = to act like the PARLEMENT ('parliament')
use N	instrumental	TÉLÉPHONER = to use the TÉLÉPHONE ('phone')
be N	stative	PRÉLUDER = to be a PRÉLUDE ('prelude')
cause N	causative	CONFUSIONNER = to cause CONFUSION ('confusion')
feel N	feeling	COMPASSIONNER = to feel COMPASSION ('compassion')

Tab. 10: Semantic patterns for verb→noun conversions.

Pattern	Name	Example
act of V-ing	action	RAPPEL = act of RAPPELER ('call back')
result of V-ing	result	AMAS = result of AMASSER ('amass')

the one that V-s	agent	MARMOTTE = animal that MARMOTTER ('mutter')
what is V-ed	patient	DÉBOURS = what is DÉBOURSER ('spend')
object to V	instrument	RÉVEIL = object to RÉVEILLER ('wake up')
place where one V-s	location	DÉCHARGE = place where one DÉCHARGER ('unload')

As can be seen by comparing the semantic patterns observed for the two types of conversions, many patterns in one direction have a counterpart in the other direction. Indeed, except for 'what is V-ed', every verb→noun pattern has a reverse noun→verb pattern, as is summarised in (5).

(5) a. do/perform N ↔ act of V-ing
    b. produce N ↔ result of V-ing
    c. act like N ↔ the one that V-s
    d. use N ↔ object to V
    e. put in(to) N ↔ place where one V-s

Because of these reverse semantic patterns, all conversion pairs that involve one of the patterns given in (5) can almost always be analysed in both directions, as shown in Table 11. Indeed, from a semantic point of view, each pair in the table can either be analysed as noun→verb conversion or as verb→noun conversion. For each one, the pattern noted in bold fonts is the one whose direction is made certain by the morphological analysis.

Tab. 11: Reverse semantic patterns.

N~V pair	N→V pattern	V→N pattern
AUBADE~AUBADER	**do a AUBADE 'serenade'**	act of AUBADER 'serenade'
RAPPEL~RAPPELER	do a RAPPEL 'recall'	**act of RAPPELER 'call back'**
RATURE~ RATURER	**produce a RATURE 'crossing-out'**	result of RATURER 'cross out'
AMAS~AMASSER	produce a AMAS 'heap'	**result of AMASSER 'amass'**
PARLEMENT~ PARLEMENTER	**act like a PARLEMENT 'parliament'**	the one that PARLEMENTER 'negotiate'
MARMOTTE~ MARMOTTER	act like a MARMOTTE 'marmot'	**the one that MARMOTTER 'mutter'**
TÉLÉPHONE~ TÉLÉPHONER	**use a TÉLÉPHONE 'phone'**	object to TÉLÉPHONER 'call'
RÉVEIL~RÉVEILLER	use a RÉVEIL 'alarm clock'	**object to RÉVEILLER 'wake up'**
REMISE~REMISER	**put into a REMISE 'shed'**	place where to REMISER 'put away'
DÉCHARGE~ DÉCHARGER	put into a DÉCHARGE 'dump'	**place where to DÉCHARGER 'unload'**

This study of the semantic patterns carried out on 626 conversion pairs has revealed that in most cases the semantic relation between the noun and the verb does not allow to decide on the directionality of conversion because of reverse semantic patterns. A few patterns seem to be reliable, though, because they have no counterparts. For instance, within verb→noun patterns, only the patient one (see DÉBOURS 'disbursement' in Table 10) seems to have no counterpart in the other direction. However, this pattern is very uncommon: it was observed in only 8 pairs only out of 166, that is, less than 5% of the data. Moreover, it could also have the ornative or instrumental patterns as counterparts. As regards noun→verb conversion, the ornative, privative, causative, stative and feeling patterns could be reliable indications of the directionality because they seem to have no reverse pattern. However, the privative pattern was observed only once and therefore is not very helpful with respect to directionality. Causative, stative and feeling cases could probably be merged with the performative pattern. Moreover, they are the least common patterns in the subset (23 pairs out of 460, *i.e.* 5% of the data) together with the privative one. The only pattern that could be a reliable clue for the directionality is the ornative one that has been observed in 58 pairs, that is 12.6% of the data. However, ornative verbs are often merged with instrumental verbs because they imply the use of the object denoted by the base noun. For example, TO SALT can be analysed as an ornative verb with the pattern 'put salt in/on'. But it can also be analysed as an instrumental verb that instanciates the pattern 'use the salt'. This is the solution that Aronoff (1980) recommends.

The analysis and comparison of all semantic patterns observed in both conversions has shown that, except for the privative pattern, which can only be found in noun→verb conversion (but is very rare), semantic patterns do not enable the identification of the conversion directionality because they all have counterparts in the opposite direction.

4.3. Noun gender

Don (2004) claimed that noun gender is a good indication of the direction of conversion in Dutch. There are two genders in Dutch: neuter and non-neuter. According to Don, verb→noun conversion, like all nominalization processes in Dutch, can only form non-neuter nouns so that when the noun is neuter, it must be the base of the conversion and the verb is derived. French also has two genders: feminine and masculine. Both of them can be found on nominalizations, as well as on nouns that are used for other derivations. When looking at the database of directional noun-verb pairs described in section 4.1., we can see that both genders are evenly distributed between the two conversions, as shown in Table 12.

Tab. 12: Noun gender and conversion.

Gender	Noun→Verb		Verb→Noun	
	#	%	#	%
feminine	195	42.4	74	44.6
fem. or mas.	1	0.2	1	0.6
masculine	264	57.4	91	54.8

Contrary to what Don has argued on Dutch, noun gender in French proves not to be associated with one direction over the other. Therefore, it cannot be used as a criterion to determine the directionality of conversion.

4.4. Verb inflection

Kiparsky (1997) relied on verb inflection to decide on the conversion type in English. According to him, an irregular verb cannot be derived from the noun and must derive from a root, together with the noun. Don (2004) made a similar statement in Dutch: noun→verb conversion can only form regular verbs, so that all pairs with irregular verb must be verb→noun conversions. As for Portuguese, Rodrigues Soares (2009) claimed that converted verbs can only bear the thematic vowel /a/. Therefore, when the verb displays the thematic vowel /i/ or /e/ it must be a verb→noun conversion.

  In French, verbs fall into three classes, named *groups*. The first group is the most important one. It includes verbs that are all regular, end in –*er*, have a past participle in –*é* and a simple past in –*a*. The *Petit Robert Électronique* dictionary includes almost 6,000 first group verbs. The second group is composed of verbs, usually regarded as irregular, that end in –*ir* and have a present participle in –*issant*. They are about 310 in the *Petit Robert Électronique*. Finally, the third group comprises all other irregular verbs. They are 374 in the *Petit Robert Électronique*. The distribution of verbs among the three groups in the directional database (see section 4.1.) is given in Table 13.

Tab. 13: Verb inflection and conversion.

Group	Noun→Verb		Verb→Noun	
	#	%	#	%
1	460	100	159	95.8
2	0	0	3	1.8
3	0	0	4	2.4

The results in Table 13 seem to correlate with those observed in Dutch and Portuguese, that is, that irregular verbs (2nd and 3rd groups) are only found in verb→noun conversion. However, these results come from the directional

database where all nouns and verbs are morphologically complex, as they are derived from other lexemes. So, the numbers in Table 13 only indicate that morphologically complex nouns cannot be converted into verbs of the second and third groups. But they say nothing about the possibility for non-complex nouns to be converted into second or third group verbs. For example, some people would analyse examples in (6) (2nd group verbs) and (7) (3rd group verbs) as noun→verb conversion.

(6) a. GAUCHE 'left' → GAUCHIR 'to reorientate the politics to the left'
     b. NORD 'north' → NORDIR 'to turn to the north' (speaking about the wind)

(7) a. DISCOURS 'speech' → DISCOURIR 'to give a speech'
     b. SECOURS 'help' → SECOURIR 'to rescue' (lit. 'to bring help')

From a semantic point of view, at least for examples in (6), it would not seem illogical to consider the nouns as the bases and the verbs as converted, because it would be very odd to define the noun GAUCHE as 'the direction towards which one reorientates a politics' or NORD as 'the direction where the wind blows'. Examples in (7) are less convincing because they can also be analysed in the opposite direction.

   To conclude, the situation in French with respect to verb inflection is less clear than in English and Portuguese. Second and third group verbs could be a hint of the direction of conversion but it is not fully reliable. Moreover, even if it were fully reliable, it only applies to very few pairs (4.2% of the directional data), so that it is not very helpful to determine the directionality of conversion.

5. Theoretical implications

The previous two sections have demonstrated that most of the time the directionality of noun~verb conversion in French cannot be identified. On the one hand, historical criteria can sometimes indicate a direction, but they are not reliable. On the other hand, synchronic criteria cannot provide a direction because each of them is compatible with both directions. These results raise problems for derivational analyses of conversion. Non-derivational analyses relying on category underspecification have been shown to be problematic in section 2. Thus, a non-derivational analysis with fully specified categories such as the one proposed by Lieber (1981, 2004) could be an interesting solution. Lieber argues that nouns and verbs do not derive from one another and are rather linked by relisting rules in the lexicon. That is, they are fully specified for categories, they are separately listed in the lexicon, and redundancy rules link them. According to the author, conversion is thus non-directional. This analysis seems to solve the directionality

problem. However, Lieber argues that conversion is directional on the semantic level: "whereas neither member of a conversion pair is structurally more basic, one member of a pair will always be semantically more basic and the other semantically derived." (Lieber 1981: 185). Thus, she adds directional semantic rules to link nouns and verbs, so eventually conversion is directional, at least on the semantic level. Yet, this solution is not satisfactory because, as section 4.2. has shown, semantic relations between nouns and verbs are almost always ambiguous between both directions in French.

In fact, this directionality problem is not specific to conversion. Corbin (1976) already noticed this problem with *–ie* and *–ique* suffixes, as in SYMÉTRIE 'symmetry' and SYMÉTRIQUE 'symmetrical'. Indeed, SYMÉTRIE is the property of being symmetrical, and SYMÉTRIQUE means 'that displays symmetry'. She noted that this kind of data is a problem for the derivational analysis she supports. Roché highlighted the same problem in various studies on different suffixes: *–ier* and *–erie* as in MERCIER 'haberdasher' and MERCERIE 'haberdashery' because MERCIER is the person who works in a haberdashery and MERCERIE is the activity of a haberdasher (Roché 2004); *–isme* and *–iste* as in FASCISME 'fascism' and FASCISTE 'fascist' because fascists are followers of fascism and fascism is the ideology of fascists (Roché 2007); or in country names-demonyms pairs, such as HONGRIE 'Hungary' and HONGROIS 'hungarian' where HONGRIE is the country of hungarians and HONGROIS are inhabitants of Hungary (Roché 2008).

In order to account for such cases, Roché talks about mutual motivation between lexemes. Umbreit (2011) also highlights mutual motivations between lexemes and extends the notion to the whole derivational family, even when one lexeme clearly derives from another, such as FISHY and FISH where the adjective derives from the noun by means of the suffix *–y*. According to Umbreit, morphological families form motivational networks where motivation between members of one family can not only be bidirectional, but also multidirectional. In a quite similar fashion, word or lexeme-based approaches to morphology have recently extended the notion of paradigms to derivation. Such paradigmatic analyses of derivation allow to describe morphological families and compare their organization across the lexicon (see Hathout and Namer 2019 for an overview). As mentioned by Štekauer (2014) there are different definitions of derivational paradigms. Bonami and Strnadová (2019), for example, define derivational paradigms as sets of aligned families sharing the same organization. In such paradigms, derivational relations are regarded as multidirectional relations between members of a family. Hathout and Namer (2019) noticed that one advantage of a paradigmatic approach to derivation is that it enables analyses of phenomena that are not easily described by traditional directional rules. Conversion, with its directionality problems, is undoubtedly one of these phenomena.

6. Conclusion

This study on French noun~verb conversions has presented and discussed the main criteria mentioned in the literature when dealing with the directionality of conversion between nouns and verbs. Each criterion has been evaluated on different subsets of a corpus containing 3,241 noun~verb pairs.

From an empirical point of view, it has been demonstrated that none of these criteria is reliable enough to determine the direction of conversion for all noun~verb pairs: dates of first attestation are not always accurate enough and often conflict with morphology, etymology does not always allow to identify a direction and both historical criteria often lead to contradictory analyses. When it comes to synchronic criteria, morphological complexity is one reliable test, but it only helpfully applies to 626 pairs over 3,241, that is to 19.3% of the data. Semantic patterns often enable reverse analyses, but some of them such as the privative (see Table 9) and the patient ones (see Table 10) correlate with only one direction. However, these patterns apply to very few cases. In individual cases, the semantic relation between the noun and the verb could also indicate a direction, as in the examples in (6), but it cannot be generalised to all pairs. Noun gender is not helpful because both genders are found in both conversions, and verb inflection seems not to be reliable. To conclude, all these criteria might help to decide on a reliable direction for few individual cases, but none of them is applicable to all noun~verb pairs. Therefore, in most cases the directionality of conversion in French seems not to be determinable. This non-directionality is not specific to conversion and can also be found with a number of suffixes. Whereas these problematic derivations challenge the traditional conception of derivation rules, paradigmatic morphology seems to offer a good framework to account for such phenomena.

References

Adouani, Abdellatif. 1989. *Les substantifs déverbaux à "suffixe zéro" en français moderne*. PhD thesis, Université Lyon 2.

Aronoff, Mark. 1980. Contextuals. In Teun Hoekstra, Harry Van der Hulst & Michael Moortgat (eds.), *Lexical Grammar*, 263–285. Dodrecht: Foris Publications.

Aronoff, Mark. 2007. In the begining was the word. *Language* 83(4). 803–830.

Bally, Charles. 1944. *Linguistique générale et linguistique française*. Berne: A. Francke.

Balteiro, Isabel. 2007. *The Directionality of Conversion in English. A Dia-Synchronic Study*. Berlin: Peter Lang.

Barner, David & Alan Bale. 2002. No nouns, no verbs: psycholinguistic arguments in favor of underspecification. *Lingua* 112. 771–791.

Bonami, Olivier & Jana Strnadová. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29. 167–197.

Borer, Hagit. 2013. *Taking Form. Structuring Sense*. Volume III. Oxford: Oxford University Press.

Corbin, Danielle. 1976. Peut-on faire l'hypothèse d'une dérivation en morphologie?. In Jean-Marie Chevalier (ed.), *Grammaire transformationnelle: syntaxe et lexique*, 47–91. Lille: Publications de l'Université de Lille 3.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Niemeyer.

Croft, William. 1991. *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago: University Press of Chicago.

Dirven, René. 1988. A Cognitive Approach to Conversion. In Werner Hüllen & Rainer Schulze (eds.), *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics*, 329–343. Tübingen: Niemeyer.

Dirven, René. 1999. Conversion as a Conceptual Metonymy of Event Schemata. In Klaus-Uwe Panther & Günter Radden (eds.), *Metonymy in Language and Thought*, 275–287. Amsterdam & Philadelphia: Benjamins.

Don, Jan. 2004. Categories in the lexicon. *Linguistics* 42(5). 931–956.

Don, Jan. 2005. On Conversion, Relisting and Zero-derivation. *SKASE Journal of theoretical Linguistics* 2(2). 2–16.

Dubois, Jean. 1962. *Étude sur la dérivation suffixale en français moderne et contemporain*. Paris: Larousse.

Farrell, Patrick. 2001. Functional shift as category underspecification. *English Language and Linguistics* 5(1). 109–130.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie*. Paris: PUF.

Goldberg, Adele. 1995. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Hale, Kenneth & Samuel J. Keyser. 1993. On Argument Structure and the Lexical Expression of Syntactic Relations. In Kenneth Hale & Samuel J. Keyser (eds.), *The view from Building 20: essays in honor of Sylvain Bromberger*, 53–109. Cambridge: The MIT Press.

Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the Pieces of Inflection. In Kenneth Hale & Samuel J. Keyser (eds.), *The view from Building 20: essays in honor of Sylvain Bromberger*, 111–176. Cambridge: The MIT Press.

Harley, Heidi. 2005. How do verbs get their names? Denominal verbs, manner incorporation and the ontology of verb roots in English. In Tova Rapoport & Nomi Erteschik-Shir (eds.), *The Syntax of Aspect*, 42–64. Oxford: Oxford University Press.

Hathout, Nabil & Fiammetta Namer. 2019. Paradigms in word formation: what are we up to? *Morphology* 29. 153–165.

Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles*. Part IV: Morphology. London: Allen & Unwin.

Kerleroux, Françoise. 1996. *La coupure invisible*. Lille: Presses Universitaires du Septentrion.

Kerleroux, Françoise. 1999. Identification d'un procédé morphologique: la conversion. *Faits de langues* 14. 89–100.

Kiparsky, Paul. 1997. Remarks on denominal verbs. In Alex Alsina, Joan Bresnan & Peter Sells (eds.), *Argument Structure*, 473–499. Stanford: CSLI Publications.

Labelle, Marie. 1992. La structure argumentale des verbes locatifs à base nominale. *Linguisticae Investigationes* XVI(2). 267–315.

Labelle, Marie. 2000. The semantic representation of denominal verbs. In Peter Coopmans, Martin Everaert & Jane Grimshaw (eds.), *Lexical specification and insertion*, 215–240. Amsterdam: Benjamins.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. Vol. 1: *Theoretical prerequisites*. Stanford: Stanford University Press.

Langacker, Ronald W. 1991. *Foundations of cognitive grammar*. Vol. 2: *Descriptive application*. Stanford: Stanford University Press.

Lieber, Rochelle. 1981. Morphological Conversion Within a Restrictive Theory of the Lexicon. In Michael Moortgat, Harry van der Hulst & Teun Hoekstra (eds.), *The Scope of Lexical Rules*, 161–200. Dordrecht: Foris Publications.

Lieber, Rochelle. 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.

Ljung, Magnus. 1977. Problems in the Derivation of Instrumental Verbs. In Herbert Brekle & Dieter Kastovsky (eds.), *Perspektiven der Wortbildungsforschung*, 165–179. Bonn: Bouvier Verlag.

Lohmann, Arne. 2017. Phonological properties of word classes and directionality in conversion. *Word Structure* 10(2). 204–234.

Marantz, Alec. 1997. No escape from Syntax: Don't try morphological analysis in the privacy of your own lexicon. *UPenn Working Papers in Linguistics* 4.2. 201–225.

Marchand, Hans. 1963. On a question of contrary analysis with derivationally connected but morphologically uncharacterized words. *English Studies* 44. 176–187.

Marchand, Hans. 1964. A set of criteria for the establishing of derivational relationship between words unmarked by derivational morphemes. *Indogermanische Forschungen* 69. 10–19.

Nyrop, Kristoffer. 1936. *Grammaire historique de la langue française. Tome III: La formation des mots*. Paris: Aphonse Picard & Fils.

Plénat, Marc. 2005. Brèves remarques sur les déverbaux en *-ette*. In Frederic Lambert & Henning Nølke (eds.), *La syntaxe au cœur de la grammaire. Recueil offert en hommage pour le 60e anniversaire de Claude Muller*, 245–258. Rennes: Presses Universitaires de Rennes.

Plag, Ingo. 1999. *Morphological productivity. Structural constraints in English derivation*. Berlin & New York: de Gruyter.

Roché, Michel. 2004. Mot construit? Mot non construit? Quelques réflexions à partir des dérivés en *-ier(e)*. *Verbum* XXVI(4). 459–480.

Roché, Michel. 2007. Logique lexicale et morphologie: la dérivation en *-isme*. *Proceedings of the 5th Décembrettes*. 45–58.

Roché, Michel. 2008. Structuration du lexique et principe d'économie: le cas des ethniques. *Actes du CMLF 2008*. 1571–1585.

Rodrigues Soares, Alexandra. 2009. Portuguese converted deverbal nouns: Constraints on their bases. *Word Structure* 2 (1). 69–107.

Sanders, Gerald. 1988. Zero derivation and the overt analogue criterion. In Michael Hammond & Michael Noonan (eds.), *Theoretical morphology*, 155–175. San Diego: Academic Press.

Štekauer, Pavol 2014. Derivational paradigms. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of derivational morphology*, 354–369. Oxford: Oxford University Press.

Tournier, Jean. 1980. La conversion: Problèmes théoriques et implications sémantiques. *Recherches en Linguistique Etrangère V*. 71–90.

Umbreit, Birgitt. 2011. Motivational networks. An empirically supported cognitive phenomenon. In Klaus-Uwe Panther & Güner Radden (eds.), *From Motivation in Grammar and the Lexicon*, 269–286. Amsterdam & Philadelphia: Benjamins.

Delphine Tribout
Université de Lille
Département Sciences du Langage
Rue du Barreau
BP 60149
59653 Villeneuve d'Ascq Cedex
France
delphine.tribout@univ-lille.fr

Varia / Miscellaneous

Tagungsankündigung / Conference Announcement

"Word-Formation Theories IV / Typology and Universals in Word-Formation V" (June 23–26, 2021; Kosice, Slovakia)

Contact Person: Pavol Stekauer
Meeting Email: pavel.stekauer@upjs.sk

Linguistic Field(s): Linguistic Theories; Morphology; Typology

Meeting Description:
Another edition of our twin conference concentrates on two areas of research: word-formation theories and word-formation typology/universals. Papers discussing any aspect of these general areas are most welcome.

Call for Papers:
Abstract of max. 300 words should be emailed to livia.kortvelyessyupjs.sk

Deadlines:
Submission of abstracts: 15 April, 2021
Notification of acceptance: 30 April, 2021
Submission of a registration form: 31 May, 2021

Plenary speakers (in alphabetical order):
Mark Aronoff, Stony Brook University, New York, USA
Jan Don, Amsterdam University, The Netherlands
Martin Everaert, Utrecht University, The Netherlands
Volker Gast, Friedrich Schiller University, Jena
Gary Libben, Brock University, USA
Brian MacWhinney, Carnegie Mellon University, USA