Light verb constructions from a cross-linguistic perspective

Light verb constructions from a cross-linguistic perspective

Edited by
Jens Fleischhauer and Anna Riccio

DE GRUYTER

We acknowledge support by the Open Access Publication Fund of the University and State Library Düsseldorf.

ISBN 978-3-11-222932-3
e-ISBN (PDF) 978-3-11-222933-0
e-ISBN (EPUB) 978-3-11-222934-7
DOI https://doi.org/10.1515/9783112229330

Library of Congress Control Number: 2024947495

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.

Printing and binding: CPI books GmbH, Leck
Cover image: Jievani Weerasinghe / iStock / Getty Images Plus

www.degruyterbrill.com
Questions about General Product Safety Regulation:
productsafety@degruyterbrill.com

Contents

List of Abbreviations

ABS	absolutive
ACC	accusative
ADD	additive
ADJ	adjective
ADV	adverb
AFF	affirmative
AGT	agent
AOR	aorist
APL	animate plural
APLL	applicative
APUD	localization ‘near’
ASSOC	associative marker
ATTR	attributive
AVC	auxiliary verb construction
CAUS	causative
COMP	comparative particle
COND	conditional
CONT	localization ‘in contact’
COP	copula
CVB	converb
DAT	dative
DC	discontinuous element
DEF	definite
DEM	demonstrative
DEP	dependent mode
DOWN	direction ‘down’
ELAT	elative
ERG	ergative
EZ	ezafe
F	feminine
FOC	focus
FUT	future tense
FV	final vowel
GEN	genitive
GER	gerund
H	human
HON	honorific
HPL	human plural
HUM	humilific
IDEOPH	ideophone
IMP	imperative
IMPF	imperfective
IN	localization ‘inside’

https://doi.org/10.1515/9783112229330-203

INCH	inchoative
INDEF	indefinite
INF	infinitive
INT	interrogative
LAT	lative
LK	linker
LVC	light verb construction
M	masculine
MNR	manner adverb
NARR	narrative tense
NEG	negation
NPL	non-human plural
NPST	non-past tense
OBJ	object
OBL	oblique
OM	object marker
PART	particle
PASS	passive
PC	pronominal clitic
PER	periphrasis form
PERF	perfect
PFV	perfective
PL	plural
POSS	possession
PRF	perfect
PRS	present tense
PRX	proximal
PST	past tense
PTCP	participle
PV	preverb
Q	question
QUOT	quotative
R	Russian borrowing
RE	refactive
REFL	reflexive
REL	relative particle
REP	reportative
RMPST	remote past tense
SBJV	subjunctive
SG	singular
SUP	localization ‘on’
SVC	serial verb construction
UP	direction ‘up’
VN	verbal noun

Jens Fleischhauer and Anna Riccio

Introduction to the volume

1 A first approximation to the term ‘light verb construction’

William Croft (2012, 1) writes that “[a] central part of the grammar of every human language is the encoding of events and their participants in a clause.” That this observation is the first sentence of a book titled ‘Verbs’ is not surprising. Verbs denotate events – or, more broadly, eventualities[1] – and determine the number and type of participants. In sentences, verbs typically assume the function of the predicate. All languages have expressions that can be used as predicates.

There is, however, no 1-to-1 correspondence between the two categories ‘verb’ and ‘predicate.’ While ‘verb’ is a lexical category, ‘predicate’ refers more to the use of an expression. Which is also evidenced by the fact that nouns and adjectives can be used as predicates as well (*John is a teacher*). In some languages (e.g., Russian), this is possible in certain grammatical contexts (e.g., present tense) even without a verbal element (copula):

(1) *Ona chorošij inžener*.
3.SG.F good.M.SG.NOM engineer.M.SG.NOM
‘She is a good engineer.’

(Geist, 2006, 133)

In addition to morphologically simplex verbs (such as English *give* or *take* in (2)), many languages also have more complex expressions that fulfill the function of a predicate. Examples can be seen in (3). Here, the verbs *give* in (3a) and *take* in (3b), together with the NPs *a kiss* and *a walk*, form a complex predicate.

(2) a. *The girl gave the ball to the boy*.
b. *The girl took the ball from the boy*.

(3) a. *The girl gave the boy a kiss.*
b. *The girl took a walk in the park.*

1 The term ‘eventuality’ goes back to Bach (1986) and is a cover term for both states and events.

Jens Fleischhauer, Heinrich-Heine-University, Düsseldorf, Germany
Anna Riccio, University of Foggia, Foggia, Italy

https://doi.org/10.1515/9783112229330-001

In (2a) *the girl* holds the role of the giver, *the boy* is the recipient, and *the ball* is what is given (theme or patient role). The verb denotes an event in which a change of possession of the theme from the giver to the recipient occurs. This is different in (3a); here no change of possession occurs. *Kiss* denotes no entity that exists independently of the event and thus can be transferred from one individual to another. Rather, *give a kiss* denotes an event of kissing, with the girl as the one who kisses and the boy as the one who is kissed.

If we replace *kiss* in (3a) with another noun – for example *answer* (4) –, the resulting complex predicate denotes an event of answering instead of an event of kissing. This shows that, in the case of the complex predicates in (3) and (4), the denoted eventuality is determined by the noun and not by the (finite) verbs *give* or *take*. At the same time, we see that the verbs in (3) and (4) do not have the same meaning as in (2). In (2) they designate an eventuality of giving or taking; in (3) the eventuality depends on the noun with which they form a complex predicate. If, on the other hand, in (2a) we replace *the ball* with *the book*, the denoted event does not change – both cases refer to a transfer event. Rather, only what is given changes. This shows that in (2a) the verb, not the noun, determines the denoted eventuality.

(4) *The girl gave the boy an answer*.

The verbs in (3) and (4) carry the same morphosyntactic features as their counterparts in (2). That is, they realize the functional categories commonly associated with verbs in English (tense, subject agreement). Thus, the verbs in (3) and (4) play the same formal role – as bearers of grammatical information – as the verbs in (2). They can thus be seen as the morphosyntactic or grammatical head of the sentence. Semantically, the verbs in (3) and (4) are defective, since they themselves do not make an event predicate (cf. Butt and Geuder, 2001, 356). Following Jespersen (1942), verb usages like those in (3) and (4) are called 'light verbs.' According to Jespersen (1942, 117), light verbs are "insignificant verbs to which the marks of person and tense are attached, before the really important idea." This "really important idea" is what is contributed by the nominal element.

Complex predicates such as *give a kiss*, *give an answer*, and *take a walk* are referred to as 'light verb constructions,' 'stretched verb constructions' (Allerton, 2003), or also 'support verb constructions' (Langer, 2004; Mel'čuk, 2022). There are also other designations, and in particular different specific terms used in different language-specific traditions. For a brief overview, see Riccio's contribution to this volume (especially page 177). The reader is directed to Heine (2020) and Pompei et al. (2023) for a brief historical overview of debate on light verb constructions.

So far, there is no definition of the term 'light verb construction' accepted in the research literature, nor a comparative concept that would allow cross-linguistic

identification of LVCs. Nevertheless, the term 'light verb construction' has proven productive for linguistic description, as it is used to refer to a type of complex predicates in typologically diverse and genetically unrelated languages. For the Germanic languages, LVCs are for example assumed in German (e.g., Harm, 2021; Fleischhauer, 2022), Dutch (Everaert and Hollebrandse, 1995), English (e.g., Wierzbicka, 1982; Ronan and Schneider, 2015; Mehl, 2017; Stojanovska-Ilievska, 2025; Sundquist, 2025), Swedish (Sundquist, 2018), and Norwegian (Hellan, 2023). For the Romance languages, LVCs are discussed with respect to French, Italian, Spanish, and Portuguese (e.g., Alba-Salas, 2002; Staudinger, 2018; Pompei, 2023; Pompei and Piunno, 2023; Wiskandt, 2025). Irish is representative of the Celtic branch, which is said to have light verb constructions (Nolan, 2014). Another Celtic language that is said to exhibit LVCs is Scottish Gaelic (Esteban, this volume). Czech is an example of a Slavic language that is said to possess LVCs (Kettnerová, 2023), but they are also proposed for older Indo-European languages such as Ancient Greek (De Pasquale, 2023), Latin (Di Salvo, 2023; Pompei, 2023), Old Swedish (Sundquist, 2018) and Old West Norse (Boldt, this volume).

Finally, LVCs are assumed to be especially frequent among the Iranian (e.g., Kurmanji Kurdish, Shabaki, Persian, Lari; Haig, 2002; Sultan, 2011; Fleischhauer and Neisani, 2020; Fleischhauer, 2020; Ourang et al., this volume, among others) and Indo-Aryan languages (Hindi/ Urdu, Bengali; e.g., Butt, 1995; Butt and Lahiri, 2013; Mohanan, 2006).Vincze et al. (2013) reports the existence of LVCs in the Uralic language Hungarian. According to Mohanan (2006), LVCs are also found in the Dravidian language Malayalam. LVCs are also proposed the exist in the languages of the Caucasus, e.g., the Nakh-Daghestanian languages Rutul (Maisak, this volume) and Udi (Harris, 2008). Outside of Indo-European, LVCs are assumed to exist in Niger-Congo languages such as Swahili (Olejarnik, 2011), Turkic languages – for instance – Turkish, Uzbek, Kirghiz (Bowern, 2004), Semitic languages such as Standard and Egyptian Arabic (Ibrahim, 2005) but also in East Asian languages like Japanese and Korean (e.g., Grimshaw and Mester, 1988; Matsumoto, 1996; Lee, 2011; Kishimoto, 2025) as well as the Sino-Tibetan languages Mandarin Chinese (e.g., Kuo, 2025) and Tibetan Lai, this volume). For the Austronesian language family, Nugraha (2024) proposes the existence of LVCs in Bahasa Indonesian. Finally, some authors argue for the existence of LVCs in Australian languages such as Nyulnyulan (Bowern, 2008b), Jaminjung (Schultze-Berndt, 2006), Bardi and Wagiman (Bowern, 2010).[2]

The cited works do not necessarily share the same theoretical conception of 'light verb constructions,' so it is unclear whether they really identify the same

2 This is not intended as an exhaustive list, but rather aims to illustrate the application of the term in relation to different languages.

grammatical phenomenon in different languages or even within the same language. At this point, we do not want to undertake a comprehensive comparison of different conceptions of the term 'light verb construction,' but rather discuss LVCs in more detail from the perspective of one particular approach, namely the one proposed by Ježek (2016). We have chosen this approach, as Ježek presents her conception of LVCs very concisely, thereby providing a good starting point for a critical discussion. At the same time, her approach appears to be representative of other approaches to LVCs (at least for example Germanic and Romance languages). The guiding question in our discussion is to what extent the properties that LVCs are supposed to exhibit under this conception are empirically warranted. Based on our admittedly modest data discussion, we argue that a (cross-linguistically useful) definition of the term 'light verb construction' should make as few formal assumptions as possible and define LVCs primarily semantically – through the determination of event reference.

2 Ježek's conception of light verb constructions

In her volume on the lexicon, Ježek (2016, 204–205) proposes the following characterization of light verb constructions. According to her, LVCs …

– are a particular type of collocations,
– and consists of a "highly polysemous" light verb and an event noun "preceded by an article and/ or, in a few cases, by a preposition."

We start with the notion of a light verb. Contrary to Jespersen, many authors (Bonial and Pollard, 2020; Bowern, 2008a, 2014; Brugman, 2001; Butt, 2003, 2010; Butt and Lahiri, 2013; Fleischhauer and Neisani, 2020; Isoda, 1991; Ježek, 2016, 2023; Pompei and Piunno, 2023; Sanromán Vilas, 2011; Stojanovska-Ilievska, 2025; among others) now assume that light verbs are not semantically empty and do not merely serve to license a noun in predicative use. Rather, it is assumed that they contribute to the LVC's overall meaning. An illustrative example from German is shown in (5). The three sentences contain different LVCs that differ only in their light verb (*stehen* 'to stand' vs. *geraten* 'to get' vs. *stellen* 'to put' (lit. cause to stand)). While (5a) expresses a state predication (cf. Fleischhauer, 2021a), which can be paraphrased as 'be in the state of being observed by someone,' we have an inchoative predication – change from being not observed to being observed – in (5b) and a causative predication – cause to be observed by someone – in (5c). If light verbs were semantically empty, the complex predicates in (5) should not differ semantically.

(5) a. *Der Verdächtige stand unter Beobachtung durch die Polizei.*
the suspect stood under observation by the police
'The suspect is under observation by the police.'
b. *Der Verdächtige geriet unter Beobachtung durch die Polizei.*
the suspect got under observation by the police
'The suspect came under observation by the police.'
c. *Das Gericht stellte den Verdächtigen unter Beobachtung durch die Polizei.*
the court put the suspect under observation by the police
'The court placed the suspect under surveillance by the police.'

Minimal pairs of LVCs, in which only the light verb varies but the meaning of the respective complex predicates differs, can be found in numerous languages (e.g., Italian *avere paura* 'to be afraid' lit. have fear vs. *prendere paura* 'to get afraid' lit. take fear vs. *fare paura* 'to frighten' lit. make fear; cf. Pisciotta and Masini 2025, 138). The semantic function of light verbs includes specifying Aktionsart features (e.g., dynamicity, inchoation, causation), volitionality, or even honorificity (cf. Fleischhauer and Neisani, 2020; Lai, this volume). It remains unclear whether light verbs systematically contribute the same meaning across all LVCs or whether their semantic contribution varies from LVC to LVC. Following the use of various authors (e.g., Butt, 2010; Butt and Geuder, 2001; Bonial and Pollard, 2020; Fleischhauer and Neisani, 2020), we speak of a light-verb usage that contrasts with a heavy-verb usage of a lexical verb.[3] The underlying assumption is that light verbs and heavy verbs do not constitute formally distinct verb classes, but merely represent different uses of the same lexical verb (e.g., Butt and Geuder, 2001; Butt and Lahiri, 2013).

Ježek's claim that light verbs are polysemous thus reflects the fact that verbs like English *give* or German *stehen* 'to stand' have at least two usage variants: a heavy-verb usage and a light-verb usage. More precisely: Verbs like English *give* and German *stehen* 'to stand' display more than one heavy-verb usage (see, for example, Newman, 1996; Butt and Geuder, 2001, 338–343 on English *give*). Possibly the verbs also have multiple light-verb usages (see, for instance, Brugman, 2001 on the issue of polysemy of light verbs).

Ježek designates LVCs as collocations, i.e., as "word combinations on which a restriction applies, for which the choice of a particular word (the collocate) to express a given meaning is influenced by a second word (the base) to which the meaning applies" (Ježek, 2016, 200). In the case of LVCs, the noun imposes a restriction on the

3 To avoid unnecessarily long formulations, we will continue to speak of 'light verbs' and 'heavy verb', thereby referring to light- and heavy-verb usages of verbs.

choice of the verb. The nominal element of an LVC cannot be combined with every light verb.[4] Thus, for instance, *unter Beobachtung* 'under observation' is compatible with various light verbs (as in (5)), but not with all. Fleischhauer (2021a, 2022) shows that in German there is a series of LVCs of the type *stehen* + *unter* NP 'stand + under NP' which have a passive meaning (cf. *unter Beobachtung stehen* in (5a), which means 'to be observed'). Some of these LVCs take *stellen* 'put' (lit. cause to stand) as a caustive light verb (5c), while others – as in (6a) – take *setzen* 'put' (lit. cause to sit) – but not *stellen* – as a causative light verb variant (6b). This restriction seems to be conventionalized (Ježek, 2016, 204) and – at least to the extent that it concerns the distribution of the causative light verbs *stellen* and *setzen* – not to be motivated by the meaning of the two verbs.

(6) a. *Das Mädchen steht unter Stress.*
the girl stands under stress
'The girl is stressed.'
b. *Die Prüfung setzt/ *stellt das Mädchen unter Stress.*
the exam puts puts the girl under stress
'The exam puts the girl under stress.'

The second component of an LVC is, in addition to the light verb, a noun. Ježek points out that the noun can occur together with an article and/ or a preposition. A preposition appears, for instance, in the German examples in (5) and (6). Contrary to occasional claims in the research literature (e.g., Heringer, 1968, 41; Helbig, 1984, 165), the preposition is not a purely formal element, but contributes to the meaning of the LVC. This becomes evident when looking at minimal pairs in which the LVCs vary only in the choice of the preposition (7). While *unter Beobachtung stehen* 'be under observation' – as noted above – has a passive meaning, *vor der Beobachtung stehen* 'be close to be observed' has an aspectual – more precisely prospective – meaning (cf. Fleischhauer and Gamerschlag, 2019; Fleischhauer, 2023b).

(7) a. *Die Partei steht unter Beobachtung durch den Geheimdienst.*
the party stands under observation by the intelligence_service
'The party is under observation by the intelligence service.'

4 See also Wiskandt (2025) for a comparative study of the combination potential of psych nouns with different light verbs in Romance and Germanic languages. This issue is also taken up in several chapters of this volume. Esteban (this volume) reports it for Scottish Gaelis, Ourang et al. (this volume) for the Aheli dialect of Lari and Lai (this volume) for Tibetan.

b. *Die Partei steht vor der Beobachtung durch den*
the party stands in_front_of the observation by the
Geheimdienst.
intelligence_service
'The party is close to being observed by the intelligence service.'

The choice of the preposition affects the interpretation of the LVC and the choice of affects the denoted eventuality. Thus, it is not only the light verb but also the other components contribute to the meaning of the LVC. This also applies to the articles, as we will see below. In the English examples *give a kiss*, *give an answer* and *take a walk* an indefinite article is present. On the other hand, there are LVCs like German *unter Beobachtung stehen* 'be under observation' (5a) and *unter Stress stehen* 'be stressed' (6a) in which the noun occurs without any article. Especially in German linguistics, there is no consensus on whether the lack of articles is characteristic for German LVCs or whether there is (more or less) free article variation. A concise summary of the discussion about German can be found in Fleischhauer (2024). We find a similar debate but also concerning English LVCs. While some authors (e.g., Alego, 1995; Plante, 2014) assume that the noun occurring in LVCs appears only with an indefinite article, other authors attest article variation in LVCs (e.g., Bonial and Pollard, 2020; Stojanovska-Ilievska, 2025; see also the results of the contrastive corpus study (English, German, Catalan, and Spanish) in Alvarez-Morera, 2023 and Alvarez-Morera et al., 2025).

The examples in (8) show that *Beobachtung* 'observation' as a component of the LVC *unter Beobachtung stehen* 'be under observation' can also be used with an indefinite article (8a) and a definite article (8b). We can therefore conclude that the absence of an article – or the fixed use of, for instance, the definite article as some authors claim – cannot be characteristic of LVCs in German. A more detailed data-based argumentation for the fact that German LVCs exhibit free article variation can be found in Fleischhauer (2024).

(8) a. "*Ich stehe unter einer gewissen besonderen Beobachtung*", *verriet*
I stand under a certain special observation revealed
der 96-Torjäger nach seinem zehnten Treffer [...].
the 96-goal_scorer after his tenth hit
"'I am under a certain special kind of observation," revealed the 96-goal scorer after his tenth goal.'
(T06/FEB.04902 die tageszeitung, 27.02.2006)

b. *Der französische Klub steht im Rahmen des Financial Fairplay*
the French club stands in.the setting the.GEN financial fairplay
unter der Beobachtung der Uefa.
under the observation the.GEN Uefa
'The French club is under UEFA's observation as part of Financial Fair Play regulations.'
(SOL15/JAN.01914 Spiegel-Online, 21.01.2015)[5]

For a brief discussion of the issue of in-/definiteness in English LVCs, see Bruening (2015, 52). Riccio's contribution to this volume presents a very detailed study of the use of indefinite and definite articles in Italian LVCs.

The possibility to vary the articles points to two properties of LVCs that Ježek also notes. First, LVCs exhibit (a certain degree of) syntactic flexibility. The individual components do not necessarily have to be adjacent to each other. In (8a) we see that between the preposition *unter* 'under' and the noun *Beobachtung* 'observation' there can be not only an article but also two modifiers that modify the noun. Furthermore, the light verb can be separated from the PP by, for instance, adjuncts (8b).

Second, the nominal elements of the LVCs can be used referentially. This is, among other things, evidenced by the fact that they can be anaphorized.[6] Take (8b) as an example. The PP *unter der Beobachtung* (lit. under the observation) introduces an observation event, which can be picked up anaphorically in a following clause with a demonstrative. In (9) the demonstrative refers to the observation event and predicates how long this observation has already lasted.

(9) *Der französische Klub steht im Rahmen des Financial Fairplay*
the French club stands in.the setting the.GEN financial fairplay
unter der Beobachtung der Uefa. Diese dauert nun schon drei
under the observation the.GEN Uefa this go_on now already three
Monate an.
months PART
'The French club is under UEFA's observation as part of Financial Fair Play regulations. This has now been ongoing for three months.'

According to Ježek there is a connection between syntactic flexibility and the referentiality of the noun. She writes: "As it happens, the constructions whose members are not fully flexible syntactically are those in which the noun is non-referential"

5 The examples come from the German Reference Corpus (Leibniz-Institut für Deutsche Sprache, 2021) and were extracted during the corpus analysis described in Fleischhauer (2024).

6 See Doron (1988) for an intense discussion of criteria distinguishing between referential and predicative uses of nominals.

(Ježek, 2016, 206). That not all LVCs in a language behave uniformly with regard to the feature of syntactic flexibility is also described by Esteban (this volume) for Scottish Gaelic. One consequence is that LVCs either do not represent a syntactically uniform category or that the term 'LVC' is restricted to either complex predicates that are syntactically flexible or those that are not. Such an approach is, for example, proposed by Fleißner and Smirnova (2025) for German.

Regarding the nominal component of the LVC, Ježek also makes an assumption that is important to discuss. She writes that the noun is an "event noun" (Ježek, 2016, 204). While non-eventive nouns refer to non-temporal entities (e.g., people, books, or abstract concepts like 'happiness'), eventive nouns refer to eventualities (see Fábregas et al., 2012; Fábregas and Marín, 2012 for criteria to distinguish eventive nouns from non-eventive ones). The assumption that the noun occurring in LVCs must be eventive is sometimes state in the literature (von Polenz, 1963, 1987; Alego, 1995; Everaert and Hollebrandse, 1995; Helbig, 2006; Langer, 2005; Ronan and Schneider, 2015; Brugman, 2001, 553; Wittenberg et al., 2014a, 61; Wittenberg et al., 2014b, 31; Bruening, 2015, 55; Bonial and Pollard, 2020, 585). Sometimes this claim is also only implicit and follows from assuming that the noun has a deverbal origin (e.g., Wierzbicka, 1982; Dixon, 1991; Shahrokny-Prehn and Höche, 2011; Plante, 2014; Wittenberg, 2016; Giparaitė and Selmistraitis, 2024. Typically, there is no explicit justification for this assumption. However, one possible reconstruction of an argument could be as follows: light verbs are semantically deficient and do not denote an eventuality. A sentence predicate, however, must denote an eventuality. Therefore, the predicative core of the LVCs – the noun – must contribute the eventuality to the predication. As a consequence, the noun has to be eventive.

In many cases, the nominal component of an LVC is indeed an eventive noun, regardless of whether it is derived by conversion (as it is the case with English *kiss*) or derivation (as with German *Beobachtung* (8), which is derived from the verb *beobachten* 'to observe' via *ung*-nominalization) or has no deverbal origin (such as the German noun *Stress* 'stress' in (6)). However, it is less clear whether constructions with non-eventive nouns should also be classified as LVCs. Ježek, for instance, treats *give an answer* as an LVC, while *give a hand* – which contains a non-eventive noun – is considered to be an idiomatic expression (Ježek, 2016, 204). Idiomatic expressions differ from LVCs concerning their semantic composition. LVCs are "built syntagmatically by means of compositional processes" (Ježek, 2016, 209). That is, the meaning of an LVC can be derived from the meanings of its components (see, for instance, Fleischhauer and Gamerschlag, 2019; Fleischhauer et al., 2019 for a formal analysis of the semantic composition of German LVCs of the type *stehen vor* NP; lit. stand in front of NP). Idiomatic expressions, in contrast, have a meaning that cannot be traced back to their components. The meaning 'die', as expressed by *kick the bucket*, cannot be derived from the meanings of the individual parts of the idiom

(e.g., Gibbs et al., 1989; Nunberg et al., 1994; Bonial and Pollard, 2020, 579. With respect to idioms, Nunberg et al. (1994) further distinguish between 'idiomatic expressions' like *kick the bucket* on the one hand and 'idiomatically combining expressions' on the other hand. Idiomatically combining expressions have components that bear an idiomatic interpretation, but once identified, the overall meaning is derived compositionally from the individual components. Some authors propose that LVCs are either idiomatically combining expressions (Samvelian and Faghiri, 2013a,b, 2014; Fleischhauer, 2020; Fleischhauer and Neisani, 2020; Fleischhauer and Gamerschlag, 2019; Nicoletti, 2025) or another type of semi-compositional construction (Bonial and Pollard, 2020) since they exhibit some flavor of idiomaticity (see also Riccio's contribution to this volume on this topic).

At this point, let us return to the question of whether the nouns appearing in LVCs must be eventive. Our impression is that such an assumption is primarily made by authors who study specific languages – e.g., Germanic or Romance languages – and thus might be influenced by this selection of languages. A counterargument to the assumption that this is the case could be that non-eventive nouns are also associated with specific events. This is most clearly represented in Pustejovsky's (1991; 1995) 'Generative Lexicon theory.' Pustejovsky assigns a qualia structure to lexical expressions, which is partially illustrated for the noun *violine* in (10).

(10) a. Formal role: a violine is a MUSICAL INSTRUMENT.
b. Agentive role: the violine is created through the event of BUILDING or, more generally, CREATING.
c. Telic role: The violin is created for PRODUCING MUSICAL SOUND.
(Pustejovsky and Batiukova, 2016, 162)

For a structured meaning representation, several qualia roles are distinguished. Via the 'Formal role,' the "basic semantic type, including features that distinguish the object within a larger domain" (Pustejovsky and Batiukova, 2016, 162) is described. For a violin, it is central that it is a musical instrument. The 'Agentive role' specifies how an object comes into being. This can be, as with the violin, very general creation events, but it can also be more specific events if a corresponding lexicalization exists (for example in German *töpfern* 'to do pottery' in the case of earthenware vessel). The 'Telic role' describes the purpose that an object serves. A violin's purpose is to enable making music with it. As Pustejovsky and Batiukova (2016, 163) write, there are two qualia roles "that encode events, Agentive and Telic […]". As the authors explain, the events associated with these roles can be inferred as 'hidden events.' They illustrate this with the example *Antonio Stradivari finished the violin*, which is to be interpreted as 'finished building the violin,' where a creation event is inferred via the Agentive role. We could similarly assume this for complex predicates like *piga marimba* 'play

the xylophone' (lit. hit xylophone) from Swahili (11). *Piga* 'hit, beat' occurs with many nouns as light verbs, including various names for musical instruments (e.g., *vigoma* 'tambourine,' *gitaa* 'guitar' and *fidla*/ *zeze* 'violin;' cf. Olejarnik, 2009, 161). Musical instruments are artifacts and as such do not denote events. However, in all complex predicates where *piga* combines with a noun for a musical instrument, the interpretation is 'play the instrument' (= use the instrument to make music). This yields a regular inference of the event specified in the Telic role (cf. the qualia structure for the musical instrument in (10)).

(11) *Ni-li-wez-a ku-pig-a marimba kidogo lakini si-ku-wez-a*
1-PST-can-FV INF-hit-FV 9.xylophone little but 1.NEG-PST-can-FV
ku-pat-a maneno.
INF-find-FV 6.words
'I could play the xylophone a little, but I couldn't find words.'
(Olejarnik, 2009, 161; glossing slightly adopted)

As in the case of unambiguous LVCs such as *give a kiss*, the denoted event is contributed by the noun. The difference between *kiss* and *marimba* 'xylophone' is merely whether the event is denoted directly by the noun itself (as with *kiss*) or inferred from the noun's meaning as a hidden event.

If we say that the noun occurring in LVCs has to be eventive, we take a property of the noun – to refer to eventualities – as definitional. This claim is defensible – see the reconstructed argument above – but still arbitrary. If we do not make this assumption, we do not immediately constrain LVCs to complex predicates with a particular semantic type of noun. For languages like Swahili and Gĩkũyũ (cf. Fleischhauer and Kihara, this volume), this definitional question is central. If we propose an affirmative answer to the question of whether the noun in LVCs must be eventive, then both languages have none – or at most few – LVCs. If we answer it negatively, then both languages exhibit (numerous) LVCs. But also for other languages, such as Persian, the answer to the question is central. The light verb *kešidan* 'pull' occurs with eventive (12a) as well as non-eventive nouns (12b) (see also Family, 2011). With nouns like *sigar* 'cigarette' and other smokeable objects, a use event is inferred (licensed by the Telic role). With nouns like *jâde* 'road,' on the other hand, a creation event is inferred via the Agentive role. With eventive nouns finally, the event is directly contributed by the noun.

(12) a. *nafas kešidan* 'to breath' (lit. breath pull)
b. *sigar kešidan* 'to smoke cigarettes' (lit. cigarette pull), *jâde kešidan* 'to build a road' (lit. road pull)

If the notion of an LVC is restricted to complex predicates with eventive nouns, the example in (12a) would be an LVC, but not the two in (12b). If we do not make this assumption, then all the examples in (12) fall under the label 'light verb construction.' And as it seems, authors working on languages such as Swahili, Persian or Udi have fewer reservations about extending the term 'light verb construction' to complex predicates with a non-eventive nominal element (see, for instance, Folli et al., 2005; Olejarnik, 2009, 2011; Karimi-Doostan, 2011; Family, 2008, 2011; Harris, 2008).[7]

If we do not restrict LVCs to constructions of light verb + eventive noun, that does not mean that the boundary to idiomatic expressions is lost. It would not follow that, for instance, *kick the bucket* is an LVC since the event 'die' could not be inferred from the qualia structure of bucket. Neither is the dying event a kind of production of buckets (agentive role), nor is it a telic role associated with buckets.

Ježek's conception of LVCs captures various properties that are certainly central to LVCs – e.g., syntactic flexibility, referentiality of the noun – but it is not clear whether these are sufficient or even only necessary criteria for identifying a complex predicate as an LVC. Ježek herself points out, as mentioned above, that there is a connection between syntactic flexibility and the referentiality of the nominal element. Thus, it seems that both features are not necessary for LVCs. On the other hand, there is the assumption that the nominal element must be eventive. However, it seems relatively arbitrary to restrict LVCs to complex predicates with eventive nouns.

Many of the assumptions about LVCs that we find in Ježek's conceptions are also present in the authors gathered in this volume. The idea that LVCs contain a non-eventive noun is explicitly mentioned only by Boldt (this volume). However, it should be noted that Boldt refers to 'prototypical light verb constructions,' thereby not excluding the possibility that non-prototypical LVCs with a non-eventive noun may exist.[8]

Greater agreement among the authors and with Ježek's conception of LVCs can be found regarding the following properties:

i. LVCs are multi-word expressions consisting of a light verb and another component.

Ourang et al. (this volume) use the term 'non-verbal element,' a designation that is particularly common in works on LVCs in Iranian languages. Depending on the language, this category includes not only nouns but also other linguistic units that

7 See Lee (2011) for Korean LVCs with non-eventive nouns. For German, the existence of LVCs with non-eventive nouns is stated in Fleischhauer (2023a).

8 Similar applies to Wittenberg (2016, 12), who discusses LVCs with deverbal nouns as "the most clear-cut cases of German and English light verb constructions."

do not behave morphosyntactically like a verb (adjectives (13), adverbs, particles, etc.; see especially the discussion in Maisak, this volume on this issue).

(13) *pahn kardan* [Persian]
wide make
'to spread/ to widen'
(Folli et al., 2005, 1370)

It is not surprising that, particularly in analyses on Iranian languages, the assumption that an LVC must contain an eventive noun is not particularly prominent. This is probably because it is generally assumed that expressions of lexical categories not associated with eventivity (i.e., adjectives, adverbs) can also occur as non-verbal elements in LVCs.

ii. The light verb and the non-verbal element form a joint predication, with the predicative content primarily contributed by the non-verbal element.

Lai (this volume, 119) describes this as follows: "[…] LVCs have a semantically specific noun carrying most of the semantics plus a semantically general verb encoding relatively little information […]." Riccio (this volume, 180), who limits her analysis to constructions with an NP as the non-verbal element, states that this NP is the semantic head of the complex predicate, while the light verb functions as the syntactic head. In regular predicate-argument constructions, the syntactic head and semantic head coincide, whereas idiomatic expressions have a syntactic head but no semantic head. This results in LVCs having a specific distribution of meaning components that distinguishes them from other predicative constructions (cf. e.g., the discussion in the chapters of Lai and Ourang et al).

iii. LVCs are a specific type of collocation.

This view is particularly prominent in Lai's collocation analysis of Tibetan LVCs but also shows up in Riccio's contribution.

Although the authors gathered in this volume do not explicitly use a common conception of LVCs, we believe that a unified definition can be developed that takes into account the aspects discussed above and allows for an operationalization of the concept of a light verb construction.

3 Towards a definition of 'light verb constructions'

The brief discussion of Ježek's conception of light verb constructions has shown that the range of phenomena we capture as LVCs depends strongly on which defini-

tional properties we assign to light verb constructions. Must the nominal element be eventive or not? May the nominal element be referential or not (as claimed by e.g., Langer, 2004; Eisenberg, 2013, 308)? Is article variation within LVCs possible or not? Depending on how we answer these questions, we select different linguistic constructions that exhibit contradictory properties. If, for instance, LVCs do not allow variation of the article, then far fewer linguistic expressions – at least in languages that have grammaticalized articles – fall into the category of LVCs than if we do not make this assumption. The German complex predicates with *stehen* 'stand' mentioned above would then not be LVCs. Such a formal criterion can, at most, be useful for a language-specific definition, since it is irrelevant for languages without grammaticalized articles (e.g., Swahili and Gĩkũyũ). Therefore, semantic criteria (referentiality, eventivity) are better suited for a possible definition, as these are independent of which functional categories a language has grammaticalized. Nevertheless, there should be good justification for why such a criterion should be definitional for light verb constructions. It might be more interesting if such properties are not taken as definitional for LVCs but, if at all, emerge as a result of an empirical investigation of LVCs. Then it would also be possible that languages differ with respect to these properties. In some languages, LVCs could be restricted to constructions with eventive nouns, in others not. In some languages, the nominal elements in LVCs could be referential, in others not. Such an approach would allow us to determine which formal properties (if any) LVCs exhibit cross-linguistically and in terms of which features they can vary. This, however, presupposes that we do not already incorporate these properties into the definition of LVCs. A first proposal for such a definition is:

(14) Light verb constructions are multi-word expressions which form complex predicates consisting of (at least) a light verb and a non-verbal element. The eventuality denoted by the complex predicate is determined by the non-verbal element rather than the verb.[9]

The definition consists of two formal criteria and one semantic criterion. The first formal criterion – LVCs are multi-word expressions (e.g., Masini, 2021) – is used to delimit this type of complex predicates from complex predicates that form a morphological unit. This second type is represented, for example, by prefixed verbs (e.g., English *outperform*), verbal compounds (e.g., English *proofread*), and morphological noun incorporation (15), which – as Mithun (1984, 2010) correctly states – can also be considered a specific type of compounding.

9 A similar definition is proposed in Fleischhauer (2020, 4), Fleischhauer (2021b, 34–35) and Fleischhauer and Turus (2021, 76).

(15) *wa-hati-itsi-a-ient-a'n-e'* (Mohawk < Iroquoian)
FACTUAL-M.PL.AGT-fish-LK-have-INCH-PERF
'They caught much fish.'

(Mithun, 2010, 40)

Composition and, in particular, incorporation play a central role in the discussion of LVCs. Various authors see parallels between compounds on one side and LVCs on the other (e.g., Ghomeshi and Massam, 1994; Vahedi-Langrudi, 1996; Mahmoodi-Bakhtiari, 2018, but also see Goldberg, 1996; Müller, 2010 on this issue). As already discussed above, it is sometimes claimed that the nominal element occurring in LVCs is not referential and, accordingly, is not – or at least only very limitedly – combinable with articles, and also exhibits other restrictions regarding functional morphology (such as number and case) that strongly resemble the semantic and grammatical behavior of incorporated nouns (See Massam, 2001 and Borik and Gehrke, 2015 for an overview on the semantic and grammatical properties of (pseudo-)incorporated nouns. For a discussion of the differences between (Persian) LVCs and psuedo-incorporation constructions see Fleischhauer, 2020, 2021b). Generally, however, LVCs do not form a single word like compounds do; instead, they constitute a multi-word expression. This can be observed in that the components of an LVC do not have to be adjacent to each other but can be separated by linguistic expressions that clearly have word status. This has already been shown above using German language data, but for illustration also see the Persian language data in (16).

(16) *mesâl e xub-i zadan*
example EZ good-INDEF hit
'to give a good example'

(Fleischhauer and Neisani, 2020, 47)

The second formal criterion within the definition states that LVCs consist of at least one verb and a non-verbal element. As we have seen, it is possible – although not mandatory – for articles and prepositions to be part of an LVC. The definition also allows a light verb to combine with an adjective or some other element which is not a verb. Excluded by the definition are all constructions in which a verb with another verb[10], for example in the form of serial verb constructions.

As a semantic criterion, the definition relies on the fact that the denoted event is determined by the noun and not by the verb. Thus, LVCs differ on the one hand from

10 By 'verb' it is meant of clear (language-specific) instance of the lexical category verb but not a deverbal noun or a participle, i.e., items which are derived from nouns but do not behave like (regular) nouns morphosyntactically (cf. the discussion on verbal nouns in Scottish Gaelic by Esteban, this volume, 29).

regular predicate-argument constructions (e.g., *to give someone something*), since in those the verb denotes directly. They also differ from idiomatic expressions, since in those the denoted event is not determined by either the verb or the noun (cf. the discussion of *kick the bucket*). However, expressions with eventive nouns (e.g., *give a kiss*) fall under the definition just as those with non-eventive nouns (e.g., Swahili *piga marimba* 'play xylophone'), because in both cases it is the noun that determines the denoted eventuality. Either directly, since the noun denotes the eventuality, or indirectly, since the denoted eventuality is inferred from the noun.

The property that the verb is semantically defective, i.e.,it does not denote an eventuality of its own (e.g., Butt and Geuder, 2001, 356), is shared by light verbs, among others, with auxiliary verbs (cf. Mohanan, 2006, 481). It is therefore not surprising that some authors view 'light verbs' as a transitional step in auxiliarization (e.g., Hook 1991, Bernander 2024). The definition in (14) is not yet sharp enough to delineate LVCs from auxiliary verb constructions, since some (quasi-)auxiliary constructions satisfy the same conditions.[11] An illustrative example, again from Swahili, can be seen in (17).

(17) a. *Baba y-angu a-na-tak-a ku-j-a.*
1.father 9-1SG.POSS 1-PRS-want-FV 15-come-FV
'My father wants to come.'

b. *Mi-ti h-uu u-na-tak-a ku-anguk-a.*
3-tree DEM.PRX-3 3-PRS-want-FV 15-fall-FV
'The tree is about to fall.'

(Narrog and Heine, 2021, 62; glossing slightly changed)

The verb *taka* means, in its use as a lexical heavy verb, 'want, like' (17a). In an auxiliary use it is used with a nominalized infinitive and serves to express prospective aspect (cf. Heine, 1994; Narrog and Heine, 2021, 62). While in (17a) the eventuality – a state of wanting – is denoted by the lexical verb, the verb in (17b) contributes a perspectivization of the event denoted by the noun. Quasi-auxiliaries as in (17b), according to the definition in (14), constitute LVCs. This raises the question whether LVCs and quasi-auxiliaries as in (17b) can be distinguished at all, and if so, by which properties. Fleischhauer and Kihara (this volume) address this question for the Bantu

11 We have not discussed the delimitation of LVCs from, for example, periphrastic causative constructions, or, in general, from raising and control constructions. Such a discussion is given in detail by Butt (2010), who identifies monoclausality as the relevant feature for delimiting LVCs from the other mentioned syntactic constructions (see, also, Mohanan, 2006, 471).

language Gĩkũyũ and argue that both exhibit different grammatical properties, which is due to quasi-auxiliaries showing a more advanced grammaticalization than LVCs.[12]

The definition in (14) explicitly refers to the fact that LVCs consists at least of a light verb and a non-verbal element. If we regard light verbs as desemanticized verbs – as opposed to the verb's heavy verb usage – that show no signs of decategorialization, we can distinguish them from (quasi-) auxiliaries, which we define by definition as desemanticized (in comparison to the heavy verb usage) and at least partially decategorialized (this is basically in line with Bowern's (2004) claim that auxiliaries but not light verbs have defective paradigms in Turkish and other Turkic languages). Thus, (17b) would no longer fall under the notion of a 'light verb construction.'[13]

Even though the definition in (14) is not explicitly endorsed by the authors gathered in this volume it nonetheless captures the essential common features that the authors rely on for LVCs in their investigations, and can thus be presented as an implicit frame of reference for the conception of LVCs used here (without implying that all authors would also necessarily accept this definition.).

4 Overview of the volume

The contributions in this volume address many of the topics mentioned above. These include questions regarding the role of the separability of the components of an LVC for the analysis of LVCs. Another question that is explored is whether the nominal element is referential and what this implies for the analysis of LVCs. However, the unifying element of the contributions is the question of how LVCs differ from other constructions, including regular predicate-argument constructions, multi-word units, and also compounds.

The contributions, on one hand, develop (language-specific) criteria for the identification of LVCs and their distinction from other predicative construction types. On the other hand, they present language-specific analyses of specific aspects, including the conditions for the emergence of LVCs in Old Norse, the interaction of in/definiteness and LVCs, as well as aspects of their meaning composition. Although

12 More concretely: LVCs exhibit desemanticization of the verb, but no decategorialization of the noun. Auxiliary verb constructions, in addition to the desemanticization of the verb, it also exhibit a (partial) decategorialization.

13 Mohanan (2006, 482) argues that LVCs can be distinguished form auxiliary verb constructions by the fact that the light verb can contribute to argument structure, whereas auxiliaries do not. We think that this might a further relevant property separating these two types of verb usages, however it might follow from the fact that auxiliaries but not light verbs show signs of decategorialization.

the authors primarily address LVCs from a single-language perspective, they also take into account similarities and differences with other languages. In addition to the Indo-European languages Italian (Romance; Riccio), Irish Gaelic (Celtic; Esteban), and Aheli (Iranian; Ourang, Amberber & Deligianni), non-Indo-European languages are also addressed: Kina Rutul (Nakh-Daghestania; Maisak), Tibetan (Sino-Tibetan; Lai), and Gĩkũyũ (Bantu; Fleischhauer & Kihara). The chapters share a common focus on the characteristics of LVCs in the respective languages under investigation. While Italian is among the languages that have been studied more extensively in this regard, the studies on other languages (e.g., Aheli and Gĩkũyũ) represent something new.

How do light verb constructions differ from serial verb constructions and auxiliary verb constructions? **Esteban** addresses this question in his study of LVCs in Scottish Gaelic. Based on the identification of semantic and morphosyntactic characteristics of LVCs, Esteban argues that LVCs represent a heterogeneous type of complex predicates. Nevertheless, uniform functions for the light verb can be identified, which are modeled within a decomposition analysis. Based on the framework of Role and Reference Grammar, the author formulates a syntactic analysis of LVCs in terms of nuclear cosubordination.

Light verb constructions in Aheli, a dialect of the Indo-Iranian language Lari, are the subject of the paper by **Ourang, Amberber & Deligianni**. The authors base their study on works related to the Persian language and argue that LVCs in Aheli can be categorized into several semantic templates. In this way, a systematic contribution of meaning from the non-verbal element is identified. In contrast to Persian, various strategies for forming LVCs are discussed: combination and (pseudo-)incorporation. The authors argue that both strategies are applied in Aheli but lead to different types of LVCs. Additionally, the paper descriptively highlights some differences between LVCs in Aheli and those in Persian.

In their contribution to complex predicates in the Bantu language Gĩkũyũ, **Fleischhauer** and **Kihara** compare different types of verb-noun combinations, in which the verb is semantically reduced compared to its full verb variant. A central result of the study is that light verb constructions do not exhibit significant grammatical differences from regular predicate-argument constructions. In contrast, linguistic units that are more likely to be considered auxiliary verb constructions show grammatical differences that indicate, on one hand, that these constructions are subject to a process of grammaticalization, and on the other hand, allow for distinguishing LVCs from auxiliary verb constructions based on grammatical characteristics.

Light verbs typically combine with a limited number of nouns to form a light verb construction. Statistical methods can be used to conduct collocation analyses. Based on the fact that the association strength between the light verb and the noun is not equally strong, but rather that the light verb attracts the noun more than

vice versa, **Lai** employs unidirectional association measures to explore how LVCs in modern Tibetan can be distinguished from other frequent verb-noun combinations using statistical distributions.

Maisak's paper describes various complex verbs, such as 'noun + light verb' and 'adjective + light verb' combinations in the Kina Rutul dialect, a Lezgic language, where verbs are formed by combining a lexical component with a light verb (e.g., 'do', 'be', 'say'). The author examines the challenges of defining complex verb, noting that while many of these can be analyzed as periphrastic constructions, some, particularly those with acategorical coverbs, function as true complex verbs, where the coverbs convey the primary lexical meaning.

The research conducted by **Riccio** explores the morpho-syntactic flexibility and semantic nuances of Italian light verb constructions, with particular attention to the variation in the use of definite and indefinite articles with event-denoting nouns. Through examples such as *fare una/la doccia* 'to take a/the shower' and *dare una/la opinione* 'to give an/the opinion', the analysis, grounded in the framework of Role and Reference Grammar, shows the intricate relationship between syntax and semantics. The findings reveal that the choice of articles significantly outlines the specificity and interpretation of events within LVCs. Moreover, the noun phrases in these constructions, while closely associated with the light verb, retain their syntactic independence, thereby exhibiting characteristics akin to pseudo-incorporation.

A historical linguistic approach to light verb constructions in Old West Norse, particularly their connection to the weakening of the Germanic weak verb system, leads **Boldt** to reflect on the key features of prototypical constructions in Germanic languages, as well as the semantic features of Proto-Germanic and Old West Norse weak verbs. The study suggests that the instability of the weak verb system led to the rise of light verb constructions in Old West Norse to express specific semantic nuances. It also highlights that certain denominal verbs did not transition into light verb constructions.

Acknowledgements

We would like to take this opportunity to thank the reviewers who critically and constructively read the contributions to this volume and provided valuable comments to the authors. Additionally, we would like to thank the De Gruyter team for their active support in realizing the project. The work on this volume is part of the research project 'Funktionsverbgefüge: Familien & Komposition' (HE 8721/1-3) funded by the German Research Foundation (DFG). Finally, we like acknowledging the financial support by the Universitäts- & Landesbibliothek Düsseldorf via an Open Access grant.

Bibliography

Alba-Salas, Josip. 2002. *Light Verb Constructions in Romance: A Syntactic Analysis*. Doctoral dissertation: Cornell University, Ithaca, NY.

Alego, John. 1995. Having a look at the expanded predicate. In Bas Aarts and Charles F. Meyer (eds.), *The verb in contemporary English: Theory and description*, 203–217. Cambridge: Cambridge University Press.

Allerton, David. 2003. *Streched verb constructions in English*. London: Routledge.

Alvarez-Morera, Georgina. 2023. *The nominal in light verb constructions: a corpus-based study in present-day English, German, Catalan and Spanish*. Doctoral dissertation: Universitat Rovira i Virgili, Tarragona.

Alvarez-Morera, Georgina, Jordi Ginebra and Isabel Oltra-Massuet. 2025. Modification in *give* light verb constructions. In Anna Riccio and Jens Fleischhauer (eds.), *Light verbs: Synchronic and diachronic studies*, 65–96. Düsseldorf: Düsseldorf University Press. https://doi.org/10.1515/9783111388878-004.

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

Boldt, Ramón. this volume. To come into bloom: on the emergence of light verb constructions in Old West Norse. In Jens Fleischhauer and Anna Riccio (eds.), *Light verb constructions: Complex predicates from a cross-linguistic perspective*, 205–235. Berlin: De Gruyter.

Bonial, Claire and Kimberly A. Pollard. 2020. Choosing an event description: What a PropBank study reveals about the contrast between light verb constructions and counterpart synthetic verbs. *Journal of Linguistics* 56(3). 577–600. https://doi.org/10.1017/S0022226720000109.

Borik, Olga and Berit Gehrke. 2015. An Introduction to the Syntax and Semantics of Pseudo-Incorporation. In Olga Borik and Berit Gehrke (eds.), *The Syntax and Semantics of Pseudo-Incorporation*, 1–43. Leiden: Brill.

Bowern, Claire. 2004. (Some notes on) light verbs and complex predicates in Turkic. *Proceedings of the Workshop in Altaic Formal Linguistics I, MIT Working Papers in Linguistics* 33–48.

Bowern, Claire. 2008a. The Diachrony of Complex Predicates. *Diachronica* 25. 161–185.

Bowern, Claire. 2008b. The reconstruction of Nyulnyulan complex predication. *Diachronica* 25(2). 186–212. https://www.jbe-platform.com/content/journals/10.1075/dia.25.2.04bow.

Bowern, Claire. 2010. The typological implications of Bardi complex predicates. *Linguistic Typology* 14. 39–70. https://doi.org/10.1515/lity.2010.002.

Bowern, Claire. 2014. Complex predicates in Australian languages. In Harold Koch and Rachel Nordlinger (eds.), *The languages and linguistics of Australia: A comprehensive guide*, 263–294. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110279771.

Bruening, Benjamin. 2015. Light Verbs are Just Regular Verbs. *U. Penn Working Papers in Linguistics* 21(1). 1–10.

Brugman, Claudia. 2001. Light verbs and polysemy. *Language Science* 23. 551–578.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications.

Butt, Miriam. 2003. The Light Verb Jungle. *Harvard Working Papers in Linguistics, Papers from the GSAS/Dudley House Workshop on Light Verbs* 9. 1–49.

Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Mengistu Amberber, Bret Baker and Mark Harvey (eds.), *Complex Predicates in Cross-Linguistic Perspective*, 48–78. Cambridge: Cambridge University Press.

Butt, Miriam and Wilhelm Geuder. 2001. On the (semi)lexical status of Light Verbs. In Norbert Corver and Henk van Riemsdijk (eds.), *Semilexical Categories: On the Content of*

Function Words and the Function of Content Words, 323–370. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110874006.323.

Butt, Miriam and Aditi Lahiri. 2013. Diachronic pertinacity of light verbs. *Lingua* 135. 7–29. https://doi.org/10.1016/j.lingua.2012.11.006.

Croft, William. 2012. *Verbs: aspect and causal structure*. Oxford: Oxford University Press.

De Pasquale, Noemi. 2023. *Making a move* towards Ancient Greek Light Verb Constructions. In Anna Pompei, Lunella Mereu and Valentina Piunno (eds.), *Light Verb Constructions as Complex Verbs*, 257–274. Berlin: De Gruyter. https://doi.org/10.1515/9783110747997-010.

Di Salvo, Francesca. 2023. Light Verb Constructions in Latin. A study on (*in*) *memoria* and (*in*) *animo habeo*. In Anna Pompei, Lunella Mereu and Valentina Piunno (eds.), *Light Verb Constructions as Complex Verbs*, 337–360. Berlin: De Gruyter. https://doi.org/10.1515/9783110747997-013.

Dixon, Robert. 1991. *A new approach to English grammar, on semantic principles*. Oxford: Clarendon.

Doron, Edit. 1988. The semantics of predicate nominals. *Linguistics* 26. 281–301.

Eisenberg, Peter. 2013. *Grundriss der deutschen Grammatik. Der Satz*. Stuttgart: J.B. Metzler.

Esteban, Avelino Corral. this volume. Light verb constructions in Scottish Gaelic. In Jens Fleischhauer and Anna Riccio (eds.), *Light verb constructions: Complex predicates from a cross-linguistic perspective*, 27–62. Berlin: De Gruyter.

Everaert, Martin and Bart Hollebrandse. 1995. The Lexical Representation of Light Verb Constructions. In *Proceedings of the Twenty-First Meeting of the Berkeley Linguistics Society*, 94–104.

Family, Neiloufar. 2008. Mapping semantic spaces: A constructionist account of the "light verb" xordæn "eat" in Persian. In M. Vanhove (ed.), *From polysemy to semantic change. Towards a typology of lexical semantics associations*, 139–162. Amsterdam: John Benjamins.

Family, Neiloufar. 2011. Verbal islands in Persian. *Folia Linguistica* 45(1). 1–30.

Fleischhauer, Jens. 2020. Predicative multi-word expressions in Persian. In *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, 552–561. Hanoi: Association for Computational Linguistics. https://aclanthology.org/2020.paclic-1.63.

Fleischhauer, Jens. 2021a. Light Verb Constructions and Their Families – A Corpus Study on German '*stehen unter*'-LVCs. In *"Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021)*, 63–69. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.mwe-1.8.

Fleischhauer, Jens. 2021b. Simplex and Complex Predicates in Persian: An RRG Analysis. In Robert Van Valin (ed.), *Challenges at the Syntax-Semantics-Pragmatics Interface*, 31–62. Newcastle upon Tyne: Cambridge Scholars Publishing.

Fleischhauer, Jens. 2022. *stehen unter*-Funktionsverbgefüge und ihre Familien. *Zeitschrift für germanistische Linguistik* 50(2). 247–288. https://doi.org/10.1515/zgl-2022-2055.

Fleischhauer, Jens. 2023a. German light verb constructions with non-eventive nouns. In Jean-Pierre Colson (ed.), *Phraseology, constructions and translation. Corpus-based, computational and cultural aspects*, 387–398. Louvain: Presses Universitaires de Louvain.

Fleischhauer, Jens. 2023b. Prospective aspect and current relevance – A case study of the German prospective '*stehen vor* NP'-construction. *Journal of Germanic Linguistics* 35(4). 371–408. https://doi.org/10.1017/S1470542722000198.

Fleischhauer, Jens. 2024. Eine datengesteuerte Untersuchung deutscher Funktionsverbgefüge. Artikelvarianz in *stehen*-Funktionsverbgefügen. *Lingue e Linguaggi* 65. 233–265. http://siba-ese.unisalento.it/index.php/linguelinguaggi/article/view/30024/24502.

Fleischhauer, Jens and Thomas Gamerschlag. 2019. Deriving the meaning of light verb constructions – a frame account of German *stehen* 'stand'. In Constanze Juchem-Grundmann, Michael Pleyer and Monika Pleyer (eds.), *Yearbook of the German Cognitive Linguistics Association, vol. 7*, 137–156. Berlin: De Gruyter Mouton. https://doi.org/10.1515/gcla-2019-0009.

Fleischhauer, Jens, Thomas Gamerschlag, Laura Kallmeyer and Simon Petitjean. 2019. Towards a compositional analysis of German light verb constructions (LVCs) combining lexicalized tree adjoining grammar (LTAG) with frame semantics. In *Proceedings of the 13th international conference on computational semantics - long papers*, 79–90. Gothenburg: Association for Computational Linguistics. https://www.aclweb.org/anthology/W19-0407.

Fleischhauer, Jens and Claudius Patrick Kihara. this volume. Light verbs and aspectual semi-auxiliaries in Gĩkũyũ. In Jens Fleischhauer and Anna Riccio (eds.), *Light verb constructions: Complex predicates from a cross-linguistic perspective*, 87–116. Berlin: De Gruyter.

Fleischhauer, Jens and Mozhgan Neisani. 2020. Adverbial and attributive modification of Persian separable light verb constructions. *Journal of Linguistics* 56(1). 45–85. https://doi.org/10.1017/S0022226718000646.

Fleischhauer, Jens and Dila Turus. 2021. Der Angeklagte steht unter Schutz, wird er aber auch geschützt? – Eine Analyse passivischer Funktionsverbgefüge des Typs 'stehen unter'. *Germanistische Werkstatt* 11. 73–81. 10.25167/pg.4671.

Fleißner, Fabian and Elena Smirnova. 2025. Funktionsverbgefüge unter Druck: Ein Vorschlag zur Typologisierung verbonominaler Konstruktionen des Deutschen. *Zeitschrift für germanistische Linguistik* 53(1). 126–165. https://doi.org/10.1515/zgl-2025-2004.

Folli, Raffaella, Heidi Harley and Simin Karimi. 2005. Determinants of event type in Persian complex predicates. *Lingua* 115. 1365–1401.

Fábregas, Antonio and Rafael Marín. 2012. The role of Aktionsart in deverbal nouns. State nominalizations across languages. *Journal of Linguistics* 48. 35–70. https://doi.org/10.1017/S1470542722000198.

Fábregas, Antonio, Rafael Marín and Louise McNally. 2012. From psych verbs to nouns. In Violeta Demonte and Louise McNally (eds.), *Telicity, change, and state*, 162–184. Oxford: Oxford University Press.

Geist, Ljudmila. 2006. *Die Kopula und ihre Komplemente*. Tübingen: Niemeyer.

Ghomeshi, Jila and Diane Massam. 1994. Lexical/syntactic relations without projections. *Linguistic Analysis* 24 (3–4). 175–217.

Gibbs, Raymond W., Nandini P. Nayak and Cooper Cutting. 1989. How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language* 28(5). 576–593.

Giparaitė, Judita and Linas Selmistraitis. 2024. Light verb constructions with deverbal nouns BITE and SNACK in native English varieties: a corpus-based study. *Studies about Languages / Kalbų studijos* 45. 89–106. https://doi.org/10.5755/j01.sal.1.45.38448.

Goldberg, Adele. 1996. Words by default: Optimizing constraints and the Persian complex predicate. *Annual Proceedings of the Berkeley Linguistics Society* 22. 132–146.

Grimshaw, Jane and Armin Mester. 1988. Light verbs and theta-marking. *Linguistic Inquiry* 19. 205–232.

Haig, Geoffrey. 2002. Noun-plus-verb complex predicates in Kurmanji Kurdish: Argument sharing, argument incorporation, or what? *Sprachtypologie & Universalienforschung* 55. 15–48. https://doi.org/10.1524/stuf.2002.55.1.15.

Harm, Volker. 2021. *Funktionsverbgefüge des deutschen: Untersuchungen zu einer Kategorie zwischen Lexikon und Grammatik*. Berlin: De Gruyter. https://doi.org/10.1515/9783110661255.

Harris, Alice C. 2008. Light verbs as classifiers in Udi. *Diachronica* 25(2). 213–241. https://doi.org/10.1075/dia.25.2.05har?locatt=mode:legacy.

Heine, Antje. 2020. Zwischen Grammatik und Lexikon. Ein forschungsgeschichtlicher Blick auf Funktionsverbgefüge. In Sabine De Knop and Manon Hermann (eds.), *Funktionsverbgefüge im Fokus*, 15–37. Berlin: De Gruyter. https://doi.org/10.1515/9783110697353-002.

Heine, Bernd. 1994. On the genesis of aspect in African languages: The proximative. *Proceedings of the 20th Annual Meeting of the Berkeley Linguistics Society* 35–46.

Helbig, Gerhard. 1984. Probleme der Beschreibung von Funktionsverbgefügen im Deutschen. In Gerhard Helbig (ed.), *Studien zur deutschen Syntax, Band 2*, 163–188. Leipzig: Verlag Enzyklopädie Leipzig.

Helbig, Gerhard. 2006. Funktionsverbgefüge – Kollokationen – Phraseologismen. In Ulrich Breuer and Jarmo Korhonen (eds.), *Wörter – verbindungen*, 165–174. Frankfurt am Main: Lang.

Hellan, Lars. 2023. Unification and selection in Light Verb Constructions. A study of Norwegian. In Pompei Anna, Mereu Lunella and Piunno Valentina (eds.), *Light Verb Constructions as Complex Verbs*, 45–70. Berlin: De Gruyter. https://doi.org/10.1515/9783110747997-003.

Heringer, Hans Jürgen. 1968. *Die Opposition von 'kommen' und 'bringen' als Funktionsverben*. Düsseldorf: Pädagogischer Verlag Schwann.

Ibrahim, Amr Helmy. 2005. Light verbs in standard and Egyptian Arabic. In Mohammad Alhawary and Elabbas Benmamoun (eds.), *Perspectives on Arabic Linguistics XVII-XVIII: Papers from the seventeenth and eighteenth annual symposia on Arabic linguistics*, 117–131. Amsterdam: John Benjamins.

Isoda, Michio. 1991. The light verb construction in Japanese. In *CLS. Papers from the General Session at the 27th Regional Meeting*, vol. 27 1, 261–275.

Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles, Part VI, Morphology*. Copenhagen: Ejnar Munksgaard.

Ježek, Elisabetta. 2016. *The Lexicon. An introduction*. Oxford: Oxford University Press.

Ježek, Elisabetta. 2023. Semantic Co-composition in Light Verb Constructions. In Anna Pompei, Lunella Mereu and Valentina Piunno (eds.), *Light Verb Constructions as Complex Verbs: Features, Typology and Function*, 221–238. Berlin: De Gruyter.

Karimi-Doostan, Gholamhossein. 2011. Separability of light verb constructions in Persian. *Studia Linguistica* 65(1). 70–95. https://doi.org/10.1111/j.1467-9582.2011.01178.x.

Kettnerová, Václava. 2023. Valency structure of complex predicates with Light Verbs. The case of Czech. In Anna Pompei, Lunella Mereu and Valentina Piunno (eds.), *Light Verb Constructions as Complex Verbs*, 19–44. Berlin: De Gruyter. https://doi.org/10.1515/9783110747997-002.

Kishimoto, Hideki. 2025. On the light verb construction in Japanese. In Anna Riccio and Jens Fleischhauer (eds.), *Light verbs: Synchronic and diachronic studies*, 233–276. Düsseldorf: Düsseldorf University Press. https://doi.org/10.1515/9783111388878-009.

Kuo, Pei-Jung. 2025. The GIVE group of the Mandarin light verbs. In Anna Riccio and Jens Fleischhauer (eds.), *Light verbs: Synchronic and diachronic studies*, 43–64. Düsseldorf: Düsseldorf University Press. https://doi.org/10.1515/9783111388878-003.

Lai, Ryan Ka Yau. this volume. Beyond bidirectional association: Distinguishing light verb constructions from other conventionalised noun-verb combinations in modern Tibetan Beyond directional association. In Jens Fleischhauer and Anna Riccio (eds.), *Light verb constructions: Complex predicates from a cross-linguistic perspective*, 117–145. Berlin: De Gruyter.

Langer, Stefan. 2004. A linguistic test battery for support verb constructions. *Lingvisticae Investigationes* 27(2). 171–184.

Langer, Stefan. 2005. A Formal Specification of Support Verb Constructions. In Stefan Langer and Daniel Schnorbusch (eds.), *Semantik im Lexikon*, 179–202. Tübingen: Narr.

Lee, Juwon. 2011. Two Types of Korean Light Verb Constructions in a Types Feature Structure Grammar. In Valia Kordoni, Carlos Ramisch and Aline Villavicencio (eds.), *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World (MWE 2011)*, 40–48. Portland: Association for Computational Linguistics.

Leibniz-Institut für Deutsche Sprache. 2021. *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2021-I (Release vom 02.02.2021)*. Mannheim: Leibniz-Institut für Deutsche Sprache. https://cosmas2.ids-mannheim.de/cosmas2-web/.

Mahmoodi-Bakhtiari, Behrooz. 2018. Morphology. In Anousha Sedighi and Pouneh Shabani-Jadidi (eds.), *The Oxford Handbook of Persian Linguistics*, 273–299. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198736745.013.10.

Maisak, Timur. this volume. Complex verbs in Kina Rutul (Nakh-Daghestanian). In Jens Fleischhauer (ed.), *Light verb constructions: Complex predicates from a cross-linguistic perspective*, 147–174. Berlin: De Gruyter.

Masini, Francesca. 2021. Multi-Word Expressions and Morphology. In Rochelle Lieber (ed.), *The Oxford Encyclopedia of Morphology*, 1597–1619. Oxford: Oxford University Press.

Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language & Linguistic Theory* 19. 153–197. https://doi.org/10.1023/A:1006465130442.

Matsumoto, Yo. 1996. *Complex Predicates in Japanese: A Syntactic and Semantic Study of the Notion 'word'*. Tokyo & Stanford: Kurosio Publishers & CSLI Publications.

Mehl, Seth. 2017. Light verb constructions in the *International Corpus of English*: onomasiological variation, identity evidence and degrees of lightness. *English Language and Linguistics* 23(1). 55–80. https://doi.org/10.1017/S1360674317000302.

Mel'čuk, Igor. 2022. Support (= Light) Verbs. *Neophilologica* 34. 1–30. https://doi.org/10.31261/NEO.2022.34.03.

Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4). 847–894. 10.1353/lan.1984.0038.

Mithun, Marianne. 2010. Constraints on compounds and incorporation. In Irene Vogel and Sergio Scalise (eds.), *Compounding*, 37–56. Amsterdam: John Benjamins.

Mohanan, Tara. 2006. Grammatical Verbs (with Special Reference to Light Verbs). In Martin Everaert and Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax, Volume I*, 459–492. Malden: Blackwell.

Müller, Stefan. 2010. Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* 46(2). 601–665. https://doi.org/10.1017/S0022226709990284.

Narrog, Heiko and Bernd Heine. 2021. *Grammaticalization*. Oxford: Oxford University Press.

Newman, John. 1996. *Give: A cognitive linguistic study*. Berlin: De Gruyter Mouton.

Nicoletti, David. 2025. Constructional semantics of Italian light verb constructions. In Anna Riccio and Jens Fleischhauer (eds.), *Light verbs: Synchronic and diachronic studies*, 199–232. Düsseldorf: Düsseldorf University Press. https://doi.org/10.1515/9783111388878-008.

Nolan, Brian. 2014. Complex predicates and light verb constructions in Modern Irish. *Spanish Journal of Applied Linguistics* 27(1). 140–167. https://doi.org/10.1075/resla.27.1.06nol?locatt=mode:legacy.

Nugraha, Danang Satria. 2024. Morphosemantic features of *memenuhi* 'meet' in the light verb constructions of Indonesian. *Linguistik Indonesia* 42(2). 461–480. https://doi.org/10.26499/li.v42i2.638.

Nunberg, Geoffrey, Ivan A. Sag and Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538. https://doi.org/10.1353/lan.1994.0023.

Olejarnik, Magdalena. 2009. *Complex Predicates in Swahili: An LFG Approach*. Doctoral dissertation: Universytet Warszawski.

Olejarnik, Magdalena. 2011. An LFG Approach to Complex Predicates in Swahili. In Piotr Banski, Beata Lukaszewicz, Monika Opalinska and Joanna Zaleska (eds.), *Generative Investigations: Syntax, Morphology, and Phonology*, 139–165. Newcastle upon Tyne: Cambridge Scholars Publishing.

Ourang, Muhammed, Mengistu Amberber and Efrosini Deligianni. this volume. The derivation of the light verb construction in Aheli. In Jens Fleischhauer and Anna Riccio (eds.), *Light verb constructions: Complex predicates from a cross-linguistic perspective*, 63–85. Berlin: De Gruyter.

Pisciotta, Flavio and Francesca Masini. 2025. A paradigm of psych-predicates. In Anna Riccio and Jens Fleischhauer (eds.), *Light verbs: Synchronic and diachronic studies*, 123–176. Düsseldorf: Düsseldorf University Press. https://doi.org/10.1515/9783111388878-006.

Plante, Sebastien. 2014. Types of nominalized verbs in light verb constructions. *McGill Working Papers in Linguistics* 24(1). 81–91.

Pompei, Anna. 2023. How light is 'give' as a Light Verb? A case study on the actionality of Latin Light Verb Constructions (with some references to Romance languages). In Anna Pompei, L. Mereu and V. Piunno (eds.), *Light Verb Constructions as Complex Verbs*, 149–200. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110747997-006.

Pompei, Anna, Lunella Mereu and Valentina Piunno. 2023. Introduction. In Anna Pompei, L. Mereu and V. Piunno (eds.), *Light Verb Constructions as Complex Verbs*, 1–18. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110747997-001.

Pompei, Anna and Valentina Piunno. 2023. Light Verb Constructions in Romance languages. An attempt to explain systematic irregularity. In Anna Pompei, L. Mereu and V. Piunno (eds.), *Light Verb Constructions as Complex Verbs*, 99–147. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110747997-005.

Pustejovski, James. 1991. The Generative Lexicon. *Computational linguistics* 17. 409–441.

Pustejovski, James. 1995. *The Generative Lexicon*. Boston: The MIT Press.

Pustejovsky, James and Olga Batiukova. 2016. *The lexicon*. Cambridge: Cambridge University Press.

Riccio, Anna. this volume. Italian light verb constructions and definiteness. In Jens Fleischhauer and Anna Riccio (eds.), *Light verb constructions: Complex predicates from a cross-linguistic perspective*, 175–203. Berlin: De Gruyter.

Ronan, Patricia and Gerold Schneider. 2015. Determining light verb constructions in contemporary British and Irish English. *International Journal of Corpus Linguistics* 20(3). 326–354. https://doi.org/10.1075/ijcl.20.3.03ron.

Samvelian, Pollet and Pegah Faghiri. 2013a. Introducing PersPred, a Syntactic and Semantic Database for Persian Complex Predicates. In Valia Kordoni, Carlos Ramisch and Aline Villavicencio (eds.), *Proceedings of the 9th Workshop on Multiword Expressions (MWE 2013)*, 11–20. Atlanta: The Association for Computational Linguistics.

Samvelian, Pollet and Pegah Faghiri. 2013b. Re-thinking Compositionality in Persian Complex Predicates. *Berkeley Linguistics Society* 39. 212–226. https://shs.hal.science/halshs-01439564v1, last access: 13/01/2025.

Samvelian, Pollet and Pegah Faghiri. 2014. Persian Complex Predicates: How Compositional Are They? *Semantics-Syntax interface* 1(1). 43–74.

Sanromán Vilas, Begoña. 2011. The unbearable lightness of light verbs. Are they semantically empty verbs? In Igor Boguslavsky and Leo Wanner (eds.), *Proceedings of the 5th International Conference on Meaning-Text Theory*, 253–262.

Schultze-Berndt, Eva. 2006. Taking a closer look at function verbs: Lexicon, grammar, or both. In Felix K. Ameka, Alan Dench and Nicholas Evans (eds.), *Catching language: the standing challenge of grammar writing*, 359–391. Berlin: De Gruyter Mouton.

Shahrokny-Prehn, Arian and Silke Höche. 2011. Rising through the registers – A corpus-based account of the stylistic constraints on Light Verb Constructions. *Corpus* 10. 239–257. https://doi.org/10.4000/corpus.2110.

Staudinger, Eva. 2018. French and Spanish 'MAKE/GIVE + FEAR"'-type LVCs – an RRG Constructional Account. In Rolf Kailuweit, Lisann Kunkel and Eva Staudinger (eds.), *Applying and expanding Role and Reference Grammar*, 237–261. Freiburg: FRIAS.

Stojanovska-Ilievska, Natasha. 2025. A corpus-based study of the English light verb constructions with *give* denoting bodily actions and physical interaction. In Anna Riccio and Jens Fleischhauer (eds.), *Light verbs: Synchronic and diachronic studies*, 15–42. Düsseldorf: Düsseldorf University Press. https://doi.org/10.1515/9783111388878-002.

Sultan, Abbas. 2011. An Account of Light Verb Constructions in Shabaki. *Acta Linguistica* 5(2). 36–56.

Sundquist, John. 2025. Taking a closer look at light verb constructions in the history of American English. In Anna Riccio and Jens Fleischhauer (eds.), *Light verbs: Synchronic and diachronic studies*, 177–198. Düsseldorf: Düsseldorf University Press. https://doi.org/10.1515/9783111388878-007.

Sundquist, John D. 2018. A Diachronic Analysis of Light Verb Constructions in Old Swedish. *Journal of Germanic Linguistics* 30(3). 260–306. https://doi.org/10.1017/S1470542717000137.

Vahedi-Langrudi, Mohammad-Mehdi. 1996. *The syntax, semantics and argument structure of complex predicates in modern Farsi*. Doctoral dissertation: University of Ottawa.

Vincze, Veronika, István Nagy and János Zsibrita. 2013. Learning to detect English and Hungarian light verb constructions. *ACM Transactions on Speech and Language Processing (TSLP)* 10(2). 1–25. https://doi.org/10.1145/2483691.2483695.

von Polenz, Peter. 1963. *Funktionsverben im heutigen Deutsch. Sprache in der rationalisierten Welt*. Düsseldorf: Schwann.

von Polenz, Peter. 1987. Funktionsverben, Funktionsgefüge und Verwandtes. Vorschläge zur satzsemantischen Lexikographie. *Zeitschrift für germanistische Linguistik* 15. 169–189. https://doi.org/10.1515/zfgl.1987.15.2.169.

Wierzbicka, Anna. 1982. Why can you have a drink when you can't *have an eat? *Language* 58(4). 753–799. https://doi.org/10.2307/413956.

Wiskandt, Niklas. 2025. Light verb constructions with experiencer objects in Germanic and Romance. In Anna Riccio and Jens Fleischhauer (eds.), *Light verbs: Synchronic and diachronic studies*, 97–122. Düsseldorf: Düsseldorf University Press. https://doi.org/10.1515/9783111388878-003.

Wittenberg, Eva. 2016. *With Light Verb Constructions from Syntax to Concepts*. Potsdam: Universitätsverlag Potsdam.

Wittenberg, Eva, Ray Jackendoff, Gina Kuperberg, Martin Paczynski, Jesse Snedeker and Heike Wiese. 2014a. The processing and representation of light verb constructions. In Asaf Bachrach, Isabelle Roy and Linnaea Stockall (eds.), *Structuring the Argument. Multidisciplinary research on verb argument structure* , 61–80. Amsterdam: John Benjamins. https://doi.org/10.1075/lfab.10.

Wittenberg, Eva, Martin Paczynski, Heike Wiese, Ray Jackendoff and Gina Kuperberg. 2014b. The difference between "giving a rose"and "giving a kiss"': Sustained neural activity to the light verb construction. *Journal of Memory and Language* 73. 31–42. https://doi.org/10.1016/j.jml.2014.02.002.

Avelino Corral Esteban

Light verb constructions in Scottish Gaelic

1 Introduction

The study of complex predicates (Butt, 1995; Alsina et al., 2001; Amberber et al., 2010; among others) and multi-verb constructions (Aikhenvald and Muysken, 2011; among others) has attracted the interest of scholars worldwide. Both descriptive and theoretical approaches to their exploration and analysis in many languages appear to agree on a number of shared properties. An example is their monoclausal structure, which is reflected in the fact that the intonational properties of the construction are the same as those of a monoverbal construction. Likewise, they have a single tense, aspect, modality, mood, evidentiality and polarity value, but their argument structure is essence quite complex as it consists of two heads that are capable of contributing semantically to the event, and there is also a single complex predicate and a single subject.

Despite the widespread accordance with the properties attributed to constructions involving complex predicates, their cross-linguistic analysis has traditionally been rather tricky, mainly due to the difficulty in establishing a clear-cut criteria that enable the identification of different types of complex predicates. The lack of clear classification criteria comes from the fact that some characteristics seem to be too difficult to apply, for example whether or not a verb contributes to the semantic content of the construction. Others may refer to morphosyntactic properties that are language-specific, for example, whether it is possible to differentiate between auxiliary and lexical verbs in English in terms of inverting their order with respect to the subject in interrogative sentences. Thus, although most studies on multi-verb constructions have drawn on a tripartite classification consisting of the types commonly referred to as serial verb constructions (SVC), the light verb construction (LVC) and the auxiliary verb construction (AVC), the specific study of these three classes of complex predicates in different languages has shown great heterogeneity of properties in each construction. This makes it hard to attain a precise definition and an accurate demarcation of each type. Hence, the interest in setting up distinguishing properties and finding a clear definition for the three main complex predicate types remain relevant today. In view of this situation and that the focus of this paper is on the study

Avelino Corral Esteban, Universidad Autónoma de Madrid, Spain

https://doi.org/10.1515/9783112229330-002

and analysis of LVCs in Scottish Gaelic (Goidelic Celtic), the organization of this paper is as follows. Section 2 attempts to distinguish LVCs from other types of complex predicates, such as auxiliary verb constructions (AVC) and serial verb constructions (SVC), in Scottish Gaelic by examining the semantic and morphosyntactic properties of their verbs. Section 3 shows a cross-linguistics comparison of the characteristics of Scottish Gaelic LVCs with those that have traditionally been proposed in other languages in order to check which of the features of Scottish Gaelic LVCs are distinctive, and which ones are recurrent. Finally, the paper aims to provide a functional characterization of LVCs within the Role and Reference Grammar framework (Van Valin and LaPolla, 1997; Van Valin, 2005).

2 The identification of light verb constructions in Scottish Gaelic

As noted by Butt (2010, 49), an identification of the different types of construction covered by the concept of 'complex predication' is not an easy task. Firstly, an SVC (Aikhenvald and Dixon, 2006; Haspelmath, 2016) is generally considered to include a sequence of lexical verbs that are combined to behave like a single predicate, in such a way that each verb describes a different facet of what should be conceptualized as a single event. Neither do they contain any marker of coordination or subordination, they cannot be negated or questioned separately from the whole construction, and are under a single intonation contour. Other studies also allude to the possibility that serial verbs can occur on their own as the main verb of a clause in other contexts, span the entire verbal paradigm, have shared values for tense, aspect and modality, do not have separate temporal or locative modifiers, share at least one argument (generally corresponding to the subject, although object sharing can sometimes also occur), and may not be strictly adjacent (Baker, 1988; Stewart, 1963; Aikhenvald, 2006, 2018; Haspelmath, 2016).

These prototypical properties would indicate that there are no examples of SVCs in Scottish Gaelic. The closest equivalents to an SVC in this language could be subject and object control, periphrastic causative, purposive constructions, and constructions including posture verbs and direct perception, as their two verbs contribute semantic content to the construction. However, only one of the two verbs is inflected and the other must occur in the form of a non-finite hybrid of noun

and verb commonly referred to as a verbal noun in Celtic linguistics,[1] which could then be considered to be syntactically dependent on the other verb. Also, there is an intervening particle between the two predicates and they are generally pronounced with more than one intonation contour, which appears to suggest that they encompass two syntactically separate domains of predication.

Secondly, fewer difficulties are found when distinguishing an SVC from an AVC if one adopts the most common properties proposed for the AVC (Anderson, 2006, 2011). In this construction, the lexical component generally occurs in a non-finite and, consequently, a syntactically dependent form, and the functional component, which is generally referred to as the auxiliary verb, encodes the core inflectional verbal features or categories such as aspect, tense, modality, mood, evidentiality, causation, and voice. Also, except in cases where the full verb has been elided but is easily recoverable from context, the auxiliary verb cannot occur alone and it is not restricted in its combinatorial possibilities. Additionally, it can be combined with every lexical verb, even with its homonymous full verb, it is complemented by a non-finite full verb, and generally spans the entire verbal paradigm, although, as in the case of modal auxiliary verbs, it can also display a defective paradigm, since it is restricted to just one tense or aspectual form (Heine, 1993; Kuteva, 2001; Anderson, 2006, 2011; Krug, 2012). Finally, the auxiliary verb is considered not to be able to add or reduce arguments and determines case and theta-role assignment. The set of verbs that have most commonly become grammaticalized as auxiliaries in Scottish Gaelic include the following: *bi* 'be', *is* 'be', *faigh* 'get', *gabh* 'take', *rach* 'go', *thig* 'come', and *theob* 'do almost'.

With respect to the two copular verbs, this language uses the substantive *bi* 'be' to express a number of aspectual distinctions, such as progressive aspect, perfect aspect, prospective aspect, tense, indirect causation, and passive, and the assertive or copula *is* 'be' to express other different types of aspect, such as habitual aspect, and mood, for example to express a counterfactual situation or wish. Also, this language can also express the passive voice or impersonal sentences through other different periphrastic ways involving verbs like *faigh* 'get', *gabh* 'take', *rach* 'go', and *thig* 'come'. Additionally, Scottish Gaelic uses the defective verb *theab* 'almost, nearly', which is severely restricted formally speaking as it has only one form, to express the avertive aspect. Besides these primary auxiliary verbs, Scottish Gaelic also has a couple of modal auxiliary verbs, *faod* 'can' and *feum* 'must', which serve to convey different types of deontic, dynamic, and epistemic modality. Thus, on the one hand,

1 This element is a construction-specific non-finite form – commonly referred to as a 'verbal noun' in the traditional terminology used in Celtic linguistics (MacAuley, 1992, 170–190, Gillies, 1993, 282–285, Adger, 2010, 305–336, Cox, 2017, Byrne, 2018, 117–130; and Lamb, In press, 398–439) – that, as suggested by its name, combines both verbal and nominal properties.

faod 'can' can express permission, ability, and probability, and on the other hand, *feum* 'must' can express strong obligation, necessity, and certainty. Finally, Scottish Gaelic also makes use of the substantive *bi* 'be' and the assertive or copula *is* 'be' plus a number of nouns, adjectives and prepositions to convey a number of modal distinctions. Thus, the main differences between an LVC and an AVC are that, unlike auxiliary verbs, light verbs are considered to retain some semantic content, may be accompanied by other elements apart from verbs, such as nouns, adjectives, prepositions or even adverbs, are combinatorially restricted, and then are capable of adding or reducing arguments and determining case and theta-role assignment.

The line of demarcation between LVCs and the other two complex predicate types is especially hard to draw, as LVCs appear to lie between SVCs and AVCs in the sense that it is difficult to decide to what extent one of its components, namely the light verb, can contribute lexical semantic content – as serial verbs do – or grammatical information – which happens with auxiliary verbs – to the construction, or even what kind of information it is able to provide. This complicates the definition of what should count as an LVC, hence the term LVC has sometimes been used to refer to different types of complex predicate cross-linguistically. In view of this, a neutral stance on the definition of LVCs could be one that considers the light verb to be an inflectable semi-lexical verbal element that is combined restrictively with one or more full lexical heads which may belong to a heterogeneous set of categories (e.g. noun, adjective, an uninflectable verb, preposition, etc.) to jointly predicate a single event (Alsina et al., 2001; Butt and Geuder, 2001; Butt, 2010; Butt and Lahiri, 2002; Heine, 2006; Amberber et al., 2010; Fleischhauer et al., 2019; Fleischhauer and Gamerschlag, 2019; Fleischhauer, 2021).

The set of verbs that are normally used as light verbs in Scottish Gaelic is as follows: *gabh* 'take', *thoir* 'give', *dèan* 'do/ make', *cuir* 'put', *faigh* 'get', *thig* 'come', *rach* 'go', *leig* 'let', *tarraing* 'draw', *buail* 'hit', and *cùm* 'keep'. Light verbs in this language have the same form and inflect exactly like their main verb counterparts and span the entire verbal paradigm, with most of them also functioning as auxiliary verbs. They can be combined with a number of categories – especially nominal (e.g. *faigh fios* 'find out' (lit. get knowledge), *tarraing dealbh do* 'draw a picture of' (lit. draw picture to), *leig osna* 'sigh' (lit. let sigh), etc.), but also verbal (e. g. *thoir gealladh do* 'make a promise to' (lit. give promising to), *faigh àrdachadh* 'get a promotion' (lit. get raising), *thig co-dhùnadh* 'take a decision' (lit. come concluding), etc.), adjectival (e.g. *dèan deas* 'get ready/ prepare' (lit. do ready), *cuir suarach* 'despise' (lit. put worthless), *thig nas fhèarr* 'get better/ improve' (lit. come better), etc.), prepositional (e.g. *gabh os láimh* 'take charge of' (lit. take in hand), *cuir fo gheasaibh* 'put a spell on' (lit. put under spell), *rach air chèilidh* 'visit' (lit. go on visit), etc.), or even adverbial (e.g. *thoir fa-near* 'take into account' (lit. give under observation), *cuir suas* 'set up/

establish' (lit. put up), etc.) – to form a large number of examples of LVCs in Scottish Gaelic.

The verb that is more productive in this construction is *dèan* 'do/ make', as it can be combined with a higher number of lexical verbs in order to emphasize the event taking place and even to convey the idea of volition:[2]

(1) *Dèan cadal gu sàmhach, mo leanabh gràdhach*!
do.make.IMP sleep.VN ADV quiet 1SG.POSS baby dear
'Sleep quietly, my dear baby!'

Given that auxiliary verbs are generally accompanied by a lexical verb, the type of LVC that is more problematic when it comes to differentiating it from an AVC is, obviously, a construction consisting of a light verb and a lexical verb in the form of a verbal noun (e.g. *thoir luigheachd do* 'forgive' (lit. give reward to), *faigh stuigeadh* 'get urged on' (lit. get inciting), *thig co-dhúnadh* 'take a decision' (lit. come concluding), etc.).[3] While verbal nouns in Insular Celtic languages have always been intriguing in that they tend to display the characteristics of both nouns and verbs, at least specifically for this construction, the verbal noun appears to show some morphosyntactic properties that make it more similar to a noun, such as its capacity to accept adjectival modification (2) and possessive pronouns (3):[4]

(2) *Rinn i seinn àlainn an-dè.*
do.PST 3SG.F sing.VN beautiful yesterday
'She sang beautifully yesterday.' (lit. She did a beautiful singing yesterday.)

2 The corresponding sentence, without the special emphasis on the action denoted by the more lexical verb and on the property of volition, is as follows:

(i) *Cadal gu sàmhach mo leanabh gràdhach*!
sleep.IMP ADV quiet 1SG.POSS baby dear
'Sleep quietly, my dear baby!'

3 Despite this, Scottish Gaelic appears to show a higher number of deverbal nouns than verbal nouns as the second element of the LVC, which could possibly be due to the close similarity between the two forms – that is the verbal noun and the noun – and the fact that this element appears to have more nominal than verbal properties in this construction.

4 The reason why this construction should be considered as an example of an LVC rather than an AVC is related to its productivity, as the verbal noun accompanying the light verb *dèan* is not unrestricted. For example it is combined with the noun *obair* "work' rather than with the verbal noun *oibreachadh* to express 'work'. Anyway, now it is more common to express these meanings with a single lexical verb, as in *Sheinn i gu h-àlainn an-dè* 'She sang beautifully yesterday.' and *Mheall thu sinn* 'You cheated us.'

(3) *Rinn thu ar mealladh.*
do.PST 2SG 1PL.POSS cheat.VN
'You cheated us.' (lit. You did our cheating.)

While it is difficult to draw a definitive conclusion from this evidence, it seems plausible to assume that verbal nouns in this construction are more nominal than verbal, given that they occur in syntactic contexts associated with nouns and govern genitive objects.[5] This would therefore confirm that this construction including a light verb accompanied by a verbal noun is clearly an instance of an LVC, not of an AVC. However, this construction is not unproblematic since, while more than 60 years have passed since Jespersen (1942, 117) coined the term 'light verb' and it is generally assumed that the less verbal component in this construction functions as its main predicate, no agreement has yet been reached as to whether the light or support verb can contribute semantically to the complex predicate, what type of content it conveys, and whether the nominal element functions as a complement or not. These issues are therefore analyzed in the remainder of this section.

2.1 Meaning of the light verb

What differentiates a light verb from a full lexical verb or an auxiliary verb, respectively, is that it does not seem to have the same semantic content as its homologous lexical verb – the non-light or 'heavy' verb – , and that it is not completely devoid of meaning and expresses information that it is not purely grammatical like aspect, tense, mood, or modality, as is the case with an auxiliary verb. Although this issue

5 While the context in which the verbal noun appears in an LVC is clearly nominal, the assumption that verbal nouns are simply nominals is not so obvious. Verbal nouns also occur in verbal contexts, in which they seem to behave like verbs (Cram, 1981; Ramchand, 1993; Adger, 1996, 2010). Thus, for example, verbal nouns appear in constructions expressing aspectual and modal distinctions accompanying an auxiliary verb and a modal verb. Also, an object precedes the verbal noun in constructions expressing the perfect aspect, the prospective aspect, the ingressive aspect, and the egressive aspect, or containing a modal expression denoting willingness, wish, and expectation, which seems to have no grammatical analogue in nominal structures since it is not possible for an argument of a simple noun, or its possessor, to occur in pre-nominal position. Besides, the object of a verbal noun can be fronted in a cleft construction, which is not possible for genitive objects in complex nominal groups. Finally, in a construction expressing progressive aspect, the verbal noun is preceded by the grammaticalized form of a preposition that has now become an aspectual particle. In view of this, one might say that the verbal noun, albeit a noun in origin, appears to have gradually developed two distinct functions owing to a tendency for verbalisation and, consequently, now seem to behave like a noun or like a verb in different syntactic environments.

is difficult to prove, this section attempts to shed light on it in order to confirm this assumption.

Following studies like Kearns (2002), an argument for the assumption that the light verb contributes little or no meaning to the complex predicate tends to be reflected in the fact that a fair number of LVCs in Scottish Gaelic are semantically equivalent to a simple lexical verb, which is, in turn, either identical in form to the eventive noun (e.g. *dèan adhradh do* 'worship' (lit. do worshiping to) = *adhradh* 'worship'; *gabh smoc* 'have a smoke' (lit. take smoke) = *smoc* 'smoke'; *leig sgread às* 'let out a scream' (lit. let scream out) = *sgread* 'scream'; etc.), or lexically related to the nominal element (e.g. *cuir iarrtas* 'make a request' (lit. put request) = *iarr* 'request'; *thig co-dhùnadh* 'take a decision' (lit. come concluding) = *co-dhùin* 'decide'; *thoir comhairle do* 'give advice to' (lit. give advice to) = *comhairlich* 'advise'; etc.) and can, consequently, be replaced by a simple verb without altering the meaning. However, there are also many other LVCs in this language whose meaning can only be expressed through unrelated simple verbs (e.g. *cuir fios gu* 'inform' (lit. put knowledge for) = *innis* 'tell/ inform'; *dèan ùrnaigh ri* 'pray' (lit. do prayer to) = *guidh* 'pray'; *faigh cron do* 'put a blame on' (lit. get fault to) = *coirich* 'blame'; *gabh fois* 'take a rest' (lit. take relaxation) = *tàmhaich* 'rest'; *rach air chèilidh* 'visit' (lit. go on visit) = *tadhail* 'visit'; *thoir sgoil do* 'educate' (lit. give school to) = *oileanaich* 'educate'; etc.).

Another reason for claiming that light verbs appear to have lost some of their denotational meaning in this construction is that some of them can be combined with the same lexical element without any difference in meaning or, more probably, with a subtle difference in meaning (e.g. *cùm taic do* (lit. keep support to)/ *thoir taic do* (lit. give support to) 'give support to'; *dèan dealbh de* (lit. do picture of)/ *tarraing dealbh de* (lit. draw picture of) 'take a picture of'; *dèan greim air* (lit. do grasp on)/ *gabh greim air* (lit. take grasp on)/ *thoir greim air* (lit. give grasp on) 'take hold of'; *dèan ionnsaigh air* (lit. do attack on)/ *thoir ionnsaigh air* (lit. give attack on) 'attack'; *dèan norrag* (lit. do nap)/ *gabh norrag* (lit. take nap) 'take a nap'; *faigh comhairle* (lit. get advice)/ *gabh comhairle* (lit. take advice) 'take advice'; *gabh d'anail* (lit. take your breath)/ *leig d'anail* (lit. let your breath)/ *tarraing anail* (lit. draw breath) 'take a break'; etc.). However, in many other examples of LVCs the replacement of a specific light verb by another light verb leads to a change in meaning (e.g. *faigh cead* 'get permission' vs. *thoir cead do* 'give permission'; *faigh teic* 'receive support' (lit. get support) vs. *thoir teic do* 'give support to'; *gabh comhairle* 'take advice' vs. *thoir comhairle do* 'give advice to'; *faigh meas* 'get respect' vs. *thoir meas do* 'respect'; etc.), which proves that the light verb is not entirely devoid of semantic content – a difference in terms of causation between the two LVCs is evident – since, if the light verb in each pair did not contribute to the meaning of the LVC, the two complex predicates should have the same interpretation.

The argument structure of the whole construction is also generally considered to be determined by the nominal element rather than the light verb. This can be seen in examples like *thoir ceum* 'take a step' (lit. give step), *thoir gèill* 'surrender' (lit. give submission), *thoir sgread* 'scream' (lit. give scream), *thoir mionnan* 'swear' (lit. give oaths), and *thoir stiùir* 'give a lead', since, although the light verb *thoir* 'give' is a three-argument verb, it takes a single NP object, which is related to the fact that the accompanying nouns are eventive predicates that lack a logical object. However, these are exceptional cases, and the influence that the nominal predicate has on the argument structure of the construction does not necessarily mean that the light verb does not have a semantic role to play, and it could be possible for a second NP to be added in these constructions (e.g. *thoir ceum a dh'ionnsaigh cuideigin* 'take a step towards someone' (lit. give step towards someone), *thoir gèill do chuideigin* 'surrender before someone' (lit. give submission to someone), *thoir sgread do chuideigin* 'give a scream to someone' (lit. give scream to someone), *thoir mionnan air cuideigin* 'swear to someone' (lit. give oaths to someone), *thoir stiùir do chuideigin* 'give a lead to someone' (lit. give lead to someone), etc.), so it could be understood as having been left implied in the examples above.

Further evidence for the semantic contribution of the light verb can be observed in examples like *gabh sràid* 'take a walk', since, although the action denoted by the eventive noun *sràid* 'walk' is realized by both people and animals, it would sound odd if we were to say that an animal could carry out the action expressed by the whole complex predicate. This difference presumably reflects the influence of the argument structure of the light verb *gabh* 'take' on the interpretation of the complex predicate.

Additionally, there are also light verbs that occur with different but semantically related lexical elements to express similar concepts (e.g. *faigh fàth de* (lit. get view of)/ *faigh plathadh de* (lit. get glance of) 'catch a glimpse of', *thoir sùil air* give look on/ *thoir balladh air* give glimpse on 'take a look at', *gabh dèidh de* (lit. take fondness of)/ *gabh nòisean de* (lit. take fancy of)/ *gabh tlachd ann* (lit. take delight in) 'take a liking to', *thoir* (*an*) *aire air* (lit. give (the) attention on)/ *tarraing aire air* (lit. draw attention on)/ *thoir feart air* (lit. give notice on)/ *thoir toghaidh do* (lit. give care to) 'pay attention to', *dèan togail ri* (lit. do lifting on)/ *dèan fiughair ri* (lit. do hope to) 'look forward to', *dèan iochd ri* (lit. do compassion to)/ *dèan trocair air* (lit. do mercy on) 'take pity on', *thoir meas do* (lit. give regard to)/ *thoir spéis do* (lit. give affection to)/ *thoir urram do* (lit. give respect to) 'respect', etc.), which suggests that the meaning of the nominal element should somehow fit into the properties of the light verb in order to jointly determine the argument structure of the complex predicate.

Finally, following Nunberg et al. (1994, 499–503), further evidence for the assumed compositionality of the meaning expressed by LVCs can be provided by the existence of systematic interpretational patterns involving the semantic contribu-

tion of the light verb and the nominal element, which are illustrated by families of LVCs such as the following: 1) grant something to someone (e.g. *thoir cead do* (lit. give permission to) 'permit', *thoir comhairle do* (lit. give advice to) 'advise', *thoir luigheachd do* (lit. give reward to) 'reward', *thoir mathanas do* (lit. give forgiveness to) 'forgive', *thoir meas*/ *spèis*/ *urram do* (lit. give regard/ affection/ respect to) 'pay respect to', *thoir misneachadh do* (lit. give encouragement to) 'encourage', *thoir sgoil do* (lit. give school to) 'educate', etc.); 2) perform a physical or mental activity (e.g. *dèan aithreachas* (lit. do penitence) 'repent', *dèan breug* (lit. do lie) 'tell a lie', *dèan cadal* (lit. do sleep) 'sleep', *dèan casad* (lit. do cough) 'cough', *dèan còmhnaidh* (lit. do dwelling) 'dwell', *dèan obair* (lit. do work) 'work', *dèan òran* (lit. do song) 'compose a song', etc.); 3) be passively involved in an event (e.g. *faigh àrach* (lit. get raising) 'be brought up', *faigh àrdachadh* (lit. get raising) 'get a promotion', *faigh cead* 'get permission', *faigh comhairle* 'get advice', *faigh prìomhachas* (lit. get priority) 'receive priority', *faigh teic* (lit. get support) 'receive support', *faigh ranuns*/ *raphuins* 'get a telling-off', etc); 4) be in the process of undergoing a change of state expressed by the NP (e.g. *rach air chall* (lit. go on losing) 'get lost', *rach am feabhas* (lit. go the goodness) 'get better/ improve', *rach am fianais* (lit. go the evidence) 'appear/ come into view', *rach am follais* (lit. go the clearness) 'come to light', *rach am mearachd* (lit. go the error) 'become mistaken', *rach às do bheachd* (lit. go out of your judgment) 'go crazy', etc.).

As a conclusion, what precisely the light verb contributes to the joint predication, and therefore exactly which parts of the predication are supposed to have been lost, is difficult to determine. Apart from a special emphasis on the action denoted by the more lexical element and on volition – as illustrated in the LVCs including *dèan* 'do/ make' plus a verbal noun in (1) – a tendency to highlight the specificity, telicity, and duration of the event can also be observed in the LVC in comparison with sentences including the corresponding main verb. Thus, on the one hand, while the use of a simple verb like *pòg* 'kiss' would refer to a generic action, we can make this information more specific by using an LVC (e.g. *Thug mi pòg bheag dhi* 'I gave a little kiss to her' (lit. gave I kiss little to her), *Thug mi pòg thana dhi* 'I gave a tender kiss to her' (lit. gave I kiss tender to her), etc.), and, on the other hand, although using a simple verb like *sgread* 'scream' can be potentially unbounded (e.g. *Sgread i airson còig mionaidean* 'She screamed for five minutes' (lit. screamed I for five minutes)), the use of an LVC can express a bounded and probably shorter event more explicitly (e.g. *Leig i sgread às an-dràsta* 'She let out a scream right now' (lit. let she scream out right now)).

Finally, apart from providing different semantic considerations, another motivation for the creation of LVCs in a language could be to fill a gap in its lexical inventory. Thus, in Scottish Gaelic many LVCs express meanings for which the language lacks a simple verb (e.g. *cuir ann an cunnart* 'put at risk' (lit. put in the risk), *dèan breag*

'tell a lie' (lit. do lie), *dèan fabhar ri* 'do a favour to', *dèan norrag* 'take a nap' (lit. do nap), *dèan stad* 'take a break' (lit. do stop), *faigh àrdachadh* 'get a promotion' (lit. get raising), *faigh prìomhachas* 'receive priority' (lit. get priority), *gabh amar* 'take a bath', *gabh cùram* 'care for' (lit. take care), *gabh stròc* 'suffer a stroke' (lit. take stroke), *leig dub le* 'give a break to' (lit. let dip/ nibble with), *leig mùin* 'urinate' (lit. let urination), *rach às do bheachd* 'go crazy' (lit. go out your judgment), *thig am feabhas* 'get better/ improve' (lit. come the goodness), *thoir meas do* 'respect' (lit. give regard to), *thoir taing do* 'thank' (lit. give gratitude to), etc.), which means that LVCs may serve to compensate for the lack of some full verbs and fill a lexical gap.

2.2 Syntactic status of the nominal element

Another aspect of LVCs that has been widely discussed concerns the syntactic relationship between the light verb and the nominal component. The remainder of this section will therefore include an analysis of LVCs in Scottish Gaelic consisting in the application of a series of morphosyntactic tests to a number of instances of LVCs in an attempt to gauge whether the noun accompanying the light verb can be considered a complement of the light verb or not. As studies on LVCs such as Kearns (2002) and Bruening (2015) show, the syntactic relationship between the light verb and the nominal component of the LVC can be analyzed through a series of morphosyntactic tests that examine the referentiality and objecthood of the nominal element.

One of these tests involves passivization, since it is generally assumed that objects are passivized with respect to their corresponding verb. Examples (4) show that some LVCs allow for their nominal element to be passivized, while others appear to resist it:

(4) a. *Chaidh sùil a thoirt air an leabhar.*
go.PST look PART give.VN on the book
'A look was taken at the book.'

b. **Chaidh pàirt a gabhail anns á choinneamh.*
go.PST part PART take.VN in the meeting
'*Part was taken in the meeting.'

Another property that is considered to be more typical of objects than adjuncts is their greater tendency to accept relativization. However, as (5) shows, it is possible to find LVCs where the combination of the light verb and the nominal element can form a relative clause and others whose nominal element cannot be relativized:

(5) a. *Bha feum air a' chead a fhuair mi.*
be.PST need on the permission REL get.PST 1SG
'The permission that I got was necessary.'
b. **Bha an teine a chuir iad air an taigh uabhannach.*
be.PST the fire REL put.PST 3PL on the house terrible
'*The fire they set to the house was terrible.'

LVCs also differ in terms of question formation, which raises the possibility for the deverbal noun to be part of a 'wh'-phrase, as is illustrated by examples (6):

(6) a. *Dè am prìomhachas a fhuair thu*?
what the priority REL get.PST.DEP 2SG
'What priority did you get?'
b. **Dè an t-anail a tharraing thu an-dè*?
what the breath REL draw.PST.DEP 2SG yesterday
'*What breath did you draw yesterday?'

Objects are very commonly pronominalized, so this should in principle not be a property of deverbal nouns in LVCs if they are not the actual objects of the light verb. However, as the following example illustrates, the eventive noun in some but not in all LVCs can be replaced with a pronoun:

(7) a. A: *An do rinn thu mearachd san eacarsaich*?
INT.AFF PFV do.PST.DEP 2SG mistake in.the exercise
'Did you make a mistake in the exercise?'
B: *Cha do rinn. Rinn mi i anns an deuchainn.*
NEG PFV do.PST.DEP do.PST 1SG 3SG.F in.the the exam
'No, I didn't. I made it in the exam.'
b. A: *An do ghabh thu cùram de do shean-phàrantan*?
INT.AFF PFV take.PST.DEP.PFV 2SG care of 2SG.POSS grandparent.PL
'Did you take care of your grandparents?'
B: **Chan do ghabh. Ghabh mi e de m´ uncail agus*
NEG PFV take.PST.DEP take.PST 1SG 3SG.M of 1SG.POSS uncle and
piuthar-athar.
aunt
'No, I didn't. *I took it of my uncle and aunt.'

A deverbal noun in an LVC is generally considered to be non-referential (and generally) indefinite, invariable in number, and even unable to accept adjectival modification. Again, as the following examples illustrate, it is possible to find LVCs behaving differently with respect to these criteria:

(8) a. *Thug mi (an) aire don phàipear-naidheachd.*
give.PST 1SG the attention to.the newspaper
'I paid attention to the newspaper.'
b. **Ghabh mi an nòisean dhith.*
take.PST 1SG the fondness of.3SG.F
'I took a liking to her.' (lit. 'I took the liking to her.')

(9) a. *Thug e mòran phògan dhi.*
give.PST 3SG.M many kiss.PL to.3SG.F
'He gave a lot of kisses to her.'
b. **Ghabh iad ùidhean anns an leabhar.*
take.PST 3PL interest.PL in the book
'*They took interests in the book.'

(10) a. *Fhuair mi deagh chomhairle bhuaithe.*
get.PST 1SG good advice from.3SG.M
'I got good advice from him.'
b. **Chuir mi an aithne luath a chèile iad.*
put.PST 1SG the acquaintance quick to other 3PL
'I introduced them to each other quickly.' (lit. 'I put the quick introduction to each other.')

It is also generally assumed that objects, like subjects and unlike adjuncts, are always obligatory (unless we are dealing with certain contexts in which they can be easily retrieved) and cannot therefore be omitted or, at least, may not accept omission as freely as adjuncts, which are both semantically and syntactically optional which does not always occur in all LVCs:

(11) a. **Thug mi duais do Pheadair ach cha tug do dh'Iain.*
give.PST 1SG reward to Peter.DAT but NEG give.PST to John.DAT
'I gave a reward to Peter but I didn't give to John.'
b. *Cha tàinig e nas fhèarr an-dè ach thig e a-màireach.*
NEG come.PST 3SG.M COMP better yesterday but come.FUT 3SG.M tomorrow
'He didn't get better yesterday but he will get tomorrow.'

Finally, while this is not a distinctive property of objects and may merely serve to highlight a closer relationship with the verb, which does distinguish objects from adjuncts, the eventive noun in an LVC also seems to be restricted regarding its use elsewhere other than in this construction. This is the situation that is commonly observed in some LVCs, such as *gaolagan* 'hug, cuddle, lovey' in *thoir gaolagan do* 'give a hug to' or *seimhig* 'laughing stock' in *dèan seimhig de* 'make a fool of', for

example, as the nouns *gaolagan* and *seimhig* are not commonly used in constructions other than these LVCs.

These tests show that the relation between the verb and the noun is not clearly one between predicate and complement – which occurs in free verb-noun combinations where the elements can continue to be combined freely with each other and with others in expressions showing no semantic idiomaticity – , as the noun does not always have the typical properties of an object. This evidence might plausibly show that LVCs, or light verbs more specifically, cannot be understood as a homogenous class. We might therefore suppose that there are in fact two – as suggested by Kearns (2002) – or even more types of LVC, as demonstrated by the lack of homogeneity in this construction. This is reflected in the fact that the light verb shows a higher degree of desemanticization in some cases than in others, and the deverbal noun seems to function as an object of a verb in some examples but not in others. This should be connected to a greater or lesser extent with grammaticalization, especially if it is assumed that there are light verbs between main verbs and auxiliary verbs in the grammaticalization cline (Hopper and Traugott, 1993, 4–7; Roberts and Roussou, 2003, 200–205). This view should not be incompatible with Butt's[6] (2003; 2010) who considers main verbs and light verbs – verbs with generic semantic specifications and flexible lexical entries that allow them to modulate the main event semantics – as two alternatives that can be used in two different contexts. This is because different LVCs containing the same light verb have different syntactic and semantic properties, which stresses the importance of interpreting the semantic contribution of both the light verb and deverbal noun jointly, so that, depending on the specific LVC in which it occurs, the same light verb may be found at different points – and showing different degrees of the inevitable process of semantic bleaching – within that intermediate stage on the scale. This would also account for the fact that a light verb is not formally divergent.

Consequently, while some LVCs may be similar to lexicalized units, which behave like lexical verbs, as their elements appear to lack autonomy and form an indissoluble unit showing certain restrictions and with a degree of semantic idiomaticity,[7] others may be seen as free combinations of a verb and a less verbal element that show no semantic idiomaticity. In addition to this, given the heterogeneity of properties that LVCs show, it would seem logical that between these two options there exists a group of expressions whose constituents show different degrees of cohesion since the verb may be more or less desemanticized and its collocational partner – generally a

6 It must be taken into account, however, that Butt and Lahiri (1998) argue that light verbs have been attested to for thousands of years, which appears to call into question the assumption that light verbs have arisen from gradual semantic bleaching.

7 As it occurs in idiomatic expressions like *caill an deò* 'die' (lit. lose the air).

deverbal noun having an inherent eventive value – appears to keep all or most of its denotational meaning.

3 Analysis of Scottish Gaelic LVCs within RRG

In view of this evidence, it seems clear that, seen from a global perspective, a light verb neither retains all its predicational content, nor is it semantically empty, and that the nominal element provides most but not all the semantic content of the construction. It would therefore follow that the light verb is generally part of a joint predication within a complex predicate and does contribute to it, as both the light verb and the more lexical element jointly determine argument structure. The real issue is, however, to determine accurately how much each of these two elements contribute and how the lexical composition takes place and should be represented. The first part of this section thus offers an analysis of the semantic composition of LVCs in Scottish Gaelic, by examining the relationship between the argument structures of the light verb and of the accompanying nominal element, which is characteristically an eventive noun, within the theory of lexical and semantic representation proposed by the RRG framework, which is based on the concepts of Aktionsart and logical structure (Foley and Van Valin, 1984, 28–74, Van Valin, 2005, 31–50).

3.1 The composition of LVCs

The current RRG typology of Aktionsart distinguishes between the following classes: states, activities, achievements, accomplishments, semelfactives, active accomplishments, and the causative versions of all Aktionsart classes (Van Valin, 2005, 32–39). All of these Aktionsart classes are defined in terms of the features of [+/- static], [+/-dynamic], [+/- punctual], and [+/- telic] and are associated with different logical structures, which can be seen in Table 1.

These logical structures are a representation that originates in the lexical configuration of the predicate of a sentence and are expanded by means of the semantic macroroles and syntactic functions; hence they constitute the main descriptive device of the linking between semantics and syntax. Bearing this in mind, a proposal will be made below to try to describe the way in which the light verb and its collocational partner combine to form a light verb construction. In this sense, the proposal stems from the assumption that both elements specify the manner and aktionsart of the event and provide information about the event itself and that, consequently, the

Tab. 1: Aktionsart types and logical structures in RRG (Van Valin, 2005, 45).

Aktionsart type	Logical structure
State	predicate'(x) or (x, y)
Activity	do'(x, [predicate'(x) or (x, y)])
Achievement	INGR predicate'(x) or (x, y)
	INGR do'(x, [predicate'(x) or (x, y)])
Accomplishment	BECOME predicate'(x) or (x, y)
	BECOME do'(x, [predicate'(x) or (x, y)])
Semelfactive	SEML predicate'(x) or (x, y)
	SEML do'(x, [predicate'(x) or (x, y)])
Active accomplishment	do'(x, [predicate$_1$'(x, (y))]) & BECOME predicate$_2$'(z, x) or (y)
Causative	α CAUSE β, where α and β are LSs of any type

meaning of LVCs can be derived from the meaning of their parts. With respect to the relationship between the light verb and the less verbal element, as the eventive noun is not a verb, it appears to require a verbal frame to be able to realized as the sentence predicate, and this is provided by the light verb according to both its syntactic properties and the semantic requirements of its argument structure, as can be illustrated by examining the syntactic and semantic and syntactic specifications of the events in which the different light verbs occur.

Thus, for example, the predicate *gabh* 'take' denotes an agentive action whereby a participant gets hold of an inanimate entity, and as such requires two arguments, namely an animate agent or recipient and an inanimate theme.[8] Thus, in many LVCs containing this light verb, the nominal element, which denotes an event involving a single argument, appears to function as the logical object of *gabh* 'take' (e.g. *gabh an caothach* 'go mad' (lit. take a madness), *gabh cuairt*/ *sràid* 'take a walk', *gabh fearg* 'get angry' (lit. take anger), *gabh fras*(*air*) 'take a shower', *gabh iongantas* 'be surprised' (lit. take surprise), *gabh òraid* 'deliver a speech' (lit. take speech), *gabh òran* 'sing a song' (lit. take song), *gabh smoc* 'have a smoke' (lit. take smoke), *gabh stròc* 'suffer a stroke' (lit. take stroke), *gabh turas* 'take a trip', etc.):

8 This analysis uses the following terminology to identify thematic or semantic roles: 1. 'agent', which refers to an animate, volitional and causative participant that controls the action expressed by the predicate; 2. 'patient', which refers to an entity being affected by the action denoted by the predicate; 3. 'theme', which refers to an entity undergoing a change of location or being exchanged in a transference; 4. 'goal', which refers to the final destination in a movement; 5. 'recipient', which refers to an animate participant that receives the entity being transferred in the action; 6. 'experiencer', which refers to an animate participant that apprehends a mental or sensory phenomenon; and 7. 'referent', which refers to an entity that is directly involved in the event caused by another entity.

(12) a. *Ghabh sinn cuairt.*
take.PST 1PL walk
'We took a walk.'
b. take: [do'(1PL,)] CAUSE [BECOME have'(1PL,*cuairt*)]
↑
walk: do'(1PL,[walk'(1PL)])

In those cases where the eventive noun requires two arguments, one of them is either realised as a prepositional object functioning as an adjunct in the LVC (13) (e.g. *gabh beachd air* 'examine' (lit. take an opinion of), *gabh brath air* 'take advantage of', *gabh cùram ri* 'take care of', *gabh dèidh de* 'take a liking to', *gabh pàirt ann* 'take part in', *gabh gaol air* 'fall in love with' (lit. take love on), *gabh greim air* 'take hold of' (lit. take grasp on), *gabh sùim ann* 'show consideration for' (lit. take attention in), etc.) or left implied (14) (e.g. *gabh eagal* 'take fright', *gabh stiùir* 'take control' (lit. take lead), etc.):

(13) a. *Ghabh iad beachd air càr.*
take.PST 3PL opinion on car
'They examined the car.'
b. take: [do'(3PL,∅)] CAUSE [BECOME have'(3PL, *beachd*)]
↗
examine: do'(3PL,[see'(3PL,car)])

(14) a. *Ghabh mi eagal san àite-obrach.*
take.PST 1SG frighten in_the place-work
'I took fright at work.'
b. take: [do'(1SG,∅)] CAUSE [BECOME have'(1SG, *meas*)]
↗
respect: respect'(<someone>, 1SG)

The predicate *thoir* 'give' denotes an agentive action that involves transferring the possession of something concrete to someone, which means that it requires three participants: an animate agent, a (generally) non-animate theme, and an animate recipient. This configuration is illustrated by most of the LVCs containing this light verb and, unlike the previous light verb, the second argument of the nominal predicate, which is an argument of the light verb, occurs as an object of the preposition *do* 'to' (e.g. *thoir breith do* 'give birth to', *thoir cead do* 'give permission to', *thoir comhairle do* 'give advice to', *thoir gealladh do* 'make a promise to' (lit. give promising to), *thoir duais do* 'reward' (lit. give reward to), *thoir mathanas do* 'forgive' (lit. give forgiveness to), *thoir meas*/ *spèis*/ *urram do* 'pay respect to', *thoir misneachadh do* 'encourage' (lit. give encouragement to), *thoir òrdugh do* 'give an order to', *thoir pliutag do* 'give

a slap to', *thoir sgoil do* 'educate' (lit. give school to), *thoir taing do* 'give thanks to', *thoir taic do* 'give support to', *thoir toghaidh do* 'pay attention to', etc.), which always introduces the recipient argument (or the indirect object) in Scottish Gaelic. We must therefore assume that it is an (oblique core) argument, rather than an adjunct, to the light verb:

(15) a. *Thug iad gaolagan dhuinn.*
give.PST 3PL cuddle to_1PL
'They gave a hug to us.'

b. give: [do'(3PL,∅)] CAUSE [BECOME have'(1PL,*gaolagan*)]
↗
hug: do'(3PL,[hug'(3PL, 1PL)])

However, it is of note that there are also some examples of LVCs where the verb *thoir* 'give' appears to have lost more of its semantic content, as the second argument of the eventive noun is preceded by a different preposition (e.g. *thoir bideadh*/ *glàmadh air* 'take a bite from' (lit. give biting/ devouring on) , *thoir damh-shùil air* 'throw a glance at' (lit. give ox-eye on), *thoir feart air* 'pay attention to' (lit. give attention on), *thoir sgol air* 'give a telling-off to' (lit. give a splinter on), *thoir ranuns air* 'give a telling-off to' (lit. give telling-off on), *thoir sùil air* 'take a look at' (lit. give look on), etc.), which may be an indication that it functions as an adjunct in this construction. This could be related to a semantic property of the nominal element, which would prevent the entity from being (at least physically) transferred and, consequently, a different preposition is needed to refer to the recipient-like argument. More difficult to explain is why some examples of LVCs with this same light verb include an eventive noun having a single argument (e.g. *thoir ceum* 'take a step' (lit. give step), *thoir leum* 'jump' (lit. give jump), *thoir mionnan* 'swear' (lit. give oaths), *thoir oidhirp* 'make an attempt' (lit. give attempt), *thoir seachad òraid* 'deliver a speech' (lit. give past speech), etc.), unless we assume, as discussed above, that the eventive noun has no logical object – the first two examples – or that the missing argument has been left implied – the last three examples – , although in this first case we should explain why the noun uses this light verb rather than one that corresponds to a two-argument verb.

The predicate *dèan* 'do/ make' generally expresses an agentive action where an inanimate entity, which can be either concrete or abstract, is brought into existence, so it requires two arguments, that is, an animate agent and an inanimate patient-like argument. This explains why the corresponding light verb can be combined with such a large number of verbal and deverbal nouns functioning syntactically as its argument (e.g. *dèan aithreachas* 'repent' (lit. do penitence), *dèan bigeireachd air* 'tease' (lit. do teasing on), *dèan braim* 'fart' (lit. do fart), *dèan breug* 'tell a lie' (lit. do lie), *dèan cabhag* 'hurry' (lit. do hurry), *dèan cac* 'defecate' (lit. do excrement),

dèan cadal ‘sleep’ (lit. do sleep), *dèan casad* ‘cough’ (lit. do cough), *dèan còmhnaidh* ‘dwell’ (lit. do dwelling), *dèan foighidinn* ‘have patience’ (lit. do patience), *dèan mùin* ‘urinate’ (lit. do urination), *dèan obair* ‘work’ (lit. do work), *dèan oidhirp* ‘make an attempt’, *dèan òran* ‘compose a song’ (lit. do song), *dèan stad* ‘take a break’ (lit. do stop), *dèan suidhe* ‘sit’ (lit. do sit), etc.):

(16) a. *Dèan cadal a-nis*!
do.2SG.IMP sleep now
‘Sleep now!’
b. do: [do’(2SG,∅)] CAUSE [BECOME exist’ (*cadal*)]
↑
sleep: do’(2SG,[sleep’(2SG)])

In cases where the eventive noun has a logical object, this is realized as an adjunct in the LVC introduced by a preposition other than *do* ‘to’ (e.g. *dèan cobhair air* ‘help’ (lit. do help on), *dèan dealbh de* ‘take a picture of’ (lit. do picture of), *dèan fabhar ri* ‘do a favour to’, *dèan feum de* ‘make use of’, *dèan gàire ri* ‘laugh at’ (lit. do laugh to), *dèan magadh air* ‘mock’ (lit. do mocking on), *dèan seimhig de* ‘make a fool of’, *dèan slìomaireachd ri* ‘flatter’ (lit. do flattery on), *dèan snodha-gàire ri* ‘smile at’ (lit. do smile to), *dèan taoim air* ‘make fun of’ (lit. do bilgewater on), *dèan tròcair air* ‘have mercy upon’ (lit. do mercy on), *dèan ùrnaigh ri* ‘pray to’ (lit. do prayer on), etc.), the only exception being *dèan adhradh do* ‘worship’ (lit. do worship to) :

(17) a. *Rinn iad feum de na h-iuchraichean*.
do.PST 3PL use of the key.PL.DAT
‘They made use of the keys.’
b. do: [do’(3SG,∅)] CAUSE [BECOME exist’ (*feum*)]
↑
use: do’(3PL,[use’(3PL,keys)])

The predicate *cuir* ‘put’ generally expresses an agentive action in which a participant moves an inanimate entity and places it in a certain place, so it requires three arguments: an animate agent, an inanimate theme, and a goal. In LVCs including this light verb the goal corresponds to one of the two logical objects of the eventive noun – the other is left implied – , which may be an animate or inanimate theme and may be introduced by different prepositions (e.g. *cuir ceist air* ‘ask a question to’ (lit. put question on), *cuir coire air* ‘put the blame on’, *cuir earbsa ann* ‘put faith on’, *cuir fàilte air* ‘greet’ (lit. put welcome on), *cuir fios gu* ‘contact/ inform’ (lit. put knowledge to), *cuir iarrtas air* ‘make a request for’ (lit. put request on), *cuir sradag ri* ‘set fire to’ (lit. put spark to), *cuir ùidh ann* ‘take an interest’ (lit. put interest in), *cuir uidhireachd air* ‘pay attention to’, etc.):

(18) a. *Chuir mi ceist oirre.*
put.PST 1SG question on_3SG.F
'I asked a question to her.'

b. put: [do'(1SG,∅)] CAUSE [BECOME be-LOC'(3SG.F,*ceist*)]
↗
ask: [do'(1SG,[say'(1SG,*ceist*)])] CAUSE [do'(3SG.F,[say'(3SG.F,<something>)])] CAUSE [BECOME aware.of'(1**sg**, <something>)]

A different version of this LVC occurs when the light verb is combined with a prepositional, rather than a nominal, element, which corresponds to the goal argument of the light verb (e.g. *cuir air aghaidh* 'give birth to' (lit. put on face), *cuir air cùil* 'forget' (lit. put on corner), *cuir air chuimhne* 'remind' (lit. put on memory), *cuir a dholaidh* 'abuse' (lit. put one's harm), *cuir air ghluasad* 'set in motion', *cuir an àite* 'replace' (lit. put the place), *cuir ann an cunnart* 'put at risk', *cuir ann an òrdan* 'put in order', *cuir às a ghabhail* 'disappoint' (lit. put out of ordeal), *cuir fo gheasaibh* 'put a spell on' (lit. put under charm), *cuir fo smachd* 'bring under control' (lit. put under control), *cuir gu dùbhlan* 'challenge' (lit. put to challenge), *cuir na theine* 'set on fire' (lit. put in the fire), *cuir san àireamh* 'take into account' (lit. put in the count), etc.).

(19) a. *Chuir mi an duilgheadas fo smachd.*
put.PST 1SG the problem under control
'I put the problem under control.'

b. put: [do'(1SG,∅)] CAUSE [BECOME be-LOC'(*smachd*, *duilgheadas*)]
↗
control: do'(1SG,[control'(1SG, *duilgheadas*)])

This prepositional element normally requires the presence of an additional nominal element and it is this element that functions as the theme argument of the light verb (19).

The predicate *faigh* 'get' commonly expresses both agentive and non-agentive actions that involve an animate entity coming into the possession of something with apparent difficulty, so it generally requires two arguments, an animate agent or recipient and an inanimate theme or referent. Thus, in many examples of this LVC, this light verb is combined with an eventive noun functioning as the theme or referent argument of the light verb (e.g. *faigh an t-sitig* 'be kicked out' (lit. get the dunghill), *faigh àrach* 'be brought up' (lit. get raising), *faigh àrdachadh* 'get a promotion', *faigh brath* 'receive notice' (lit. get information), *faigh cead* 'get permission' (lit. get permission), *faigh chùrsadh* 'get reprimanded' (lit. get cursing), *faigh comhairle* 'get advice', *faigh donnag* 'receive a blow' (lit. get blow), *faigh prìomhachas* 'receive

priority', *faigh ranuns*/ *raphuins* 'get a telling-off', *faigh stuigeadh* 'get urged on' (lit. get inciting), *faigh teic* 'receive support', *faigh togail* 'feel elated' (lit. get lifting), etc.). While it seems obvious that the eventive noun in these LVCs requires two arguments, one of them, namely the agent, is not generally explicitly mentioned. As regards the other argument, namely the referent, it can generally be identified with the agent argument of the light verb:

(20) a. *Gheibh mi taic (bho mo charaidean).*
get.FUT 1SG support from 1SG.POSS friend.PL.DAT
'I will get support from my friends.'
b. get: [do'(1SG,∅)] CAUSE [BECOME have' (1SG,*taic*)]
↗
support: do'(<my friends>,[support'(<my friends>,1SG)])

In cases where the light verb is combined with an eventive noun with an explicit object, this is realized as a prepositional object and functions as an adjunct to the light verb (e.g. *faigh aiteal de* 'catch a glimpse of' (lit. get glimpse of), *faigh buaidh air* 'have an effect on' (lit. get effect on) , *faigh cinnt air* 'check' (lit. get certainty on), *faigh cuidhteas de* 'get rid of' (lit. get quittance of), *faigh plathadh de* 'catch a glimpse of' (lit. get glance of), etc.). As can be observed in (21), now it is the agent, rather than the referent of the eventive noun, that is coreferential with the agent of the light verb:

(21) a. *Fhuair sinn aiteal den nighean.*
get.PST 1PL glimpse of_the girl
'We caught a glimpse of the girl.'
b. get: [do'(1PL,∅)] CAUSE [BECOME have'(1PL,*aiteal*)]
↗
glimpse: do'(1PL,[see'(1PL,*nighean*)])

An interesting example is *faigh cron do* 'damage' (lit. get harm to), which appears to add a third argument, since the logical object of the eventive noun is preceded by the preposition *do* 'to'. This could be explained by arguing that there is an even more complex event in which the light verb *faigh* 'get' normally denotes an action that involves achieving something difficult to attain and, once it has been achieved, it is transferred to someone.

The predicate *thig* 'come' normally expresses an agentive action whereby a participant firstly leaves a place and then reaches a new destination, but it requires only two arguments, namely an animate agent and a goal, as the source is generally only implied. The LVC including this light verb appears to respect this configuration since the light verb is combined with a prepositional element, which functions as

its goal (e.g. *thig gu aois* 'mature' (lit. come to age), *thig gu bith* 'come into existence/ become reality' (lit. come to existence), *thig gu buil* 'come true' (lit. come to result), *thig gu ìre* 'grow up' (lit. come to degree of growth), *thig san aois òg* 'become young' (lit. come to the age young), etc.):

(22) a. *Thig i gu aois.*
come.FUT 3SG.F to age
'She will mature'
b. come: do'(3SG.F,[come'(SG.F)]) & INGR be-loc'(*aois*,3SG.F)
↗
be.mature: be'(3SG.F,[mature'(3SG.F)])

Additionally, the prepositional element may also be accompanied by a nominal element that corresponds to the logical object of the preposition and functions as an adjunct in the LVC (e.g. *thig air aghaidh le* 'get on with' (lit. come on face with), *thig air chèilidh air* 'visit' (lit. come on visit on), *thig gu còrdadh le* 'come to an agreement with', *thig gu eòlas air* 'come to know' (lit. come to knowledge on), etc.):

(23) a. *Thàinig iad air chèilidh orm.*
come.PST 3PL to visit on_1SG
'They paid a visit to me.'
b. come: do'(3PL,[come'(3PL)]) & INGR be-loc'(*chèilidh*,3PL)
↑
visit: do'(1SG,[visit'(3PL, 1SG])

Apart from this meaning, *thig* 'come' can also be used to express a process, or more accurately an accomplishment, by which a participant – a patient or experiencer argument – is brought into a specific state or condition or undergoes a mental or sensory event, which is realized by either an eventive nominal predicate (e.g. *thig am feabhas* 'get better/ improve' (lit. come the goodness), *thig am fianais* 'appear/ come into view' (lit. come the evidence), *thig an àite* 'replace' (lit. come the place), *thig an comhair* 'come near' (lit. come the direction), *thig co-dhùnadh* 'take a decision' (lit. come concluding), etc.) or an eventive adjectival predicate (e.g. *thig beò* 'come alive/ survive', *thig nas fhèarr* 'get better/ improve' (lit. come better), etc.) in the LVC, which refers to a change of state, rather than a goal:

(24) a. *Thigeadh e beó air measan agus glasraich.*
come.COND 3SG.M alive on fruit.PL.DAT and vegetable.PL.DAT
'He would survive off fruit and vegetables.'

b. come: BECOME be.alive'(3SG.M)
↑
be.alive: be'(3SG.M,[alive'(3SG.M)])

The predicate *rach* 'go' can express the same meanings as *thig* 'come', namely an agentive action involving a change of location (e.g. *rach a chadal* 'go to sleep', *rach air adhart* 'proceed/ go forward' (lit. go on advance), *rach air chèilidh air* 'pay a visit to' (lit. go on visit on), *rach air falbh* 'go away' (lit. go on leaving), *rach ann an dàil le* 'get in touch with' (lit. go in a meeting with), *rach cuide ri* 'go along with', etc.) and a process – an accomplishment – leading to obtaining a new state or condition or experiencing an event through the senses (e.g. *rach air chall* 'get lost' (lit. go on losing), *rach am feabhas* 'get better/ improve' (lit. go the goodness), *rach am fianais* 'appear/ come into view' (lit. go the evidence), *rach am follais* 'come to light' (lit. go the clearness), *rach am mearachd* 'become mistaken' (lit. go the error), *rach às do bheachd* 'go crazy' (lit. go out of your judgment)), which shows the presence of two arguments (an animate agent and a goal) or a single argument (a patient or experiencer), respectively.

In the first sense, *rach* is always combined with a prepositional element, which, although it does not denote a location, may serve as a goal (25).

(25) a. *Chaidh mi air falbh.*
go.PST 1SG on leave.VN
'I went away.'
b. go: do'(1SG,[go'(1SG)]) & INGR be-LOC'(*falbh*,1SG)
↑
leave: do'(1SG,[leave'(1SG)])

In the second sense, the light verb is followed by a nominal, adjectival, or prepositional element indicating the process that the only argument of the construction undergoes (26).

(26) a. *Theid e am feabhas gu luath.*
go.FUT 3SG.M the goodness ADV fast
'He will get better quickly.'
b. go: BECOME be.better'(3SG.M)
↑
be.better: be'(3SG.M, [better'(3SG.M)])

The predicate *leig* 'let' denotes an agentive action in which a participant grants permission to carry out another action, so it requires two participants, an animate agent and an inanimate theme. It is of note that this light verb is generally combined

with a single element, which corresponds to an eventive noun with no logical object (e.g. *leig d'anail* 'take it easy/ take a breath' (lit. let your breath), *leig glaodh às* 'give a shout' (lit. let shout out), *leig sgread às* 'let out a scream' (lit. let scream out)). In these LVCs the eventive noun corresponds to the inanimate theme argument of the light verb that expresses the event being permitted, and the person carrying out the action denoted by the eventive noun must by necessity be understood as being coreferential with the agent of the light verb:

(27) a. *An uair sin, leig m'athair osna.*
the time that let.PST 1SG.POSS_father sigh
'Then, my father sighed.'
b. let: [do'(*athair*, ∅)] CAUSE [do'(*athair*,[sigh'(*athair*)])]
↑
sigh: sigh'(*athair*)

A version of this LVC includes an eventive noun including an explicit logical object, which functions as an adjunct in the construction, as it is preceded by prepositions other than *do* 'to' (e.g. *leig an t-srian le* 'give free rein to' (lit. let the bridle with), *leig dub le* 'give a break to' (lit. let dip with), *leig glamhadh gu* 'snap at' (lit. let snap to), etc.). Similarly as in the previous version, the light verb and the eventive noun must share an argument too:

(28) a. *Leig sinn glamhadh thuca.*
let.PST 1PL snap to_3PL
'We snapped at them.'
b. let: [do'(1PL,∅)] CAUSE [do'(1PL,[snap.at'(1PL,3PL)])]
↑
snap.at: do'(1PL,[snap'(1PL,3PL)])

The predicate *tarraing* 'draw' generally expresses an agentive action that entails moving an entity from a source, which is not commonly stated explicitly, by pulling it towards oneself, so it basically requires two arguments, an animate agent and an inanimate theme. This can be observed in all the examples of LVCs containing this light verb and an eventive noun, which functions as one of its arguments and generally has no logical object (e.g. *tarraing anail* 'draw breath', *tarraing conas à* 'pull somebody's leg' (lit. draw quarrel from), *tarraing osna* 'sigh' (lit. draw sigh), *tarraing srann* 'snore' (lit. draw snore), etc.):

(29) a. *Tharraing sinn srann a-raoir.*
draw.PST 1PL snore last_night
'We snored last night.' (lit. We draw a snore last night).

b. draw: [do'(1PL,∅)] CAUSE [BECOME be.away.from'(<somewhere>, *srann*)]
↗
snore: do'(1PL, snore'(1PL))

In cases where the eventive noun has an object, this is always introduced by a preposition other than do and therefore functions as an adjunct to the light verb (e.g. *tarraing aire air* 'pay attention to' (lit. draw attention on), *tarraing dealbh de* 'draw a picture of' (lit. draw picture of), *tarraing stràbh à* 'tease/ poke' (lit. draw strack from), etc.):

(30) a. *Tarraingidh i dealbh den t-sealladh-tìre.*
draw.FUT 3SG.F picture of_the landscape
'She will take a picture of the landscape.'
b. draw: [do'(3SG.F,∅)] CAUSE [BECOME be.away.from'(<somewh.>, *dealbh*)]
↗
photograph: do'(3SG.F,[photograph'(3SG.F,*sealladh-tìre*)])

The predicate *buail* 'hit' expresses an agentive action whereby a participant, generally animate, gives a blow to another participant, either with the hand or with an instrument, so it requires two arguments – an agent and a patient.

(31) a. *Bhuail mi am pathadh madainn an-diugh.*
hit.PST 1SG the thirst morning today
'I became thirsty this morning.'
b. hit: [do'(1SG,∅)] CAUSE [BECOME affected'(*pathadh*)]
↗
be.thirsty: be'(1SG,[thirsty'(1SG)])

ustrated in all the examples of LVCs that include this light verb where it is accompanied by a nominal element with no logical object (e.g. *buail am pathadh* 'become thirsty' (lit. hit the thirst; (31)), *buail basan* 'clap hands' (lit. hit palms), etc.), which corresponds to the patient argument of the light verb.

By contrast, in those cases where the eventive noun has a logical object, this is realized as a prepositional object and functions as an adjunct in the LVC (e.g. *buail breab air* 'give a kick to' (lit. hit kick on), *buail bròg air* 'hit with a shoe' (lit. hit shoe on), *buail dòrainn air* 'torment' (lit. hit torment on), etc.). The problem with this analysis is that of understanding why the eventive noun does not use a different light verb, more specifically a light verb requiring three arguments so that the logical object of the eventive noun may be realized as an argument of the light verb. This could be accounted for by arguing that, in this sense, the light verb *buail* 'hit' can also express an agentive action that entails hitting an entity from a source, which is not

commonly stated explicitly, and, as a consequence of the impact, the entity changes its location. This being the case, it would require three arguments, an animate agent, an inanimate theme, and a goal, which would be realized by the object of the eventive noun:

(32) a. *Bhuail thu breab orm an-dè.*
hit.PST 2SG kick on_1SG yesterday
'You gave a kick to me yesterday.'
b. hit: [do'(2SG,∅)] CAUSE [BECOME be-LOC'(1SG,*breab*)]
↗
kick: do'(2SG,[kick'(2SG,1SG)])

Finally, the predicate *cùm* 'keep' is commonly used to express either an event in which a participant continues to possess something, so it requires two participants, that is, an animate theme and an inanimate referent, or continues to have an entity in a specified location, position, or condition, so it requires three participants, namely an animate agent, an inanimate theme and a goal. Consequently, LVCs containing the light verb *cùm* can appear in two different guises. In the first case (33), similarly as with the light verb *cuir* 'put', the light verb is followed by an eventive noun, which would be its referent argument, and the eventive noun includes a logical object that is realized as an adjunct indicating the goal of the action of the event denoted by the noun (e.g. *cùm taic ri* 'lend support to' (lit. keep support to), *cúm cridhe ri* 'cheer up' (lit. keep heart to), etc.). In the second case (34), the light verb is combined with a prepositional predicate – which corresponds to its goal – and the logical object of the prepositional predicate is realized as the argument of light verb (e.g. *cùm air chuimhne* 'remember/ keep in mind' (lit. keep on memory), *cùm air mheomhair* 'memorise' (lit. keep on memory), *cùm fo breithneachadh* 'keep under review' (lit. keep under consideration), etc.):

(33) a. *Chùm mi taic riut an-cómhnaidh.*
keep.PST 1SG support to_2SG always
'I always lent support to you.'
b. keep: [do'(1SG,∅)] CAUSE [have'(1SG,*taic*)]
↑
support: do'(1SG,[support'(1SG,2SG)])

(34) a. *Cùmaidh sinn thu air chuimhne.*
keep.FUT 1PL 2SG on memory
'We will keep you in mind.'

b. keep: [do'(1PL,∅)] CAUSE [be-at'(*cuimhne*, 2SG)]

↑

remember: remember'(1PL,2SG)

A summary of the thematic structure of the light verbs is included in Table 2. We can see that some of the semantic configurations of the predicates that appear as light verbs in LVCs are the same or very similar – especially if the third argument is omitted with the verbs *thoir* 'give', *cuir* 'put', and *leig* 'let' – which suggests that some of these verbs could be replaced by others with the same or similar semantic properties without altering the meaning of the LVC (e.g. *gabh*/ *faigh comhairle* 'take advice' (lit. take/ get advice), *thoir*/ *dean ionnsaigh air* 'make an attack on' (lit. give/ do attack on), *dèan*/ *leig mùin* 'urinate' (lit. do/ let urination), *tarraing*/ *leig osna* 'sigh' (lit. draw/ let sigh), *dèan*/ *tarraing dealbh de* 'take a picture of' (lit. do/ draw picture of), *thig*/ *rach am feabhas* 'get better/ improve' (lit. come/ go the goodness), *thoir*/ *dèan*/ *gabh greim air* 'take hold of' (lit. give/ do/ take grasp on), *gabh*/ *dèan norrag* 'take a nap' (lit. take/ do nap), *gabh*/ *thig stròc* 'suffer a stroke' (lit. take/ come stroke), etc.). Furthermore, the fact that not only eventive nouns but also some light verbs share many semantic properties appears to show that LVCs are able to express the same or a very similar meaning with a different structure (e.g. *gabh beachd air* (lit. take opinion on) = *thoir sùil air* (lit. give look on) 'take a look at', *gabh cùram ri* (lit. take care to) = *thoir aire air* (lit. give attention on) 'take care of', *thoir togail cridhe do* (lit. give raising heart to) = *cuir a chridhe ri* (lit. put the heart to) 'cheer up', *thoir*

Tab. 2: Semantic properties of light verbs.

Light verb	Thematic roles	Aktionsart class
gabh 'take'	agent/ recipient + theme/ referent	Causative accomplishment
thoir 'give'	agent + theme + recipient	Causative accomplishment
dèan 'do/ make'	agent + patient/ theme	Causative accomplishment
cuir 'put'	agent + theme + goal	Causative accomplishment
faigh 'get'	agent/ recipient + theme/ referent	Causative accomplishment
thig 'come'	1) agent + goal	Activity/ active
accomplishment	2) patient/ experiencer	Accomplishment
rach 'go'	1) agent + goal	Activity/ active accomplishment
	2) patient/ experiencer	Accomplishment
leig 'let'	agent + theme	Causative accomplishment
tarraing 'draw'	agent + theme (+ source)	Causative accomplishment
buail 'hit'	1) agent + patient	Causative accomplishment
	2) agent + theme + goal	Causative accomplishment
cùm 'keep'	1) theme + referent	Causative state
	2) agent + theme + goal	Causative state

feart air (lit. give notice on) = *cuir uidhireachd air* (lit. put attention on) 'pay attention to', *dèan stad* (lit. do stop) = *leig d'anail* (lit. let your breath) 'take a break', *thig air aghaidh le* (lit. come on face with) = *rach cuide ri* (lit. go along with) 'get on with', *thoir taic do* (lit. give support to) = *dèan frìdeam air* (lit. do sufficiency on) 'give support to', *thoir cuireadh do* (lit. give invitation to) = *tarraing cuireadh gu* (lit. draw invitation to) 'invite', etc.

Finally, the opposite seems to occur when the same eventive noun is combined with two light verbs with different semantic properties, which means that the same event is understood but with a special emphasis on a different participant, namely agent or theme/ recipient, as we can see in LVCs using *thoir* 'give' and *faigh* 'get' (e.g. *thoir taic do* 'give support to' vs. *faigh taic* (lit. get support) 'receive support', *thoir ranuns air* 'give a telling-off to' (lit. give telling-off on) vs. *faigh ranuns* 'get a telling-off', *thoir cead do* 'give permission to' vs. *faigh cead* 'get permission', etc.).

This evidence appears to demonstrate that the light verb is not totally devoid of meaning and that it contributes to the overall understanding of the event expressed by the complex predicate, especially because the choice of a particular light verb can modify the meaning of the construction. However, the light verb does not contribute its own separate domain of predication. Rather, light verbs should be seen as contributing to the argument structure of the complex predicate in such a way that the thematic grid of the eventive noun contains roles which are generally compatible with the semantic roles of the light verb, hence, as they are very similar but not necessarily identical, the two argument structures seem to merge together to express the event denoted by the complex predicate.

The evidence given above indicates that a LVC formation is clearly compositional, although, at least in Scottish Gaelic, it is not a regular process, as, in some cases, the selection of the light verb seems somewhat arbitrary. These two positions could be reconciled by arguing that there are different types of LVCs, some with a higher degree of compositionality than others and are therefore more restrictive as to the choice of light verb. As discussed above, it is also possible to identify some groups of LVC in which the same light verb can be combined with several nouns that share part of the same meaning, and the same noun can be combined with more than one light verb that could adapt to the meaning denoted by the noun.

Finally, the evidence that light verbs contribute to the construction semantically also suggests that light verbs are closely related to their corresponding main verb in the lexicon, as is noted by (Butt, 2003, 16; Butt, 2010, 71–74. We might therefore assume only one lexical entry, which tends to give rise to the interpretation of the verb as a main verb or as a light verb depending on the context.

3.2 The representation of LVCs

The second part of this section aims to provide a functional characterization of Scottish Gaelic LVCs within the RRG framework. So far, LVCs have not received a lot of attention in this linguistic model possibly due to their dual lexical-syntactic nature; consequently, some examples of LVCs – especially those that are strongly lexicalized – appear to resist a standard analysis of clausal syntax. However, given its strong typological orientation, the RRG theory may offer a valuable insight for the analysis and representation of this complex predicate type. While it seems clear that LVCs should be considered as syntactically formed complex predicates resulting from a merged logical structure, given that they have such a heterogeneous nature, it seems logical to assume that not all LVCs can have the same syntactic structure, mainly because their light verbs show a higher or lower degree of grammaticalization, and the nominal element has more of the properties belonging to either a complement or a nucleus of the construction. Thus, this section will explore different possibilities for the analysis and representation of LVCs.

RRG bases its study and analysis of complex sentences on two related notions, that is, nexus relations and juncture types (Van Valin, 2005, 183–188, Bentley et al., 2023, 68–73). The first concept refers to the different syntactic relations between the units in a complex construction, namely coordination, cosubordination, and subordination. These three syntactic linkage relations characterize the syntactic relationship between the units in such a way that coordination involves the joining of two or more units of the same syntactic status and, with independent operators, cosubordination entails the combination of two or more units of the same status but sharing operators, and subordination involves the structural dependency of one unit on another (peripheral subordination) or the embedding of one unit within another (daughter subordination). The second concept corresponds to the nature of the units being linked, that is, nuclear, core, clausal, and sentential. Finally, the combination of nexus relations and juncture types leads to a number of nexus-juncture or linkage types, which will be applied to the analysis and representation of LVCs in Scottish Gaelic.

The first option available to enable us to analyze and represent LVCs is to understand them as a regular verb-noun combination, that is, a simple sentence containing a full lexical verb functioning as the only predicate and its complement(s). In this sense, the light verb is then not or hardly delexicalized and keeps all, or most of, its semantic content. Consequently, the nominal element can be considered to function as a direct core argument of the light verb. This option is only correct if the light verb – for example *thoir* ‘give’ in this example – conveys the prototypical meaning

of the verb,[9] which involves the transference of an entity to a recipient, and the nominal element is considered to be its object, but it is then difficult to explain why the interpretation of the eventuality depends solely on the verb (Fig. 1).

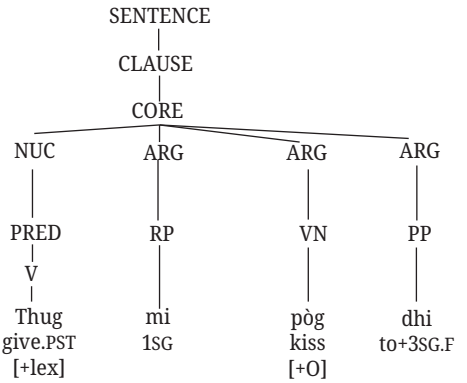

Fig. 1: Representation of an LVC as a free verb-noun combination.

The second option involves a light verb that has lost most of its semantic content so that it functions as a nucleus – that is, the grammatical head of the construction – but is almost unable to predicate; hence the primary predicate of the sentence is the nominal element, in this example *pòg* 'kiss', as it provides the semantic core of the LVC. This analysis of an LVC as an instance of the nuclear subordinate linkage type also matches the semantic information in the logical structure, but the problem is to understand in what sense the verbal nucleus is structurally subordinate to the nominal element functioning as the predicate (cf. Figure 2).

Another possibility would be to consider an LVC as an instance of nuclear co-subordination whereby two predicational nuclei, namely a slightly desemanticized light verb and a nominal element, which may lack some prototypical object properties, share nuclear operators such as aspect and nuclear negation and merge their argument structures to form a complex predicate, which could also be expressed by a single verb akin to the eventive noun, in this case *pòg* 'kiss'. This analysis would match the semantic information in the logical structure, which presupposes that the action denoted by the predicate 'kiss' requires two arguments that correspond to a first person agent and a third person theme (cf. Figure 3).

9 It should be noted here that it is possible to replace *thoir* 'give' with a lexical verb like *caith* 'cast/throw' in this context without altering the meaning (e.g. *caith pòg air* 'give a kiss to').

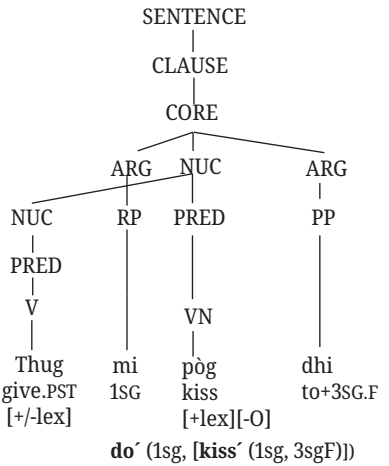

Fig. 2: Representation of an LVC as nuclear subordination.

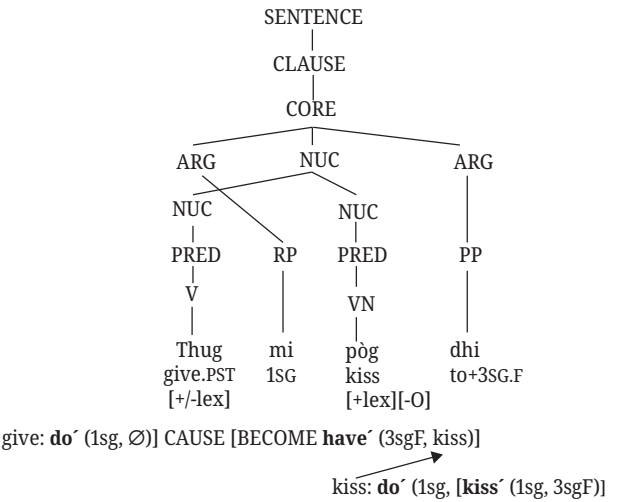

Fig. 3: Representation of an LVC as nuclear cosubordination.

The main problem with this analysis, as pointed out by Fleischhauer (2021, 36), is that in cosubordination "units of equivalent size are joined together in a coordination-like relation, but share some grammatical category" (Van Valin, 2005, 187) and, as the more lexical element in Scottish Gaelic appears to be a noun phrase, rather than a nucleus since it may sometimes be accompanied by a determiner, be pronominalized and accept adjectival modification, this would mean that the two predicating units

that combine to form a single complex nucleus are not symmetrical. However, the term 'size' may be understood to refer to the level of juncture, rather than to the difference between a word and a phrase, and, given that in this construction two nuclei combine to form a single complex nucleus – and all of the arguments of the component predicates are pooled to form a single set of core arguments for the resulting complex nucleus – , two units of the same level would be combining and this would represent an unmarked or symmetrical linkage (Van Valin, 2005, 191, 198).

In summary, depending on the semantic properties of the light verb and its relation with the deverbal noun, we might assume that different types of LVCs exist, which could obviously lead to different representations. On the one hand, following Fleischhauer (2021),[10] the first option, which corresponds to a regular predicate-argument construction, could be an adequate way to represent LVCs if we assume the existence of a mismatch between syntactic and semantic composition in these constructions, as the light verb is the syntactic head of the construction but, though being lexically defective, it also contributes to the predicational content, and the nominal element is the semantic head contributing the major part of the meaning but is an argument of the verb without being its object. On the other hand, in line with other studies (Nolan, 2014; Saeddi, 2016; Staudinger, 2018), the third option could show an effective way of explaining the behavior of most Scottish Gaelic LVCs, more in particular of the resolution of argument sharing within complex predication, as the argument structure of the complex predicate does not appear to be directly licensed by the noun or the verb alone; rather, it generally results from the combination of the semantics of the light verb and the eventive noun and follows from the co-composition or merging of the logical structures for the two elements.

Thus, for example, from this perspective, when the meaning denoted by a simple verb like *pòg* 'kiss' is expressed through an LVC, it is necessary to find a light verb that can fit into the syntactic and semantic properties of the lexical verb, which is frequently realized by an eventive noun in the LVC. Firstly, unless both predicates allow for the omission of one of its arguments syntactically, there should be a correspondence between the number of participants required by the light verb and those of the lexical verb. Thus, given that the lexical verb is now realized as a deverbal noun that conforms with the structural properties of the logical structure of the light verb and becomes one of its arguments – generally the one with the unpaired thematic role – only a three-place predicate such as *thoir* 'give' or *leig* 'let', but not other predicates like *dèan* 'do/ make' or *rach* 'go' (with these predicates the logical

10 An important difference between the representation of LVCs shown in 1 and that proposed by Fleischhauer is that, according to this author, the status of the nominal as the main predicational element is not indicated in the syntactic structure (Fleischhauer, 2021, 52).

object of 'kiss' should be realized as an adjunct), can be chosen as the appropriate light verb for this specific event. Secondly, out of all three-place predicates, only a predicate with an identical distribution of thematic roles can be chosen as light verbs. This is the reason why a light verb like *thoir* 'give', which needs a recipient argument, takes precedence over others like *cuir* 'put' or *tarraing* 'draw', which require a locative argument:[11]

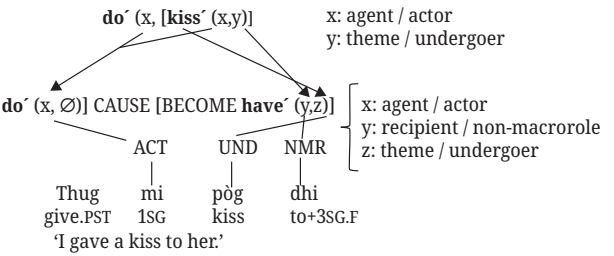

Fig. 4: Co-composition of the logical structure of the complex predicate.

Furthermore, the merging of the logical structures of the light verb and the nominal predicate into the logical structure of the causative accomplishment represented by the event denoted by the complex predicate confirms that it respects the Completeness Constraint (Van Valin, 2005, 129–130), which governs the linking between the syntactic and semantic information by guaranteeing the same number of arguments in the sentence as well as in the logical structure of the predicate.[12]

Finally, in view of this analysis, it seems logical to assume that the co-composition or merging of the logical structures of the two predicational elements takes place lexically rather syntactically. If this account of Scottish Gaelic LVCs is within the domain of nuclear cosubordination, then it must be assumed the two argument structures of the two component logical structures are pooled to form the composite arguments of the structure of the resulting logical structure, which maps into a single core and, consequently, into a monoclausal syntactic structure. This means that the formation of LVCs changes the logical structure and Aktionsarten properties of the primary event – denoted by *pòg* 'kiss' – and affects the argument structure of the predicate and the actor and undergoer assignment – the second macrorole argument of *póg* 'kiss' becomes a non-macrorole argument of *thoir* 'give' and the

11 The second argument of a predicate like 'kiss' appears to be more semantically similar to a recipient than to a locative.

12 This reflects the unmarked or default macrorole assignments for Scottish Gaelic whereby the leftmost argument, *mi* 'I", is the actor, and the rightmost argument, *pòg* 'kiss', is the undergoer.

second macrorole of the latter corresponds to the deverbal noun expressing the primary event (Van Valin and LaPolla, 1997, 389–392, Van Valin, 2005, 158–160.

In summary, the argument structure of an LVC is resolved pre-syntactically in the lexicon by fusing the logical structures of the light verb and the nominal predicate in a principled manner and, although there seem to be different options for their analysis of their representation, which could depend on the existence of different types of LVCs, the most accurate proposal appears to be the construction displaying nuclear cosubordination.

4 Conclusion

After an initial discussion of complex predicates in Scottish Gaelic, which reveals the absence of SVCs and the great productivity shown by AVCs, this paper offers an analysis of LVCs that demonstrates their great heterogeneity with regard to both the semantic properties of the light verb – a different degree of desemanticization or verb lightness – and the syntactic behavior of the noun denoting an eventuality – a different degree of non-referentiality and objecthood – even within LVCs that use the same light verb. This diversity explains the difficulty in defining and establishing cross-linguistic properties of LVCs, such as the type of information expressed by the light verb and the syntactic status of its collocational partner. This seems to suggest that there could be more than one type of LVC and that therefore an analysis and representation of this construction should be carried out by taking into account the distinctive properties of each type. However, said process, owing to its extension and complexity, requires a study of greater length.

Even with this lack of homogeneity, it also seems clear that, despite being a semantically bleached version of a main verb, the light verb is not totally devoid of semantic content, although the main predicational content is provided by the nominal element. This is because, although its meaning is not necessarily transparent and is consequently hard to characterize, it specifies information about the aktionsart and the manner of the event by giving more prominence to certain participants (agent, patient, recipient, etc.) as well as bringing out some nuances of meaning like causation and volition, for example, and serves to structure and/or modulate the event structure of the complex predicate, as is reflected in its influence on valence or argument structure. We might therefore conclude that the formation of the logical structure of the entire complex predicate arises from the interaction of the semantic properties of the argument structures of the component logical structures of the light verb and of the nominal predicate in such a way that, for the merging of the two argument structures, the thematic roles of the light verb and the eventive noun

must be compatible and the linking between the syntactic and semantic information must conform with the Completeness Constraint. In conclusion, the combination of the light verb and the nominal predicate in an LVC behaves like a syntactic and semantic unit that may be represented in different ways, depending on the specific configuration of each type of LVC, although the most logical option appears to involve analyzing it as an instance of nuclear cosubordination. Understanding an LVC as an instance of this linkage type accurately reflects the correlation between the syntactic and semantic properties of the construction, as is reflected in the linking algorithm that maps semantics into syntax. In this view, an LVC would be seen as an instance of a lexical, rather than a syntactic, phenomenon, hence it is formed pre-syntactically in the lexicon, as the merging of the argument structure of the verbal and the nominal predicate affects the complex predicate's aktionsart class and its associated composite logical structure. In connection with the idea of grammaticalization, it seems logical to think of LVCs – rather than light verbs – as different types of lexicalized units in a continuum, which differ with respect to whether they form an inseparable unity and whether they show idiomaticity.

Bibliography

Adger, David. 1996. Aspect, agreement and measure phrases in Scottish Gaelic. In Robert D. Borsley and Ian Roberts (eds.), *The syntax of the Celtic languages: A comparative perspective*, 200–222. Cambridge: Cambridge University Press.

Adger, David. 2010. Gaelic Syntax. In Moray Watson and Michelle Macleod (eds.), *The Edinburgh Companion to the Gaelic Language*, 304–351. Edinburgh: Edinburgh University Press.

Aikhenvald, Alexandra. 2006. Serial verb constructions in typological perspective. In Alexandra Aikhenvald and Robert M. W. Dixon (eds.), *Serial Verb Constructions*, 1–68. Oxford: Oxford University Press.

Aikhenvald, Alexandra. 2018. *Serial verbs*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198791263.001.0001.

Aikhenvald, Alexandra and Robert M. W. Dixon. 2006. *Serial verb constructions: a cross-linguistic typology*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199279159.001.0001.

Aikhenvald, Alexandra and Peter C. Muysken. 2011. *Multi-verb Constructions: A View from the Americas*. Leiden: Brill.

Alsina, Alex, Joan Bresnan and Peter Sells. 2001. *Complex predicates*. Stanford: CSLI Publications.

Amberber, Mengistu, Brett Baker and Mark Harvey. 2010. *Complex Predicates. Cross-linguistic Perspectives on Event Structure*. Cambridge: Cambridge University Press.

Anderson, Gregory. 2006. *Auxiliary Verb Constructions*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199280315.001.0001.

Anderson, Gregory. 2011. Auxiliary verb constructions (and other complex predicate types): A functional–constructional overview. *Language and Linguistics Compass* 5(11). 795–828. https://doi.org/10.1111/j.1749-818X.2011.00311.x.

Baker, Mark. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Bentley, Delia, Ricardo Mairal Usón, Wataru Nakamura and Robert D. Van Valin. 2023. *The Cambridge Handbook of Role and Reference Grammar*. Cambridge: Cambridge University Press.

Bruening, Benjamin. 2015. Light Verbs are Just Regular Verbs. *U. Penn Working Papers in Linguistics* 21(1). 1–10.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications.

Butt, Miriam. 2003. The Light Verb Jungle. *Harvard Working Papers in Linguistics, Papers from the GSAS/Dudley House Workshop on Light Verbs* 9. 1–49.

Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Mengistu Amberber, Bret Baker and Mark Harvey (eds.), *Complex Predicates in Cross-Linguistic Perspective*, 48–78. Cambridge: Cambridge University Press.

Butt, Miriam and Wilhelm Geuder. 2001. On the (semi)lexical status of Light Verbs. In Norbert Corver and Henk van Riemsdijk (eds.), *Semilexical Categories: On the Content of Function Words and the Function of Content Words*, 323–370. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110874006.323.

Butt, Miriam and Aditi Lahiri. 1998. The Status of Light Verbs in Historical Change. Paper presented at DIGS 5, York.

Butt, Miriam and Aditi Lahiri. 2002. *Historical stability vs. historical change*. Unpublished manuscript. https://ling.sprachwiss.uni-konstanz.de/pages/home/butt/main/papers/stability.pdf, last access 26/9/2023.

Byrne, Michel. 2018. *Gràmar na Gàidhlig*. Isle of Lewis: Stòrlann.

Cox, Richard A. V. 2017. *Geàrr-Ghràmar na Gàidhlig*. Perthshire: Clann Tuirc.

Cram, David. 1981. *Scottish Gaelic syntax: a transformational approach*. Unpublished manuscript.

Fleischhauer, Jens. 2021. Simplex and Complex Predicates in Persian: An RRG Analysis. In Robert Van Valin (ed.), *Challenges at the Syntax-Semantics-Pragmatics Interface*, 31–62. Newcastle upon Tyne: Cambridge Scholars Publishing.

Fleischhauer, Jens and Thomas Gamerschlag. 2019. Deriving the meaning of light verb constructions – a frame account of German *stehen* 'stand'. In Constanze Juchem-Grundmann, Michael Pleyer and Monika Pleyer (eds.), *Yearbook of the German Cognitive Linguistics Association, vol. 7*, 137–156. Berlin: De Gruyter Mouton. https://doi.org/10.1515/gcla-2019-0009.

Fleischhauer, Jens, Thomas Gamerschlag, Laura Kallmeyer and Simon Petitjean. 2019. Towards a compositional analysis of German light verb constructions (LVCs) combining lexicalized tree adjoining grammar (LTAG) with frame semantics. In *Proceedings of the 13th international conference on computational semantics - long papers*, 79–90. Gothenburg: Association for Computational Linguistics. https://www.aclweb.org/anthology/W19-0407.

Foley, William and Robert D. Van Valin. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.

Gillies, William. 1993. Scottish Gaelic. In Martin Ball and James Fife (eds.), *The Celtic Languages*, 145–228. London): Routledge.

Haspelmath, Martin. 2016. The serial verb construction: Comparative concept and cross-linguistic generalizations. *Language & Linguistics* 17(3). 291–319. https://doi.org/10.1177/2397002215626895.

Heine, Antje. 2006. *Funktionsverbgefüge in System, Text und korpusbasierter (Lerner) Lexikographie*. Frankfurt am Main: Peter Lang.

Heine, Bernd. 1993. *Auxiliaries. cognitive forces and grammaticalization*. Oxford: Oxford University Press.

Hopper, Paul and Elizabeth Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165525.

Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles, Part VI, Morphology*. Copenhagen: Ejnar Munksgaard.

Kearns, Kate. 2002. Light verbs in English. https://api.semanticscholar.org/CorpusID:116138945, last access 27/9/2024.

Krug, Manfred. 2012. Auxiliaries and grammaticalization. In Bernd Heine and Heiko Narrog (eds.), *The Oxford Handbook of Grammaticalization*, 547–558. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199586783.013.0044.

Kuteva, Tania. 2001. *Auxiliation: An enquiry into the nature of grammaticalization*. Oxford: Oxford University Press.

Lamb, William. In press. *Scottish Gaelic: A Comprehensive Grammar*. London: Routledge.

MacAuley, Donald. 1992. The Scottish Gaelic Language. In Donald MacAuley (ed.), *The Celtic Languages*, 137–248. Cambridge: Cambridge University Press.

Nolan, Brian. 2014. Complex predicates and light verb constructions in Modern Irish. *Spanish Journal of Applied Linguistics* 27(1). 140–167. https://doi.org/10.1075/resla.27.1.06nol?locatt=mode:legacy.

Nunberg, Geoffrey, Ivan A. Sag and Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538. https://doi.org/10.1353/lan.1994.0023.

Ramchand, Gillian. 1993. *Aspect and argument structure in Modern Scottish Gaelic*. Doctoral dissertation: Stanford University.

Roberts, Ian and Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.

Saeddi, Zari. 2016. *Complex predicates in modern Persian*. Sheffield: Equinox.

Staudinger, Eva. 2018. French and Spanish 'MAKE/GIVE + FEAR'"-type LVCs – an RRG Constructional Account. In Rolf Kailuweit, Lisann Kunkel and Eva Staudinger (eds.), *Applying and expanding Role and Reference Grammar*, 237–261. Freiburg: FRIAS.

Stewart, John M. 1963. Some restrictions on objects in Twi. *Journal of African Languages* 2(2). 145–149.

Van Valin, Robert. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511610578.

Van Valin, Robert and Randy LaPolla. 1997. *Syntax*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166799.

Muhammed Ourang, Mengistu Amberber, and Efrosini Deligianni

The derivation of the light verb construction in Aheli

1 Introduction

Aheli is a dialect of the Lari language which is predominantly spoken in the Ahel city of the Lamerd County, Fars Province, Iran. The dialect is spoken by nearly 3500 people as a mother tongue (Ourang, 2022, 2). The Lari language belongs to the South-Western group of Iranian languages which includes a number of dialects spoken in Iran (in southern Fars, western Hormozgan and eastern Bushehr) as well as in a few Persian Gulf countries such as the United Arab Emirates (UAE), Qatar and Kuwait (cf. Figure 1). The total number of Lari speakers is estimated to be between 150 to 200 thousand (Moridi, 2007, 64).

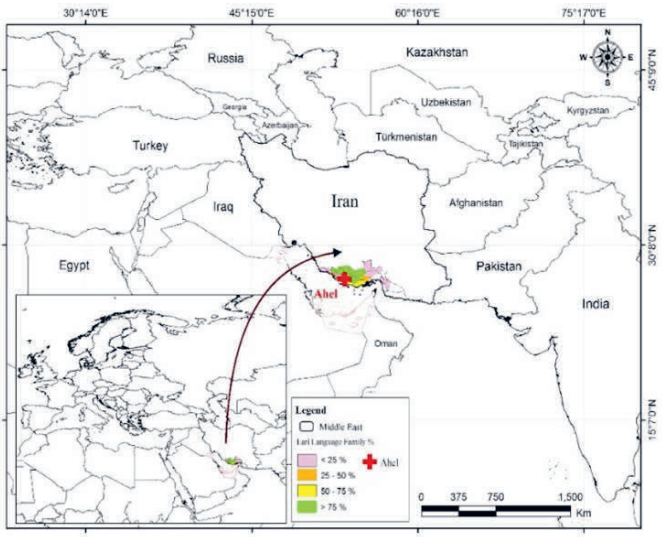

Fig. 1: The areas where the Lari language family is spoken in the Middle East (Ourang, 2022, 8).

Muhammed Ourang, University of New South Wales, Sydney, Australia
Mengistu Amberber, University of New South Wales, Sydney, Australia
Efrosini Deligianni, Independent researcher, Thessaloniki, Greece

https://doi.org/10.1515/9783112229330-003

The Lari language[1] is divided into a number of dialects including Lari, Gerashi, Evazi, Khonji, Bastaki and Bikheyi (Eghtedari, 2005; Kalbasi, 1990), Banarouyi, Bikheyi (Molčanova, 1982), Aheli, Buchiri, Ashnezi, and Kangani (Salami, 2004–2007, 2009, 2011) among others.

Since the dialects of the Lari language are scattered across a large area of Iran, there is some regional linguistic variation as would be expected. The variation[2] mostly occurs in phonology (e.g., in the phonological inventory) and morphology. For instance, phonologically, there is some variation in the following phonemes: /ɣ/, /ʒ/, /o:/, /ü/ and /æ:/. The Lari dialect has all of these phonemes except /æ:/ (Assarian, 2016, xxvi), whereas the Khonji dialect has only /ɣ/ and /ʒ/ (Khonji, 2009, 24), and the Aheli dialect has only /æ:/. The Buchiri dialect has none of these phonemes (Rashidi, 2008, 30). Morphologically, Aheli appears to be the only dialect which marks the past progressive by using a particle (i.e., *kœj*/*kœr* PROG) preceding the past imperfective (*kœj m=œ-ʃet* 'I was kneading [it]' (PROG 1SG=IMPF=knead.PST)). In other dialects, the tense-aspect is marked by prefixes and suffixes. See also Khonji (2009, 87–89) for more information. However, despite these and other variations, there is a high degree of mutual intelligibility between the dialects of the Lari language (Khonji, 2009, 16).

Our paper is organised as follows: Section 2 gives an overview of the scholarship on the Lari language and Aheli dialect, Section 3 provides a brief description of the research methodology, Section 4 delves into the Aheli LVCs semantic templates based on Megerdoomian's classification and gives the LVC formation formulae including combination and pseudo-incorporation. The last section of the paper, section 5, concludes the paper and offers some suggestions for further research. This paper is an effort to examine the derivation of the LVC in Aheli, which are formed by a NVE (usually a noun, an adjective, a preposition, etc.) and an LV such as (*va*)*kerdæ* 'to make', *vavʊzæ* 'to become/to get', *zætæ* 'to hit'. The paper aims to answer the following questions: (a) which LVs participate in the LVC, (b) what classes of NVE participate in the LVC and (c) what makes the Aheli LVC different from the LVC in other Iranian languages like Persian in terms of grammatical and semantic features.

1 It should be mentioned that the term 'Lari language' is used to include all the dialects spoken in the areas shown in Figure 1. Therefore, when a specific dialect of the Lari language is mentioned, the following terms are used: 'the Buchiri dialect', 'the Bastaki dialect', 'the Lari dialect', 'the Aheli dialect', etc.

2 Much more research needs to be done in order to determine the precise nature of inter-dialectal variation.

2 Previous studies on Lari and Aheli

Previous studies on the Lari language have made a significant contribution to the scholarship of the Lari language family. Researchers such as Kamioka and Yamada (1979); Eghtedari (1964, 1992, 2005); Molčanova (1982); Skjaervo (1989); Salami (2004–2007, 2009, 2011); Moridi (2006, 2007); Mahmoodian (2007); Dabir-Moghaddam (2013) and Dianat (2016), to name a few, have examined the phonology, etymology, noun and verb morphology, vocabulary and syntax of the Lari language, and have offered some important insights into the Lari language and its dialects. However, there are some methodological limitations to the previous studies relating to the documentation (e.g., data based on limited fieldwork), analysis (e.g., lack of standard transcription and glossing conventions), and presentation of research findings. Ourang (2022) is the first reference grammar of the Aheli dialect of the Lari language. The grammatical analysis is situated within the framework of descriptive linguistics and is based on extensive fieldwork interviews (see Section 3 below). The grammar provides a background on the Aheli dialect and a sketch of the sound system, morphology and syntax. This work, which is the first reference grammar of the Aheli dialect, has also engaged the Aheli community in the research project (especially monolingual women) by using some non-linguistic prompts such as pictures and short video clips to encourage them to talk about their experiences in farming (e.g., palm tree harvesting and picking dates), wedding ceremonies and childbearing, to name a few. In addition, this work presents a detailed description of the phonology, noun morphology and verb morphology, which provides crucial material for developing resources for teaching the Aheli dialect to children.

3 Research methods

This research draws on data which were collected by the first author through fieldwork which involved interviewing mainly Aheli monolingual speakers (in total 45 interviews, equal to over 10 hours of recordings) during February-March 2018. The interviews were in one-to-one (30 interviews) or group formats (15 interviews). The former lasted around 15-20 minutes each while the latter had a longer duration of around 30-40 minutes each. A total number of 55 monolingual speakers (35 women and 20 men) were interviewed. The majority of the consultants were women. This is because women play a pivotal role in the community and are actively engaged in its daily activities such as harvesting, raising children, cooking, etc.

These data have been complemented by the first author's native speaker intuition when appropriate. For example, the first author's native speaker intuition was

applied to clarify ambiguities in the interviewees' speech, for example, by inquiring about a specific term like *ʧʊngu* 'small spoon for feeding a baby'. In addition, the native speaker intuition helped to complement the fieldwork data by asking follow-up questions to clarify subtle semantic and syntactic acceptability judgments, including, for example, the use of numerous particles such as the present progressive *kæj* PROG. The interviews were conducted using a semi-structured interview format, which delivered recorded conversations, stories, historical narratives, songs and folktales from a total number of 55 monolinguals. The semi-structured interview used several themes including *bæʃkar* 'planting', *davat* 'wedding', *betʃ* 'child-bearing', *nu pʊtæ* 'bread-baking', *ʤuri* 'harvesting', *ʧe:run* 'churn', *ʃælva* 'folklore', i.e., themes that are culturally relevant for Aheli speakers.[3]

4 The light verb constructions (in Aheli)

It is well known that Persian has a productive light verb construction which complements its small number (around 200) of independent heavy verbs. Thus, it is not surprising that the LVC in Persian has been extensively studied by researchers including Dabir-Moghaddam (1997); Karimi (1997); Megerdoomian (2004, 2012); Pantcheva (2008), among others. Eshaghi and Karimi Doostan (2023) have developed the first corpus[4] for Persian light verb constructions which includes 6000 LVCs formed by 21 light verbs (LVs) such as *kærdæn* 'to do', *ʃʊdæn* 'to get', *daʃtæn* 'to have', *dadæn* 'to give', *gereftæn* 'to catch', etc. to name a few in order of frequency of occurrence in their corpus. Nevertheless, far too little has been conducted on the light verb construction in the dialects of the Lari language, especially in Aheli, which highlights the need for high-quality research in this area. The only research paper on compound verbs is Dianat (2023), which examines LVCs based on Dabir-Moghaddam's (1997) categorisation of compounding and incorporation. The author has presented the types of non verbal elements (NVEs; adjectives, nouns, prepositional phrases, pronouns, etc.), which participate in the compounding process and then addressed the incorporation process based on categories of direct object, indirect object and prepositional phrase. We will shed light on the research findings of this paper in the subsequent sections, when we analyse our Aheli data.

This paper aims to examine the LVC with a particular attention to the semantic templates of Megerdoomian (2004), and more specifically the transitivity of a verb

3 The ethics approval for the research was granted by the HREAP B: Arts, Humanities & Law (University of New South Wales, Sydney, Australia) in Sep. 2017 under No. HC17642.

4 Their corpus is called 'Light Verb Construction of Persian Language (LCP)'.

when it functions as a full verb as opposed to when it participates in a LVC, the list of verbs which only appear as light verbs but do not function as full verbs, and finally some important differences between the Aheli LVC and the relatively well-studied Persian LVC.

As Korn (2013, 33) pointed out, the light verbs in Persian typically belong to a small, closed class of verbs, while the nonverbal elements belong to an open class. LVs host the inflection of the verbal phrase including syntactic features of the verbs (such as tense, mood and person), transitivity and voice (Sheintuch, 1976 as cited in Korn, 2013, 32). In Aheli, light verbs such as *kerdæ* 'to make' or *vakerdæ* 'to make, to do' combine with an NVE to derive complex predicates (CPs) such as *saji vakerdæ* 'to repair' (lit. correct make).[5] Other examples of CPs in Aheli are *ʤuri kerdæ* 'to harvest' (lit. harvesting do), *tiva zætæ* 'to separate the hay residues' (lit. residuals hit). See Section 4.2.1 for more examples of LVCs. The NVEs are usually followed by a light verb, a list of which is presented below:[6]

Tab. 1: List of light verbs in Aheli.

kerdæ	'to do/make'	*vakerdæ*	'to do/make'
bʊlezæ	'to cut'	*zætæ*	'to hit'
xærdæ	'to eat'	*ræftæ*	'to go'
*hund**	'to have'	*dæzæ*	'to give'
kefezæ	'to pull'	*ʊftæzæ*	'to fall'
geretæ	'to take'	*ændæ*	'to come'
bʊzæ	'to become'	*ævʊrdæ*	'to bring'
bʊrdæ	'to take'	*vavʊzæ*	'to become, to get'
vaxærdæ	'to drink'		

5 Karimi Doostan (2005, 1746) claims that the NVE in Persian can occur with 'Ezafe'. The term 'Ezafe' is commonly used in the literature on Iranian languages to refer to a clitic that links the head noun to its dependents such as a noun, or an adjective (Stilo, 2012; Korn, 2017). In Aheli, 'Ezafe' has two allomorphs: *=e* which appears after the consonant-ending head nouns and *=j* which follows a head ending in a vowel.

6 It should be noted that 'to have' is an exceptional full verb for which there is no corresponding verb in Aheli. Instead, *hund** 'to be' is used to derive the verb form corresponding to 'to have'. The asterisk shows that the form is unattested, but its existence is inferred from other attested forms.

4.1 Semantic categorisation of LVs in Aheli

Megerdoomian (2004, 26) defines LVCs in Persian as consisting of a preverbal element which can be a noun, an adjective or an adverb, and which combines with a light verb or "semantically bleached" verb to form a single predicate in terms of argument structure and semantic interpretation. The question of whether the light verb has independent semantic meaning is controversial. For example, Bruening (2015, 51) which is a formal study of light verb constructions cross-linguistically, states that the light verb does not contribute that much to the meaning of the LVCs, while the non-verbal element carries the semantic meaning of the construction. Telegdi (1951) (as cited in Korn, 2013, 52) mentions that there are some cases where both the LV and the nonverbal element can equally contribute to the meaning of the complex predicate, while there are some other cases where the LV's function is that of a verbaliser, that is, to convert the nonverbal element into an expression equivalent to a full verb. By the same token, Fleischhauer (2020, 69) explains that the LVs are semantically reduced verbs which contribute to the semantic meaning of the complex predicate. He then goes on to state that there is evidence that the meaning of an LV is not similar to the corresponding heavy verb because the LV does not denote an event of its own (compare for instance, the Aheli LV *xærdæ* in *xʊssæ xærdæ* 'to grieve' (lit. grief eat) and the heavy verb *xærdæ* 'to eat'. Arguing that LVs are not semantically empty, Megerdoomian (2004, 28–30) categorises the LVC templates[7] in Persian into change-of-state alternation verbs, activity verbs and instrument verbs. Adopting her classification, we describe the LVC templates in Aheli in the following section, 4.1.1 to 4.1.3.

4.1.1 Change of state alternation verbs

Change of state verbs have been thoroughly researched in the literature on the Persian LVCs. The templates for these verbs in Aheli are presented in the following sections. In these templates, BECOME and CAUSE represent change of state and causation respectively:

Inchoative: *vavʊzæ* 'BECOME pred (y)'

The example below shows the inchoative template in context:

7 The template is based on the lexical semantic representation of verbal predicates (Megerdoomian, 2004, 25).

(1) *zemi gœɪ hʊʃk ni-vavu-Ø*
ground never dry NEG[IMPF]-become.PST-3SG
'the ground never became dry' [XdjPkr; Djouri Ga Davat]

Causative: *vakerdœ* 'x (Actor/he) CAUSE y (Undergoer/palm date) BECOME pred (dry out)'

The example below corresponds to the causative template:

(2) *arma hʊʃk ʃ=œ-vak-e*
ripe_date dry 3SG=IMPF-make.PST-PTCP
'he used to dry out the palm date' [FmjShms; Arma Vazbeneza]

In the above example, *ʃ=œ-vak-e* is the conjugated third-person form of the infinitive *vakerdœ* 'to make'. Syntactically, the LV construction can appear in clauses with one argument or two arguments. In (1), the single argument *zemi* 'ground' is a theme, while *arma* 'ripe date' in (2) functions as an undergoer and the agent is the 3rd person singular 'he', which is not expressed by a noun phrase but is rather marked on the light verb.

4.1.2 Activity verbs

Megerdoomian (2004, 28) describes the template of activity verbs in Persian as: 'X (Actor/we) ACT'. An example from Aheli is given below for *bœʃkar kerdœ* 'to till' (lit. tillage do). It should be mentioned that the NVE element in the construction is an eventive noun.

(3) *ʧe: ruz bœʃkar m=œ-kerd-e*
forty day tillage 1PL=IMPF-do.PST-PTCP
'we used to till [the land] for 40 days'
[XljAkbr; Bashkar Kerda]

Some more examples of the non-verbal elements that occur in activity verbs are presented below:

(4) a. *ʤuri kerdœ*
harvesting do
'to harvest'
b. *tiva zœtœ*
residue hit
'to separate the hay residues from crops'

c. *ʃəʊr kerdœ*
thinking do
'to consult'

Here is another example for the activity light verb *ʃəʊr kerdœ* 'to consult' (lit. consultation make) in context:

(5) *xunevazœ=j beɪn ʃəʊr œ ʧœp-ʊ-dəʊr=e xʊ=ʃ*
family=EZ bride consultation to left-and-round=EZ self=3SG.PC
ʃ=œ-kerd-e
3PL=IMPF-make.PST-PTCP
'the bride's family used to consult with their relatives' [lit. people around them]

4.1.3 Instrument verbs

The pre-verbal element in this template is interpreted as an instrument of the event. As Megerdoomian (2004, 29) states, the instrument verbs indicate a repetitive action. The template that best describes instrument verbs is: 'X (Actor/they) ACT y (Undergoer) with <Instrument>'. Here is an example from Aheli:

(6) *degœ ta bis rʊz aʊsi ʃ=œ-zœt-e*
then till twenty day winnowing_fork 3PL=IMPF-hit.PST-PTCP
'then they used to winnow for 20 days' [FmjShms; Bashkar]

In the example above, the non-verbal element *aʊsi* 'winnowing fork' is an instrument which is combined with the light verb *zœtœ* to make *aʊsi zœtœ* 'to winnow'. Other examples of nouns that express tools or instruments are shown below:

(7) a. *hʊngi keʃezœ*
broom pull
'to broom'
b. *mœntil zœtœ*
iron_rod hit
'to dig up with an iron rod'
c. *kʊmba kerdœ*
combine_harvester do
'to combine'

4.2 The template for the LVCs

Perhaps the most comprehensive account of LVC formation in Persian is to be found in the work of Dabir-Moghaddam (1997). Reviewing previous works conducted on LVCs in Persian, Dabir-Moghaddam has identified two major types of compound-verb formation: 'combination' and 'incorporation'. Nonetheless, Fleischhauer (2020, 51–59) argues that there are specific criteria which distinguish LVCs from pseudo-incorporation constructions (PICs) in Persian. Applying Fleischhauer's arguments, we show that LVCs in Aheli are different from PICs in terms of the denoted eventuality, where the NVE determines the content in LVCs, while the verb determines the denoted situation in PICs. Furthermore, this is assumed to be due to the fact that the verbal head of the LVCs is semantically light whereas the verbal head of the PIC is a heavy verb. Moreover, the two are different with regard to nominal morphology. Thus, we use the term 'pseudo-incorporation' instead of 'incorporation', but we acknowledge that this issue requires further research.

In the following sections, we describe the templates which show how LVCs are formed using Dabir-Moghaddam's (1997) analysis with some modifications to better accommodate Aheli data.[8] LVCs made through combination are presented in section 4.2.1 and verbs formed via pseudo-incorporation are illustrated in Section 4.2.2.

4.2.1 Combination

Compound verbs formed through combination in Aheli are presented below. In this category, NVEs can be an adjective, a noun, a prepositional phrase, an adverb and a past participle.

8 Monshizadeh et al. (2022, 218) argue that what authors like Dabir-Moghaddam (1997) have categorized as a '[light] verb' in some 'compound verbs' is in fact a (semantic or grammatical) 'predicate' rather than a [light] verb. For instance, *xʊrdæn* in *yæza xʊrdæn* 'to eat' (lit. food eat) is a grammatical predicate. Similarly, *bær-daʃtæn* 'to pick up' (lit. PREF-have) is a semantic predicate. For more details, see Monshizadeh et al. (2022, 228–233).

Adjective + LV

The light verbs participating in this combination process are *kerdœ* 'to do/make', *vakerdœ* 'to do', *bʊzœ* 'to get' and *vavʊzœ* 'to become'. The light verbs which are used to form the compound verbs as well as their transitivity status are presented in (8).[9]

(8) a. *nœk kerdœ*
good do
'to make' [Transitivity: +]

b. *ʃeri kerdœ*
sweet do
'to put make-up' [Transitivity: +]

c. *amʊxtœ kerdœ*
accustomed do
'to make someone accustomed to something' [Transitivity: +]

d. *nœk vakerdœ*
good do
'to repair' [Transitivity: +]

e. *bezar vavʊzœ*
awake become
'to wake up' [Transitivity: -]

f. *gœp bʊzœ*
big get
'to grow' [Transitivity: -]

Here is an example with *nœk* 'good' which functions as a predicative adjective in (9a) but as the non-verbal component of a compound verb in (9b):

(9) a. *xalu xeɪli nœk hen-Ø*
maternal_uncle very good be.NPST-3SG
'the uncle is [a] very good [man]'

b. *mœgœ nœvat nœk m=œ-kerd-e*
always candy_crystal good 1PL=IMPF-make.PST-PTCP
'we always used to make candy crystal'
[XlFmk_SrwFG; Sarau Xashar]

As demonstrated in these examples, the meaning of the compound verb is not always predictable from the meaning of its constituents, namely the non-verbal element

9 In (8) , 'Transitivity: +' means the LVC is 'transitive' while 'Transitivity: -' means the LVC is 'intransitive'. The criterion for transitivity is that an LVC selects a noun phrase as a direct object.

and the light verb. This can be demonstrated further in the following example where *ʃeri* 'sweet' as an NVE element combines with the 3rd singular form of *kerdæ* 'to make' and makes *ʃeri kerdæ* 'to put make-up' (but lit. to make someone sweet):

(10) *pæsin jæ pir-zæn=i æ-ja-Ø ʊ bem ʃeri*
evening one old-woman=INDEF IMPF-come.NPST-3SG and bride sweet
æ-ku-Ø
IMPF–do.NPST-3SG
'an old woman comes and puts make-up on the bride in the evening'
[ZNB_ML; Dava:t Kavi]

The context influences the interlocutor's understanding of the LVCs. For instance, in (10) above, *ʃeri kerdæ* literally means 'to make sweet', but it is a non-compositional LVC which literally means 'to put make-up'. Thus, if the listener is not aware of the context in which this LVC is used, i.e., wedding ceremonies, they might assume that the speaker is talking about 'making sweets' for the wedding rather than the actual intended meaning 'making the bride beautiful'. It is worth noting that this Aheli LVC has now been replaced by the equivalent Persian LVC *arajeʃ kærdæn* which literally means 'make-up do'.

Noun + LV

A large proportion of compound verbs consist of a noun followed by a light verb such as *kerdæ/vakerdæ* 'to make/to do'. Other light verbs which frequently occur are *zætæ* 'to hit', *dæzæ* 'to give', *geretæ* 'to take', *keʃezæ* 'to pull', the equivalent light verb for 'to have', *xærdæ* 'to eat' and *vaxærdæ* 'to drink'.

(11) a. *ʤuri kerdæ*
harvesting do
'to harvest' [Transitivity: -]
b. *mʊrʊk kerdæ*
piling do
'to pile up' [Transitivity: +]
c. *pat kerdæ*
sorting do
'to sort' [Transitivity: +]
d. *kævi bʊlezæ*
engagement cut
'to conclude the marriage contract' [Transitivity: -]

e. *saz zætæ*
instrument hit
'to play an instrument' [Transitivity: -]

f. *ʤar zætæ*
call hit
'to call (someone's name)' [Transitivity: -]

g. *lʊpʊk zætæ*
deception hit
'to deceive' [Transitivity: +]

h. *mʊs zætæ*
fist hit
'to punch' [Transitivity: -]

i. *tælak dæzæ*
divorce give
'to divorce' [Transitivity: +]

j. *bu geretæ*
smell take
'to turn fetid' [Transitivity: -]

k. *əʊ keʃezæ*
water pull
'to rinse' [Transitivity: +]

l. *xævær hund**
news have
'to be informed' [Transitivity: -]

m. *lʊpʊk xærdæ*
deception eat
'to be deceived' [Transitivity: -]

n. *xʊssæ xærdæ*
grief eat
'to grieve' [Transitivity: -]

o. *pitʃ vaxærdæ*
twist drink
'to be twisted' [Transitivity: -]

In the following examples, *kævi* 'engagement' functions as a nominal predicate of the copula verb 'to be' to derive a construction that is equivalent to English 'have' (12), but when it occurs with the light verb *bʊlezæ* 'to cut', it functions as the non-verbal component of a compound verb in (13):

(12) *mʊ sœva kœvi ʊm=hen*
I tomorrow engagement 1SG=be.NPST
'I have my engagement tomorrow'

(13) *ʃ=œ-j kœvi bʊ-vʊl-en*
3PL=IMPF-want.NPST engagement SBJV-cut.NPST-3PL
'they want to conclude the marriage contract' (lit. they want to cut the engagement)
[Znb_ZnbMlDtsh; Nou Sheta]

In the example above, the compound verb has a rather idiomatic interpretation: 'cut an engagement' means 'conclude the marriage contract' (somewhat similar to the expression 'cut a deal' in English). This is in line with the compound verbs formed through the process of combination in Persian where the meaning of the compound verbs is less transparent when compared to those compound verbs formed through pseudo-incorporation as we shall see later in Section 4.2.2. Fleischhauer and Neisani (2020, 53) argue that Persian separable LVCs are compositional and believe that it is important to consider the compositionality as a key factor in the separability of an LVC by an adjective modifier. Applying this analysis to Aheli, we can see that some LVCs can be separated by adding an attributive modifier between the NVE (mainly a nominal preverb) and the LV. For instance, consider the LVC *kœvi bʊlezœ* 'to engage; to conclude the marriage contract' in (14) below:

(14) *Æli kœvi ʊʃ=bʊli*
Ali engagement 3SG=cut.PST
'Ali had an engagement'

As can be seen in (15) below, the adjective 'beautiful' separates the two components of the LVC to modify the construction.

(15) *Æli kœvi-je ʃux-i ʊʃ=bʊli*
Ali engagement-EZ beautiful-INDEF 3SG=cut.PST
'Ali had a beautiful engagement'

This is in line with Nunberg et al. (1994, as cited in Fleischhauer and Neisani, 2020, 46–47) who maintains that the LVC's components have identifiable/compositional meaning if they license internal modification. In a similar example, the *zenœ vagerœtœ* 'to get married' (lit. woman include) is an LVC the meaning of which cannot be understood from the separate meanings of the constituents nor an internal modification of the LVC is licensed (cf. **zenœ-œj nœk vagerœtœ* woman-EZ good include; intended meaning: 'to get married with a good wife'). Nevertheless, further research is required to determine what kind of attributive modification is allowed

in the Aheli LVC and whether or not modification by an adverb (which has scope over the whole LVC) or an adjective (scoping over only the nominal NVE) results in the same interpretation of meaning as shown by Fleischhauer and Neisani (2020) for Persian LVCs.

Prepositional phrase (PP) + LV

The combination of a prepositional phrase and a light verb can form a compound verb. Examples are presented in (16):

(16) a. *æ dʊnja ændæ*
to world come
'to be born' [Transitivity: -]

b. *æ jaz ævʊrdæ*
to remembrance bring
'to remind' [Transitivity: +]

c. *æ la vanesæ*
to waste put
'to waste' [Transitivity: +]

d. *æ bar ræsenæzæ*
to fruit deliver
'to raise' [Transitivity: +]

e. *æ tæʃ næzæ*
to fire put
'to cook' [Transitivity: +]

f. *æ dur ævʊrdæ*
to out bring
'to bring out'[10] [Transitivity: +]

The transitivity of the compound verbs formed by a PP is contingent on the transitivity of the light verb. For instance, *vanesæ* 'to put' is transitive, thus *æ la vanesæ* 'to waste' is transitive too. This is similar to what Dabir-Moghaddam (1997, 37) refers to when he talks about the transitivity of compound verbs in Persian.

As we can see, *æ* 'to', which functions as a preposition occurs with a noun and the resulting prepositional phrase combines with a light verb to derive a compound

10 Dianat (2023, 123) has included this LVC under the category of 'adverb + LV' which seems unproblematic. The LVC is comprised of *æ* 'to' and *dur* 'out' which makes a prepositional phrase *æ dur* 'outside'.

verb. The examples below show *æ* 'to' as a regular preposition for location in (17a) and as part of the non-verbal component of the LVC in (17b):

(17) a. *æ:ma æ me:ræk æ-raft-e-ss-em*
we to me:ræk IMPF-go.PST-PTCP-be.NPST-1PL
'we used to go to me:rak [an old castle]'
[FmRs_AmjSknFmRs; Davat]

b. *væxt-i betʃ æ dʊnja jæmu-Ø pir-zæn-ija karan*
time-RES baby to world [IMPF]come.PST-3SG old-woman-PL Qur'an
ʃ=æ-xʊnd-e
3PL=IMPF-read.PST-PTCP
'when a baby was born, old women used to read Qur'an'
[BvMl_FmMlBvMl; Ancient Ahel]

Adverb + LV

A list of adverbs which combine with a light verb to make a compound verb is given in (18):

(18) a. *vapæs dæzæ*
back give
'to give back' [Transitivity: +]

b. *ʤʊləʊ ʊftæzæ*
front fall
'to advance, to take over' [Transitivity: -]

c. *va gæʃtæ*
again roam
'to return' [Transitivity: -]

Regarding the prefix/adverb *va*, it should be mentioned that in some cases like (19) below, it functions as an adverb meaning 'again' when it occurs independently:

(19) *gʊtæ* → *va-gʊtæ*
say again-say
'to say' vs. 'to say again'

Another example is *letæ* 'to pour' → *va-letæ* 'to pour again' where *va-* is a morpheme equivalent to *re-* in English. However, there are examples where *va-* is a semantically vague prefix which attaches to a stem and forms a prefixed verb: *va-næzæ* 'to organize' (lit. PREF-put) in the following example, where the meaning of the derived prefixed verb 'to organize' is different from that of the stem *næzæ* 'to put'.

(20) *jœk velœ=j mʊja ʊm=vanœz-e*
one time=INDEF date_palm_tree 1SG=organize.PST-PTCP
'once, I have organized the [bunches of] palm tree'
[AbdR_MjtXrshFG; Bashkar & Djouri]

Similarly, *va-* as a prefix with no clear semantic meaning can attach to *geretœ* 'to catch' and form *va-geretœ* 'to include, to encompass'. Dianat (2023, 123) has included some adverbs of manner like *ʤœjxa* 'immediately' as an NVE which can combine with LV *ʧedœ* 'to go' and form the LVC *ʤœjxa ʧedœ* 'to go immediately'. This account seems questionable, if we take into account that the adverb *ʤœjxa* 'immediately' can modify any other heavy verbs like *xœtœ* 'to sleep', *letœ* 'to pour', etc., which do not participate in LVCs.

Past participle + LV

Compound nouns formed by combining a past participle and a light verb (mainly *bʊzœ* 'to get/be') are mostly borrowed from Persian. The resulting compound verb is intransitive. Some examples are given in (21):

(21) a. *kʊʃtœ bʊzœ*
killed get
'to be killed' [Transitivity: -]
b. *ʤʊləʊ ʊftœzœ*
front fall
'to advance, to take over' [Transitivity: -]

4.2.2 Pseudo-incorporation

In this section we discuss the second type of compound verb formation, namely pseudo-incorporation. As Dabir-Moghaddam (1997, 41) argues about what he calls incorporation process, "the direct object loses its grammatical endings and incorporates with the [heavy] verb to create an intransitive compound verb which is a conceptual whole". The difference between combination and pseudo-incorporation is that the resulting compound verbs in the former may turn out to be transitive or intransitive whereas the compound-verb formations in the latter are fully intransitive. Furthermore, the compound verbs which occur from incorporation are semantically transparent while in noun-verb combinations, the [light] verb "is lexicalised to

serve as an aktionsart marker[11] which results in a compound verb with an idiomatic meaning" (Dabir-Moghaddam, 1997, 46–50). For instance, the compound verb *ʧaʃt xærdæ* 'to eat lunch' (lit. lunch eat) which is formed through pseudo-incorporation, has a transparent meaning, while a compound verb like *xʊssæ xærdæ* 'to grieve' (lit. grief eat) has a relatively idiomatic meaning which is more difficult to infer from its constituents.

Direct object

In the direct object type of incorporation in Persian, endings such as the postposition *-ra*, the indefinite marker *-i* or the plural marker suffix and possessive pronominal suffix are removed from the direct object and attach to the light verb to form a compound verb as shown in the examples in (22) from Dabir-Moghaddam (1997, 41–42) with some modifications in transcription and glossing rules:

(22) a. *bætʃe-ha ɣæza-ef-an-ra xʊrd-ænd*
child-PL food-3PC-PL-OM eat.PST-3PL
'the children ate their food'

b. *bætʃe-ha ɣæza xʊrd-ænd*
child-PL food eat.PST-3PL
'the children did food-eating'

Unlike Persian where a direct object loses its marker *-ra*, the compound verbs in Aheli can only lose the indefinite markers, possessive pronominals or plural suffixes. Example (23a) shows that the proclitic (3SG.POSS) is added to the noun *ʧaʃt* 'lunch' while the object is incorporated into the light verb in (23b) and forms the compound verb *ʧaʃt dæzæ* 'to give lunch' (lit. feed lunch). As mentioned by Dabir-Moghaddam (1997, 43), the compound verb after pseudo-incorporation constitutes "a conceptual whole" in which the noun is understood generically.

11 In Aheli, *jar zata* (lit. shouting hit) and *jar kesheza* (lit. shouting pull), have the same meaning: 'to shout'. The difference is that the light verb *zata* 'to hit' indicates 'disconnectedness, force and instantaneity of the action' while *kesheza* 'to pull' highlights 'the prolongation and duration of the action'. Arguing that the verbal element shows the mode of the activity, Dabir-Moghaddam (1997, 46–47) believes that light verbs show how the speaker conceptualises the manner of action. Therefore, it is unfounded to call the verbal element a 'light verb'. However, to conform to research tradition on Iranian languages, we will call it 'light verb' in the context of this research.

(23) a. *mʊ ʧaʃt=ʊʃ ʊm=dœ*
I lunch=3SG.PC 1SG=give.PST
'I gave [her] lunch'
[DjXzRsh; Mesa Chu Sha Buze]

b. *mʊ ʧaʃt ʊm=dœ œ hœɪfəʊ*
I lunch 1SG=give.PST to Haifow
'I gave lunch to Haifow'

Other examples of compound verbs formed by the pseudo-incorporation process (all intransitive) in Aheli are listed in (24) below:

(24) a. *entezar kefezœ*
waiting pull
'to wait' [Transitivity: -]

b. *əʊ-mivœ geretœ*
water-juice take
'to make fruit juice' [Transitivity: -]

c. *mʊja kœrezœ*
palm_tree plant
'to plant a palm tree' [Transitivity: -]

Persian and Aheli treat the grammatical markers similarly, i.e., the imperfective marker, subjunctive marker, and negation marker all attach to the LV. In the Persian example (25a) below, which is from Dabir-Moghaddam (1997, 51) with modifications in transcription style, the negation marker *nœ-* attaches to the light verb *kœrd* 'did'. Similarly, in the Aheli example (25b), the negation marker *ne-* attaches to the LV:

(25) a. *Ali Mina-ra delxʊr nœ-kœrd-Ø*
Ali Mina-OM annoying NEG-do.PST-3SG
'Ali did not annoy Mina'

b. *mʊ ʧaʃt ʊm=ne-dœ œ hœɪfəʊ*
I lunch 1SG=NEG-give.PST to Haifow
'I did not give lunch to Haifow'

However, the two languages exhibit different behaviour when pronominal clitics (PC) are used. Similar to other grammatical markers, Persian PCs functioning as subjects attach to the LV as shown in (26a) but moving the PC, *-œnd* '3PL', to the NVE makes the sentence ungrammatical as shown in (26b):

(26) a. *anha delxʊr=œm kœrd-œnd*
they annoying=1SG.OBJ do.PST-3PL
'They annoyed me'

b. **anha delxʊr=æm-ænd kærd*
they annoying=1SG.OBJ-3PL do.PST
'[Intended meaning] They annoyed me'

In contrast, subjectival PCs in Aheli can move and attach to the NVE. In the examples below, the pronominal clitic (PC) *=ʃu* '3PL', which functions as the subject typically attaches to the LV, as in (27a), but can move and join the NVE as shown in (27b) without any change in meaning:

(27) a. *ʊn-ja delxʊr ʃu=kerd=ʊm*
that-PL annoying 3PL=do.PST=1SG.OBJ
'They annoyed me'
b. *ʊn-ja delxʊr=ʃu kerd=ʊm*
that-PL annoying=3PL do.PST=1SG.OBJ
'They annoyed me'

Furthermore, as seen in the Persian example (26a), the direct object *=æm* '1SG' attaches to the NVE *delxʊr* 'annoying' whereas in the Aheli equivalent, the direct object *=ʊm* '1SG' joins the LV, as shown in (27). Moreover, the imperfective marker *-æ* in Aheli can be omitted when it occurs before a vowel-ending NVE, as shown below:[12]

(28) a. *mʊ mægæ əʊ-mivæ æ-ger-ʊm*
I always water-fruit IMPF-take.NPST-1SG
'I always make the fruit juice'
b. *mʊ mægæ əʊ-mivæ ger-ʊm*
I always water-fruit [IMPF]take.NPST-1SG
'I always make the fruit juice'

Dabir-Moghaddam (1997, 53) argues that in Persian, an adverb can modify the compound verb as a unit, but it cannot interpose between the elements of a compound verb whereas in Aheli, the placement of an adverb between the NVE and LV is permissible. As shown in (29) below, the adverb *ziyazi* 'abundantly' appears between the NVE *iraz* 'complaint' and the LV *gere* 'take':

(29) *betʃ mægæ iraz-e ziyazi ger-e*
child always complain-EZ abundantly [IMPF]take.NPST-3SG
'the child always complains a lot'

12 Although IMPF is optional in (28b), we put it in square brackets to show that the position of the prefix has been deleted.

Prepositional phrase

Sometimes a prepositional phrase combines with a light verb as a result of which the preposition is left out. In the examples below, the preposition *œli* 'on', which occurs in (30a), is omitted in (30b).

(30) a. *ʊn œli zemi ʃœst-Ø*
he on ground sit.PST-3SG
'He sat on the ground'
b. *ʊn zemi ʃœst-Ø*
he ground sit.PST-3SG
'He sat on the ground'

Contrary to Persian, where pseudo-incorporation is productive, the process is limited in productivity in Aheli, where most instances of compound verbs are formed through the combination process (Section 4.2.1).[13] However, similar to Persian, combination produces compound verbs which are more non-compositional in meaning compared to compound verbs made through pseudo-incorporation, which have more transparent meaning (compare *zemi ʃœssœ* 'sit down' (lit. ground sit) in (30b) with *pif vaxœrdœ* 'to be twisted' (lit. screw drink) in (11o). Here is another example where the preposition *œ* 'to' is left out in (31b):

(31) a. *Æli œ zemi ʊftœ-Ø*
Ali to ground fall.PST-3SG
'Ali fell to the ground'
b. *Æli zemi ʊftœ-Ø*
Ali ground fall.PST-3SG
'Ali fell to the ground'

5 Conclusion

This paper aimed to examine the derivation of the light verb constructions (LVCs) in Aheli, a dialect of Lari which is an Iranian endangered language. The LVCs are formed by a non-verbal element (NVE) followed by a light verb (LV) drawn from a small class of verbs such as *kerdœ* 'to make/do' (*ʤuri kerdœ* 'to harvest'), *zœtœ* 'to hit' (*tiva zœtœ* 'to separate the residues from crops'), *œndœ* 'to come' (*œ dʊnja*

13 This is in line with Dressler (2006, 23) who mentions that compounding is the most universal means of forming morphologically complex words across languages.

ændæ 'to be born'). It was shown that Aheli LVCs can be semantically divided into (a) change of state verbs (*huʃk vakerdæ* 'to dry out'), (b) activity verbs (*ʃəʊr kerdæ* 'to consult') and (c) instrument verbs (*kʊmba kerda* 'to combine the crops'). Moreover, it was shown that LVCs in Aheli can be formed through compounding or pseudo-incorporation. Compounding involves the use of a noun, an adjective, a prepositional phrase, an adverb, or a past participle followed by a light verb. In the case of pseudo-incorporation, a direct object loses its grammatical endings and is incorporated with a light verb as in the example *entezar keʃezæ* 'to wait' (lit. waiting pull). It has also been shown that some LVCs have compositional meaning, i.e., a modifier like an adjective can split the NVE and the LV. For instance, *ʃəʊr kerdæ* 'to consult' (thinking do) has a compositional meaning as we can add *æ ʧæp-ʊ-dəʊr* 'relatives' (lit. to left-and-round) between the constituents and form *ʃəʊr æ ʧæp-ʊ-dəʊr kerdæ* 'to consult with relatives' (lit. thinking left-and-round=EZ do). Conversely, an LVC like *ʃeri kerdæ* 'to put make-up' (lit. sweet make) does not permit the insertion of an adjective like *næk* between NVE and LV. Thus, a LVC like *ʃeri næk kerdæ* (lit. sweet good tmake) 'intended meaning: to put up good make-up on someone' is ungrammatical in Aheli.

Although there are some similarities between LVCs in Aheli and Persian with regard to semantic templates and compound verbs produced by combination, LVCs produced by pseudo-incorporation in Aheli behave differently. For instance, pronominal clitics (PCs) which function as a subject can only attach to the LV in Persian while they can be attached to both LV and NVE in Aheli, as shown in examples (27a) and (27b) above. Furthermore, in Persian, the same PCs attach to the NVE when they function as a direct object (26a) whereas in Aheli the PCs can attach to the LV as illustrated in examples (27a – 27b) above. The findings of this research contribute to our understanding of the word formation processes and morpho-syntactic properties of Aheli, as well as that of other Lari dialects (such as Evazi, Khonji, Bastaki, among others). More specifically, research into the LVC in Aheli contributes to the better understanding of complex predicates in Iranian languages. Nonetheless, there are some questions that will be left for future research, including, (a) which attributive modifications are allowed in Aheli LVCs?, and (b) whether the modification by an adverb (which has scope over the whole LVC) or an adjective (scoping over only the nominal NVE) derive the same meaning for the whole construction, as shown by Fleischhauer and Neisani (2020) for Persian LVCs?

In addition, the criteria which distinguish LVCs from pseudo-incorporation constructions (PICs) in Aheli are worth examining in future studies. It is hoped that this paper has provided a brief overview of the LVC in Aheli and some of the formal strategies used in the derivation of the complex predicate. A detailed study of the semantics and formal strategies of the construction and an examination of comparative data from other dialects of Lari will be useful to arrive at a more

comprehensive picture of this construction, a task which we hope to undertake in future research.

Bibliography

Assarian, Mina. 2016. *Bar-rasī-ye vīžegī-hā va ajzā-ye dastgāh-e fe'l dar zabān-e lārī be onvān-e zabān-e īrānī-ye dar xatar [The study of features and categories of verbal system in Lari language as an Iranian endangered language]*. Master's thesis: Payame Noor University, Tehran.

Bruening, Benjamin. 2015. Light Verbs are Just Regular Verbs. *U. Penn Working Papers in Linguistics* 21(1). 1–10.

Dabir-Moghaddam, Mohammed. 1997. Compound verbs in Persian. *Studies in the Linguistic Sciences* 27(2). 25–59.

Dabir-Moghaddam, Mohammed. 2013. *Typology of Iranian languages (Vol. 1)*. Tehran: The Organization for Researching and Composing University Textbooks in the Humanities.

Dianat, Leila. 2016. Assimilation process in the Lari language. *Journal of Iranian Language & Linguistics* 1(1). 53–72. 10.22099/JILL.2016.4014.

Dianat, Leila. 2023. Compound verbs in Lari language. *Journal of Iranian Language & Linguistics* 8(1). 117–131. 10.22099/jill.2023.46627.1337.

Dressler, Wolfgang. 2006. Compound types. In Gary Libben and Gonia Jarema (eds.), *The representation and processing of compound words*, 23–44. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199228911.003.0002.

Eghtedari, A. 1964. Lahje-ye fīšvarī [The Fishvar accent]. *The Journal of Farhang-e Iran Zamin* 1. 71–92.

Eghtedari, A. 1992. *Lārestān-e kohan va farhang-e Lārestanī [The ancient Larestan and Larestani glossary]*. Tehran: Jahan Moaser Publication.

Eghtedari, A. 2005. *Zabān-e lārestānī: Jostār-ī dar zabān va farhang-e mardom [The Larestani language: A research in people language and culture]*. Tehran: Hamsaye Publication.

Eshaghi, Mahdie and Gholamhosein Karimi Doostan. 2023. A Corpus of Persian Light Verb Constructions. *Journal of Researches in Linguistics* 14(1). 173–198. 10.22108/JRL.2023.135758.1685.

Fleischhauer, Jens. 2020. Predicative multi-word expressions in Persian. In *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, 552–561. Hanoi: Association for Computational Linguistics. https://aclanthology.org/2020.paclic-1.63.

Fleischhauer, Jens and Mozhgan Neisani. 2020. Adverbial and attributive modification of Persian separable light verb constructions. *Journal of Linguistics* 56(1). 45–85. https://doi.org/10.1017/S0022226718000646.

Kalbasi, I. 1990. Sāxt-vāže dar gūyeš-e lār [The morphology of the Lari dialect]. *The Journal of Farhang* 6. 189–198.

Kamioka, Kōji and M. Yamada. 1979. *Lārestāni studies 1, Lari basic vocabularies*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Karimi, Simin. 1997. Persian Complex Verbs: Idiomatic or Compositional? *Lexicology* 3(1). 273–318.

Karimi Doostan, Gholamhossein. 2005. Light verbs and structural case. *Lingua* 115. 1737–1756. https://doi.org/10.1016/j.lingua.2004.08.002.

Khonji, L. 2009. *Negareš-e tafsīl-ī bar zabān-e lārestanī va gūyeš-e xonjī [Comprehensive study of the Larestani language and the Khonji dialect]*. Shiraz: Ilaf Press.

Korn, Agnes. 2013. Looking for the middle way: Voice and transitivity in complex predicates in Iranian. *Lingua* 135. 30–55. https://doi.org/10.1016/j.lingua.2013.07.015.

Korn, Agnes. 2017. Notes on the nominal system of Bashkardi. *Transactions of the Philological Society* 115(1). 79–97. https://doi.org/10.1111/1467-968X.12087.

Mahmoodian, Saeed. 2007. *Encyclopedia Larestanica*. Richmond, VA: Brandylane Publishers.

Megerdoomian, Karine. 2004. A semantic template for light verb constructions. In *Proceedings of the First Workshop on Persian Language and Computer*, 25–26. Tehran: Tehran University Press.

Megerdoomian, Karine. 2012. The status of the nominal in Persian complex predicates. *Natural Language & Linguistic Theory* 30(1). 179–216. https://doi.org/10.1007/s11049-011-9146-0.

Molčanova, E. 1982. Gūyeš-hā-ye lārī [The dialects of Lari]. *Journal of Nemeye Farhangestan* 5(2). 183–187.

Monshizadeh, Mojtaba, Ehsan Changizi and Babak Sharid. 2022. The endless story of "compound verbs": An introduction to conceptology and terminology of complex verbs in Persian. *Journal of Researches in Linguistics* 13(2). 217–236. 10.22108/jrl.2022.131148.1611.

Moridi, B. 2006. Bar-rasī-ye tābu va hosn-e ta'bīr dar čand vāže az gorūh-e gūyeš-hā-ye lārī [Analysis of taboo and euphemism in some words in a group of the Lari dialects]. *The Journal of Persian Language & Literature* 2(5). 61–72.

Moridi, B. 2007. Bar-rasī va moqāyese-ye gūyeš-e lārī bā daštestānī [Analysis of the Lari dialect and comparison with Dashtestāni dialect]. *Orientalia* 4. 64–72.

Nunberg, Geoffrey, Ivan A. Sag and Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538. https://doi.org/10.1353/lan.1994.0023.

Ourang, Muhammed. 2022. *A reference grammar of Aheli: A dialect of the Lari language, Iran*. Doctoral dissertation: The University of New South Wales, Sydney.

Pantcheva, Marina. 2008. Noun preverbs in Persian complex predicates. *Tromsø Papers on Language & Linguistics: Nordlyd* 35. 19–45.

Rashidi, Ali. 2008. *Bar-rasī-ye sāxtārī-ye gūyeš-e būčīrī [A Structural study on Buchiri dialect]*. Master's thesis: The University of Sistan & Baluchestan, Zahedan.

Salami, A. 2004–2007, 2009, 2011. *Ganjīne-ye gūyeš-šenāsī-ye fārs [The treasury of the dialectology of Fars] (Vols. 1-6)*. Tehran: Iranian Academy of Persian Language and Literature.

Sheintuch, Gloria. 1976. Periphrastic verb formation in Persian: a dynamic process. In W. Christie (ed.), *Current Progress in Historical Linguistics*, 137–146. Amsterdam: North-Holland Publishing.

Skjaervo, Prods O. 1989. Languages of southeast Iran: Lārestānī, Kumzārī, Baškardī. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 363–369. Wiesbaden: Ludwig Reichert Verlag.

Stilo, Donald. 2012. Case in Iranian: From reduction and loss to innovation and renewal. In Andrej L. Malchukov and Andrew Spencer (eds.), *The Oxford Handbook of Case*, 700–715. Oxford: Oxford University Press. 10.1093/oxfordhb/9780199206476.013.0049.

Telegdi, Zsigmond. 1951. Nature et fonction des périphrases dites 'verbes composés' en Persan. *Acta Orientalia Academiae Scientiarum Hungaricae* 1(2/3). 315–338.

Jens Fleischhauer and Claudius Patrick Kihara

Light verbs and aspectual semi-auxiliaries in Gĩkũyũ

1 Introduction

Gĩkũyũ (E51[1]) is a Bantu language spoken in Central Kenya. The language shows many properties typical of Bantu languages: at the clausal level, it is a morphologically rich head-marking language (cf. Nichols, 1986). It can express up to two arguments by bound argument markers prefixed to the verbal stem. The language has a complex tense-aspect system mostly expressed by verbal pre- and suffixes. In addition, the language has a set of verbal suffixes – usually called 'extensions' in the Bantuist literature – expressing voice, i.e., causative, passive, and applicative.

Although the language is morphologically rich, it makes frequent use of periphrastic predicative constructions of the type 'Verb + Noun'. The noun can either be a lexical noun or an infinitive. Although infinitives show some properties of verbs (e.g., they license certain verbal affixes), they also behave like nouns and should be considered deverbal nouns (cf. Mugane, 2003). Infinitives are derived by adding the infinitive marker *kũ-* 'noun class 15' to a verbal stem. In predicative 'noun + verb' constructions, the verb forms a predicational unit with a noun which can be characterized as a complex predicate. Examples are shown in (1).[2] All these constructions have in common that the eventuality, i.e., the state or event denoted by the predicate, is introduced by the noun rather than the verb.[3] The complex predicate in (1a) is an instance of a prospective aspect construction and can be paraphrased as 'it is about to rain'. The raining event is contributed by the verbal noun *kũura* 'rain'. The example in (1b) is a phasal verb construction which denotes the beginning of an

1 'E51' is the so-called Guthrie number, a mainly geographically-based classification system of the Bantu languages proposed by Guthrie (1967) and updated by Maho (2003).

2 Complex predicates in Gĩkũyũ are not restricted to the types exemplified in (1) but due to reasons of space we restricted ourselves to these constructions. The Gĩkũyũ examples are derived from the introspection of the second author and have been validated with the institutions of other native speakers.

3 We use the term 'eventuality' in the sense of Bach (1986) as a cover term for both states and events.

Jens Fleischhauer, Heinrich-Heine-University, Düsseldorf, Germany
Claudius Patrick Kihara, Chuka University, Chuka, Kenya

https://doi.org/10.1515/9783112229330-004

event of singing. Finally, in (1c) the complex predicate 'beat phone' is paraphrased as 'make a call'.

(1) a. *Nĩ=kũ-r-end-a kũ-ur-a.*
FOC-17-PRS-want-FV 15-rain-FV
'It is about to rain.' (lit. It wants to rain.')[4]
b. *Ma-kĩ-rik-a kũ-in-a.*
2-NARR-start-FV 15-sing-FV
'And they started to sing/singing.'
c. *Mũ-tumia nĩ=a-ra-hũr-a thimo.*
1-woman FOC-1-PRS-beat-FV 9.phone
'The woman is making a call.' (lit. The woman is beating the phone)

At first sight, it looks as if the three constructions in (1) have a lot in common. However, we argue below that the construction in (1c) is grammatically closer to regular predicate-argument constructions like in (2) than to the two complex predicates in (1a) and (1b).

(2) *Mũ-tumia nĩ=a-ra-hũr-a ngui.*
1-woman FOC=1-PRS-beat-FV 9.dog
'The woman is beating the dog.'

The example in (1c) is an instance of a light verb construction (LVCs), the two examples in (1a) and (1b) represent semi-auxiliary constructions. On the basis of grammatical data, we argue that LVCs cannot be identified on the basis of their grammatical properties, whereas semi-auxiliary constructions show grammatical restrictions – compared to corresponding regular predicate-argument constructions – which are the result of an ongoing auxiliarization process. The data at hand do not present any evidence for light verbs undergoing a similar process of auxiliarization.

The paper is structured as follows: In section 2, we introduce the relevant grammatical background on Gĩkũyũ simplex verbs. For illustration purposes, we use the verb *hũra* 'beat' which has both a light use as in (1c) as well as a heavy use in (2). Section 3 introduces light verb constructions headed by *hũra*. The focus will be on the question of whether we can identify any grammatical differences between *hũra*'s light and heavy uses. The discussion reveals that the two systematically differ in only

4 There is some debate concerning the morphological as well as functional status of *nĩ*. Functionally, it is either analyzed as being an assertion marker and/or a focus marker (cf. Clements, 1984; Bergvall, 1987; Schwarz, 2003; Kihara, 2016, a.o.). Morphologically, it is either treated as a free particle, an affix or a clitic (cf. Barlow, 1960 and the above-mentioned literature). As the exact analysis of this element is not relevant for the current analysis, we adopt a uniform representation as a focus clitic.

one respect: the nominal element of an LVC cannot be pronominalized by a bound argument marker whereas the object argument of 'heavy' *hũra* can. In section 4, we turn to the prospective aspect construction illustrated in (1a). The central question to be asked is whether we can identify grammatical differences between *enda*'s use as a lexical full verb meaning 'want' and its occurrence within a prospective construction. This time, the comparison reveals stricter differences between the two uses. The phasal verb construction illustrated in (1b) is the subject of section 5. The constructions verbal element *rika* is also used as a lexical full verb meaning 'get into'. A comparison of the two uses reveals the same differences as identified between *enda*'s two uses. Section 6 presents a systematic discussion of the grammatical properties of the verbs and their nominal elements. These differences allow separating light verbs from semi-auxiliaries, as we term the relevant non-full verb uses of *enda* and *rika*. A cross-linguistic comparison of the Gĩkũyũ *hũra*-LVCs with light verb constructions in other languages is undertaken in section 7. The section is split into two subsections. In 7.1 we compare the Gĩkũyũ LVCs with similar LVCs in the genetically related language Swahili. Section 7.2 presents a brief comparison of some specific grammatical properties – direct object properties of the LVC's nominal element – between Gĩkũyũ and the Iranian language Persian. The paper ends with a conclusion in section 8.

2 Simplex verbs in Gĩkũyũ

Before we turn to a discussion of the different types of complex predicates mentioned in the introduction, we start with a brief discussion of simplex verbs. The reason is twofold: First, we need to distinguish complex predicates from simplex predicates – which is especially relevant for light verb constructions – and second, the discussion of simplex predicates allows introducing the relevant morphosyntactic background on Gĩkũyũ verbs. In the following discussion, we restrict ourselves to those aspects of verbal morphology which are relevant for our discussion and leave out other aspects, for instance reciprocal marking, as they are irrelevant for the discussion at hand.

Gĩkũyũ is head-marking at the clausal level (Nichols, 1986), i.e., the relation between the verb – the head of the clause – and its arguments – the dependent elements – is indicated by bound argument markers on the verb. The only obligatorily realized bound argument marker is the one realizing the subject argument. It can co-

occur with a coreferential independent subject referential phrase (RP).[5] However, a bound object marker cannot co-occur with an independent object RP within the same clause (Kihara, 2016, 55, Fleischhauer, 2023a, 166).[6] This is illustrated by the example in (3). The third person singular subject is realized by the bound argument marker *a-*, an independent pronoun or subject RP is not necessary. The only information given is that the subject referent belongs to noun class 1, e.g., the referent introduced by the noun *mũtumia* 'woman' in the preceding discourse (cf. example (2)). The object is realized by the bound argument marker *mĩ-* indicating an object argument of noun class 9. The optional independent RP *ngui* 'dog' is separated by comma intonation form the preceding sentence. Thus, it is right-dislocated and positionally realized outside of the clause (cf. Lambrecht, 2001 for the notion of dislocation).

(3) *Nĩ=a-ra-mĩ-hũr-a, ngui.*
FOC-1-PRS-9-beat-FV 9.dog
'S/he is beating it, the dog.'

The bound object marker is in complementary distribution to the reflexive prefix *ĩ-* (4). This marker is invariant and does not change depending on the subject argument's noun class.

(4) *Nĩ=a-ra-ĩ-hũr-a.*
FOC=1-PRS-REFL-beat-FV
'S/he is beating herself/himself.'

Gĩkũyũ has an elaborate tense and aspect system. In the temporal system, the language has a basic 'past – present – future' distinction which is accompanied by so-called remoteness distinctions in the past and future tense (cf. Cable, 2013). Whereas tense is realized by prefixes, aspect is realized by suffixes. The aspectual system is less elaborate than the tense system, however the language possesses three aspectual suffixes: *-ir* 'perfective', *-ag* 'imperfective' and *-ĩt* 'perfect'.

Finally, Gĩkũyũ has a rich inventory of voice affixes which operate on a verb's argument structure. Applicative and causative marking add an additional argument, passive and medium reduce one argument. Passive voice is expressed by adding *-o*

5 The notion of a referential phrase (RP) is taken form Role and Reference Grammar (RRG, Van Valin, 2008) and compensates the notion of a nominal phrase (NP) or determiner phrase (DP) used in other syntactic frameworks.

6 We adopt Van Valin's (2023, 67) analysis of bound argument markers in head-marking languages and assume that markers which can co-occur with a coreferential RP function as pronominal anaphora, whereas does which cannot function as pronouns.

as a final suffix, replacing the verb's final vowel (5a). Applicative (5b) and causative (5c) are realized by the (non-final) suffixes *-ĩr* and *-ithi* respectively.

(5) a. *Ngui nĩ=ĩ-ra-hũr-ir-wo nĩ mũ-tumia.*
9.dog FOC=9-PST-beat-PFV-PASS by 1-woman
'The dog was beaten by the woman.'
b. *Mũ-tumia nĩ=a-ra-hũr-ĩr-a nyina ngui.*
1-woman FOC=1-PRS-beat-APPL-FV 1.mother 9.dog
'The woman is beating the dog for the mother.'
c. *Mũ-thuri nĩ=a-ra-hũr-ithi-a mũ-tumia ngui.*
1-man FOC=1-PRS-beat-CAUS-FV 1-woman 9.dog
'The man caused the woman to beat the dog/ The man helped the woman the beat the dog.'

Although Gĩkũyũ licenses maximally one bound argument marker, this marker can refer to any argument. It can be one of the verb's direct arguments – as in (3) – or an applied argument, i.e., an additional argument introduced by an applicative marker, as in (6). Thus, Gĩkũyũ is a 'symmetric object language' as any non-subject argument can show direct object properties (Fleischhauer, 2023a, 166).[7]

(6) *Mũ-tumia nĩ=a-ra-mũ-hũr-ĩr-a ngui.*
1-woman FOC=1-PRS-1-beat-APPL-FV 9.dog
'The woman is beating the dog for him/her.'

In the remainder, we use to properties for identifying (direct) object arguments in Gĩkũyũ.[8] First, a direct object argument can be realized by a bound object marker (cf. (3) and (6)), and, second, a direct object argument can become the subject under passivization (5a). We will show below that these two properties are crucial for distinguishing LVCs from (semi-)auxiliary constructions.

3 Light verb constructions

As already mentioned in section 1, the verb *hũra* 'beat' is used both as a lexically full verb as well as a light verb. In its light use, *hũra* combines with different nouns to form a complex predicate. A few examples are listed in (7). Similar light verb

7 For the distinction between symmetric and asymmetric object languages, see, for instance, (e.g. Bresnan and Moshi, 1990).

8 The standard criteria for identifying object arguments in Bantu have been introduced by Hyman and Duranti (1982).

constructions are attested in, for instance, Swahili (cf. Krifka, 1995, 1401, Olejarnik, 2009, 2011; Martin, 2019) and Digo (Nicolle, 2023, 125). We present a brief comparison of Gĩkũyũ LVCs headed by *hũra* with Swahili *piga*-LVCs in section 7.1.

(7) *hũra nguo* 'to wash clothes' (lit. beat clothes), *hũra njĩra* 'to walk on a route/ path' (lit. beat path), *hũra mbica* 'to take a photo' (lit. beat photo), *hũra riboti* 'to report an incident' (lit. beat report), *hũra bathi* 'to iron clothes' (lit. beat iron box), *hũra thimu* 'to make a call' (lit. beat phone), *hũra kanua* 'to click' (lit. beat mouth), *hũra ndore* 'to fart loudly' (lit. beat fart), *hũra kiisi* 'to kiss' (lit. beat kiss)

Similar to its heavy use, light *hũra* requires the realization of two referential phrases: the subject argument and the phrasal element contributing the main predicational content. A telling example is the LVC *hũra thimũ* which literally means 'beat phone'. Although a literal interpretation is possible, the LVC is interpreted as 'to phone'. The noun refers to an artefact – a phone – which gives rise to event inference on the basis of the noun's meaning. Artefacts are designed for specific purposes and in the case of the noun 'phone' – and its correspondent in Gĩkũyũ – the referent is designed to call someone. Therefore, the noun licenses inference to an event of calling someone/phoning. At this stage, we do not go further into the semantics of the light verb construction but we would like to highlight that the nominal element contributes to the complex predicate's meaning. Crucially, the noun does not refer to an object but contributes the denoted eventuality (for further discussion of this issue see the discussion of non-eventive nouns in the introduction to the volume). Although the nominal element forms a semantically complex predicate with *hũra* 'beat', the light verb constructions does not show any grammatical differences to the regular predicate-argument construction discussed in the previous section.

As the examples in (8) show, it is possible to have a bound argument marker coreferential with the noun *thimũ*. Crucially, the noun has to be right-dislocated as discussed above. The form of the noun does not change depending on number but the number distinction is reflected in the choice of the bound argument marker. As the contrast between (8a) and (8b) indicates, nominal number affects the interpretation. The plural noun introduces a pluralic event (many calls), whereas the singular noun allows either for a habitual interpretation or an excessive one.

(8) a. *Mũ-tumia nĩ=a-ra-i-hũr-a, thimũ.*
1.woman FOC=1-PRS-10-beat-FV 10.phone
'The woman is making many calls./ The woman is calling many numbers.'

b. *Nĩ=a-ra-mĩ-hũr-a, thimũ.*
FOC=1-PRS-9-beat-FV 9.phone
'S/he is (really) calling the phone regularly./ He is calling extensively.'

Since the light verb forms a tight semantic unit with the nominal element, the noun cannot be replaced by a reflexive marker. Reflexive marking presupposes applicative marking, the reflexive marker instantiates the applied argument. But this is the only grammatical restriction on LVCs observed in our study.

(9) *Mũ-tumia nĩ=a-ra-ĩ-hũr-ĩr-a thimũ.*
1.woman FOC=1-PRS-REFL-beat-APPL-FV 9.phone
'The woman is calling herself/ dialing her own number. / The woman is calling other people herself.'

As Gĩkũyũ is a symmetrical object language, a bound object marker can also be licensed by an applied argument. This is shown in (10) which illustrates the combination of the LVC with the applicative marker *-ĩr*. The applicative marker introduces the recipient of the call, i.e., the person who is called, or a beneficient of the action. It could be realized with an independent RP or, as in (10), by a bound object marker.

(10) *Mũ-tumia nĩ=a-ra-mũ-hũr-ĩr-a thimũ.*
1.woman FOC=1-PRS-1-beat-APPL-FV 9.phone
'The woman is calling him/her. / The woman is calling on his/her behalf.'

The LVC *hũra thimũ* 'to call' only allows the applied object to become the subject under passivization (cf. 11). Thus, in this specific case does passivization presupposes applicative marking.

(11) *Mũ-tumia nĩ=a-ra-hũr-ir-wo thimũ nĩ nyina.*
1.woman FOC=1-PRS-beat-APPL-PASS 9.phone by 1.mother
'The woman is being called by her mother.'

We also regularly find a possibly partially lexicalized usage, in which the recipient of the call can be realized by a bound argument marker, while the noun *thimũ* 'phone' is not realized (12).

(12) *Nĩ=n-gũ-kũ-hũr-ĩr-a.*
FOC=1-FUT-1-beat-APPL-FV
'I will call you.'

However, *hũra thimũ* is not representative for all *hũra*-LVCs as the examples in (13) show. The examples illustrate passivization of the LVC *hũra baĩni* 'to fine' (lit. beat fine). In (13a), an active use of the LVC is shown. The sentence contains two non-subject RPs: *baĩni* 'fine' and *mũndũ* 'person'. Thus, the syntactic valency is – compared to *hũra*'s use as a lexical full verb – increased by one. Both postverbal RPs can become subject under passivization, as the examples in (13b) and (13c) show.

(13) a. *I-goti nĩ=ri-a-hũr-ir-e mũ-ndũ baĩni.*
5-court FOC=5-PST-beat-PFV-FV 1-person 9.fine
'The court fined the person.'
b. *Mũ-ndũ nĩ=a-a-hũr-ir-wo baĩni nĩ i-goti.*
1-person FOC=1-PST-beat-PFV-PASS 9.fine by 5-court
'The person was fined by the court.'
c. *Baĩni y-a-hũr-ir-wo mũ-ndũ nĩ i-goti.*
9.fine 9-PST-beat-PFV-PASS 1-person by 5-court
'The fine was charged on the man by the court.'

Although *baĩni* has one object property, it cannot be pronominalized by a bound argument marker. *Mũndũ*, on the other hand, can be pronominalized as (14) shows.

(14) *I-goti nĩ=ri-a-mũ-hũr-ir-e baĩni.*
5-court FOC=5-PST-1-beat-PFV-FV 9.fine
'The court fined him/her.'

This brief comparison of the two LVCs revealed LVC-specific restrictions concerning passivization. However, it also proofs that the nominal element shows (at least some) direct object properties. In addition, we observe that the LVCs syntactic valency does not (necessarily) coincide with *hũra*'s valency. As a lexical full verb, *hũra* takes two arguments – one subject and one object argument (e.g. (2)) – but there are two non-subject RPs in (13) without the verb undergoing any valency increasing operation. Although the question how the argument structure of an LVC is created is a highly relevant grammatical issue, we will not address it in the current paper as our focus is on other grammatical properties.

With respect to the other instances of verbal morphology – tense, aspect and causative marking – light *hũra* does not differ from its heavy use. We will not illustrate this for the complete paradigm of tense and aspect forms but restrict ourselves to the one example in (15a). In contrast to the preceding examples, the verb bears future tense marking. The example in (15b) shows the morphological causative *-ithi*[9] on the light verb *hũra*. The causative marker introduces a new argument – the causer –

9 The causative suffix *-ithi* is bimorphic as evidenced by the fact that it is not necessarily realized adjacently, i.e., the perfective aspect suffix is positioned between *ith* and *i*. Diachronically, *-ithi* is a combination of causative *-ith* and the transitivizer *-i* (Good, 2005, 10–11, Good, 2007, 208–209. Synchronically, *-i* does not make an independent semantic contribution in examples like (10). However, it functions as a transitivizer – also called 'short causative' – in other contexts. In examples like (15b), we simply gloss the discontinuous part of the morpheme as 'DC' – discontinuous element – and conceive it (synchronically) as being a part of final part of the bimorphic ('long') causative *-ithi*.

and demotes the previous subject argument to direct object status. In (15), the agent – the one calling – is expressed by a bound object marker.

(15) a. *Mũ-tumia nĩ=a-kũ-hũr-a thimũ.*
1-woman FOC=1-FUT-beat-FV 9.phone
'The woman will call.'

b. *A-tũ-hũr-ith-ir-i-e thimũ ci-a tũhũ.*
1-2-beat-CAUS-PFV-DC-FV 10.phone 10-ASSOC useless
'S/He had made us make useless calls.'

To summarize: Although the empirical basis is rather limited, we do not find any general grammatical differences between a verb's heavy use and its use as a light verb. The most promising property – which requires further empirical validation on the basis of a broader data sample – is the that the nominal element does not always show direct object properties.

4 Prospective constructions

Gĩkũyũ – like many other languages (cf. Heine, 1994) – uses periphrastic constructions of the type 'verb + noun' for the expression of prospective aspect.[10] Prospective aspect relates the subject referent's current state to a subsequent eventuality (e.g., Comrie, 1976, 52). In (16), for instance, the referent of *ngaari* 'car' is in a state preceding a crash. The actual realization of this eventuality is – as discussed above – not entailed.

Gĩkũyũ expresses prospective aspect by a combination of the verb *enda* 'want' and an infinitive. Infinitives are – as already mentioned above – derived by adding a marker of noun class 15 to a verbal stem. The verbal stem minimally consists of the lexical root – *-gũ-* – in (16) and the final vowel *-a*. However, the infinitive can bear further verbal markers like the bound object marker or voice markers. The marker for class 15 is a regular bound subject marker and the infinitives behave in all relevant aspects like nouns (cf. Mugane, 2003). The prospective construction in (16) resembles the LVCs discussed in the previous section as the verb and its non-subject argument form a predicational unit.[11]

10 For an analysis of (aspectual) auxiliary + verb constructions in the three Bantu languages Swahili, Rangi, and SiSwati, see Gibson and Marten (2016).

11 The German prospective '*stehen vor* NP'-constructions (lit. stand in front of NP) similarly resembles LVCs within the languages. The resemblance is even greater than in Gĩkũyũ as *stehen* 'stand' is also attested as a light verb in other LVCs (cf. Fleischhauer, 2023c). However, desiderative verbs seem

(16) *Ngaari ĩ-kũ-end-ag-a kũ-gũ-a.*
9.car 9-PST-want-IMPF-FV 15-fall-FV
'The car was about to crash.' (lit. The car wanted to fall.)

The verb *enda* is also used as a regular full verb expressing "a desire that the complement proposition be realized" (Noonan, 2007, 132). In its use as a desiderative predicate – illustrated in (17) –, *enda* translates as 'want'. We refer to this use as the 'desiderative use', resp. 'desiderative interpretation' of *enda*. As a full verb, *enda* can take a verbal complement. *Ũine*, in (17), is a verb bearing pronominal subject marking. The verb's final vowel is *-e* which is, among other things, used in subjunctive contexts. However, the final vowel cannot be analyzed as a subjunctive marker as it occurs in other contexts as well, i.e., perfective aspect forces the final vowel to be *-e* rather than *-a*.

(17) *Nĩ=nd-ĩr-end-a ũ-in-e.*
FOC-1SG-PRS-want-FV 2SG-sing-FV
'I want you to sing (right now).'

(Wittke, 2015, 80)

The example in (17) has unambiguously a desiderative interpretation: the speaker wants the addressee to sing. Other sentences, like the one in (18), are ambiguous between a desiderative and a prospective interpretation. The sentence can either mean that the subject referent wants to go to work or that the subject referent is about to go to work. In both cases, it does not follow that the subject referent will go to work. In the first case, it is just the subject referents wish to go to work. Whereas in the second case, it might be expected that – if nothing intervenes – the subject referent is close to going to work. It is not contradictory to say: I was about going to work but I did not go since I broke my leg while putting on my shoes.

(18) *Tũ-r-end-a gũ-thiĩ wĩra.*
2-PRS-want-FV 15-go 9.work
(i.) 'They want to go to work.'
(ii.) 'They are about going to work.'

Evidence that we are dealing an aspectual prospective construction[12] rather than a periphrastic (immediate) future tense construction is that fact that the finite verb

to be more frequently attested as the source of a prospective construction than spatial expressions (cf. Heine, 1994, 44).

12 We adopt Dik's (1997, 221) division between 'aspect proper' and 'relational aspect'. 'Aspect proper' covers perfective as well as imperfective aspect, whereas 'relational aspect' comprises the two notions of 'perfect' and 'prospective'.

can be marked for past tense to refer to a prospective situation in the past. In (19), it is expressed that at some point in the past, it was about to rain. Note that the noun class marker for class 15 and 17 are homophonous; whereas noun class 15 realized nominalized infinitives, 17 is a locational noun class deriving nouns referring to a location 'here'.

(19) *Nĩ=kũ-r-end-ag-a kũ-ur-a ira.*
FOC=17-PST-want-IMPF-FV 15-rain-FV yesterday
'It was about to rain yesterday.'

The sentences in (16) and (19) are unambiguous examples of a prospective interpretation of *enda* as an inanimate subject referent cannot have any wishes. Whereas (19) has a locational subject argument – the referent of the bound subject marker is interpreted as 'here' (noun class 17) –, the one in (16) has a concrete subject referent. Thus, a desiderative interpretation for *enda* is definitely ruled out in these cases (exceptions are fictional contexts like, for instance, in the Pixar movie *Cars*).

As a verbal component of the prospective construction, *enda* exhibits stronger restrictions regarding its complement than in its use as a full verb. In the prospective construction, the subject argument of *enda* has to be coreferential with the (unrealized) subject of its infinitival complement (cf. Fleischhauer and Gamerschlag, 2019, 146). This is different for the desiderative use of *enda* as the example in (17) shows. The subject of *enda* is different from the subject of its complement clause. A consequence of the coreferentiality restriction is that in its prospective function, *enda* only combines with infinitives but not – unlike *enda*'s desiderative interpretation (cf. (17)) – with finite verbs. However, desiderative *enda* does not obligatorily take complement clauses but can also take RP complements (20). As (20b) shows, the object of the verb can be realized by a bound object marker (the RP *ngaari ĩno* 'this car' is left-dislocated and separated by an intonation break from the remaining clause).

(20) a. *Ndĩ-r-end-a ngaari ĩ-no.*
1SG-PRS-want-FV 9.car 9-DEM
'I want this car.'
b. *Ngaari ĩ-no, nĩ=ndĩ-ra-mĩ-end-a.*
9.car 9-DEM FOC=1SG-PRS-9-want-FV
'This car, I want it.'

Pronominalization indicates that the non-subject argument bears direct object properties. It is therefore not surprising that passivization is possible as well (21). In this case, passivization does not reduce syntactic valency but affects the semantic role of the subject argument. Whereas in the active sentence the subject of *enda* is the one

having a wish, passivization makes the subject argument the logical object of the wishing, i.e., the one of which it is wished that s/he is singing.

(21) *Ndĩ-r-end-wo ny-in-e.*
1SG-PRS-want-PASS 1SG-sing-FV
'I am expected to sing.' (lit. I am wanted to sing)

Passivization is even possible if desiderative *enda* takes an infinitival complement (22). The bound object marker *kw-* is coreferential with the right-dislocated infinitival clause *kũina* 'singing'.

(22) *Kũ-in-a n-di-kw-end-ĩt-e.*
15-sing-FV 1-NEG-15-want-PERF-FV
'Singing, I don't want/like it.'

Whereas *nyine* in (21) and *kũina* in (22) qualify as clausal complements of (desiderative) *enda*, *kũgũa* 'to fall' – in the prospective construction in (19) – does not . Noonan gives the following characterization of 'complementation': "By complementation, we mean the syntactic situation that arises when a notional sentence or predication is an argument of a predicate. For our purposes, a predication can be viewed as an argument of a predicate if it functions as the subject or object of that predicate" (Noonan, 2007, 52). *Kũgũa* 'to fall' (in (16)) does not bear any object properties, especially it cannot be pronominalized by a bound object marker (23). Thus, the combination of *enda* and *kũgũa* is not an instance of complementation.

(23) **Kũ-gũ-a, ngaari ĩ-gũ-kũ-end-ag-a.*
15-fall-FV 9.car 9-PST-15-want-IMPF-FV
Intended: 'The crash, the car was about to do it.'

Furthermore, the prospective interpretation is lost under passivization as the example in (24) shows. *Ngaari* 'car' becomes subject of the sentence, *gũa* 'fall' is not realized as an infinitive but bears a bound subject marker for noun class 9 which is coreferential with *ngaari*. Crucially, the sentence does not have a prospective interpretation but is interpreted as 'the subject is expected to do' which is that the same interpretation we observed for the desiderative interpretation of *enda* in (21). We conclude that the prospective construction is incompatible with passive voice in Gĩkũyũ.

(24) *Ngaari ĩ-kũ-end-ag-wo ĩ-gũ-e.*
9.car 9-PST-want-IMPF-PASS 9-fall-FV
'The car was expected/required to crash.'

When it comes to other voice morphology – causative, applicative – we observe that neither desiderative nor prospective *enda* license these markers. Therefore, we cannot say that the two uses of *enda* differ with respect to this property.

The prospective construction shows an interesting interaction with aspect. As already mentioned, prospective aspect does not entail the actual realization of the prospective eventuality. Taking (25) as an example, it is possible that it actually rained. Thus, can be continued with *na nĩ=kũraurire* 'and (indeed) it rained'.

(25) *Ira nĩ=kũ-r-end-ag-a kũ-ur-a.*
yesterday FOC=15-PST-want-IMPF-FV 15-rain-FV
'Yesterday, it was about to rain.'

If we switch aspect from the imperfective to perfective (26a) or perfect (26b), adding *na nĩ=kũraurire* 'and it rained' results in a contradiction. Thus, perfect(ive) enforces a non-realizational interpretation of the construction.

(26) a. *Ira nĩ=kũ-r-end-ir-e kũ-ur-a.*
yesterday FOC=15-PST-want-PFV-FV 15-rain-FV
'Yesterday, it was about to rain.'
b. *Ira nĩ=kũ-r-end-ĩt-e kũ-ur-a.*
yesterday FOC=15-PST-want-PERF-FV 15-rain-FV
'Yesterday, it was about to rain.'

More specifically, in the context of past perfect(ive), the prospective construction has an avertive interpretation. Kuteva et al. (2019, 858) dub a linguistic construction 'avertive' if is refers to "a verb situation which was on the verge of taking place but did not take place" (similarly Kuteva, 2001, 78). In Gĩkũyũ, an avertive interpretation results from a specific combination of tense (past) and aspect (perfect(ive)) and is, interestingly, not restricted to the prospective construction. An avertive interpretation also arises with desiderative *enda* as (27) shows. The examples show a conjunction of two main clauses. The first clause states that the subject referent wanted to buy a car, whereas the second conjunct expresses that s/he finally did so. The conjunction of the two sentences is only licit if *enda* bears imperfective aspect marking (27a). If *enda* is marked for perfect aspect, adding the sentence *na nĩ=ndagũrire* 'and I bought it' is odd (27b). Thus, realization of the eventuality expressed by the infinitive serving as the complement of *enda* – irrespective whether the verb is used as a verb of desire or within a prospective construction – is prohibited in a past perfect(ive) context.

(27) a. *Nd-e-end-ag-a kũ-gũr-a ngaari na nĩ=nd-a-gũr-ir-e.*
1-PST-want-IMPF-FV 15-buy-FV 9.car with FOC=1-PST-buy-PFV-FV
'I wanted to buy a car and I bought one.'

b. *Nd-e-end-ĩt-e kũ-gũr-a ngaari #na nĩ=nd-a-gũr-ir-e.*
1-PST-want-PERF-FV 15-buy-FV 9.car with FOC=1-PST-buy-PFV-FV
'I wanted to buy a car and I bought it.'

The avertive interpretation is not a specific property of *enda*'s use as a prospective (semi-)auxiliary but it attested with desiderative *enda* as well. Therefore, we do not assume that prospective *enda* shows specific aspectual restrictions which are not attested with desiderative *enda* as well. However, the two differ in a number of respects as we have shown in this section.

Although the two uses of *enda* clearly differ with respect to their grammatical behavior, we retained the glossing 'want' for both uses. The reasons are simply that we assume, first, *enda* has not developed into a full-fledged auxiliary yet and, second, it is not obvious what the (lexical) meaning of prospective *enda* is. However, we will adopt a different strategy for the verb *rika* to which we turn in the next section.

5 Phasal verbs

Longacre defines 'phasals' as an expression which "indicates whether an action is beginning, continuing, or ending" (Longacre, 1976, 238). Phasal verbs are a specific subtype of phasals and lexically encode whether an eventuality is conceived of as, for instance, beginning. In the literature on phasals – sometimes also referred to as 'phasal aspect' – three subtypes are commonly distinguished: 'inceptive', 'durative' and 'terminative', referring to the beginning, continuation and ending of an eventuality (e.g., Longacre, 1976, 238; Noonan, 2007, 139–140; Croft, 2022, 559–560).

In this section, we focus on the Gĩkũyũ verb *rika* which – as a lexical full verb – means 'to get into' (28a).[13] However, the verb is also used for the expression of inceptive aspect focusing on the beginning of an eventuality (28b). As we can identify two clear semantic contributions in the two different uses, we provide to different glosses for the verb 'get.into' and 'begin', respectively.

(28) a. *Nĩ=ma-a-rik-ir-e mũ-taro.*
FOC=2PL-RMPST-get.into-PFV-FV 3-trench
'They got into the trench.'
b. *Ma-a-rik-ir-e kũ-in-a orĩo.*
2PL-RMPST-begin-FV 15-sing-FV immediately
'They began to sing immediately.'

13 *Rika* is not the only phrasal verb in Gĩkũyũ; other verbs in this group include, *anjia/ambia* 'start/begin', and *rĩkia* 'complete/finish', among others.

Benson (1964, 386) argues in favor of a polysemy analysis of *rika* and attributes the two distinct meanings to the same verbal stem. We depart from this analysis and propose that inceptive *rika* is a semi-auxiliary. Thus, it is still on its way of becoming a grammaticalized auxiliary. However, inceptive *rika* differs from the lexical full verb in a number of grammatical respects very similar to the differences observed between the two uses of *enda* in the preceding section.

Grammatical differences between the two uses of *rika* are again found with respect to the status of the construction's obligatory nominal element. As a lexical full verb, *rika*'s non-subject argument bears direct object properties. It can be pronominalized by a bound argument marker (29a) and can also become subject under passivization (29b). Different from *enda*, *rika* does not take sentential complements.

(29) a. *Nĩ=ma-a-ũ-rik-ir-e.*
FOC=3PL-RMPST-3-get.into-PFV-FV
'They got into it (e.g., the trench).'
b. *Nĩ=ũ-a-rik-ir-wo.*
FOC=3-RMPST-get.into-PFV-PASS
'It was dipped into/ It was got into.'

In the phasal verb construction, the infinitive – *kũina* 'to sing' in (28a) – does not show any object properties. It can neither be realized by a bound argument marker nor can it become the subject of a passive sentence. However, the infinitive can take an object argument (30a). This object can also be realized by a bound object marker (30b) and it becomes subject under passivization (30c). Although passive is marked on the infinitive, the new subject is expressed by a bound subject marker on *rika*.

(30) a. *Nĩ=ma-a-rik-ir-e kũ-in-a rw-ĩmbo.*
FOC=3PL-RMPST-begin-PFV-FV 15-sing-FV 11-song
'They began to sing a song.'
b. *Nĩ=ma-a-rik-ir-e kũ-rũ-in-a.*
FOC=3PL-RMPST-begin-PFV-FV 15-11-sing-FV
'They began to sing it.'
c. *Nĩ=rũ-ra-rik-ir-e kũ-in-wo nĩ mũ-tumia.*
FOC=11-RMPST-begin-PFV-FV 15-sing-PASS by 1-woman
'It was begun to be sung by the woman.'

There is an interesting contrast between phasal verb constructions and prospective aspect constructions when it comes to passivization. Whereas the phasal verb construction realizes the passive morpheme at the infinitive, it is realized on the finite verb in the case of *enda* (31). However, as already shown in the preceding section, the prospective construction is incompatible with passive marking. The addition of the

passive marker results in the desiderative interpretation of *enda* (31b). Interestingly, the desiderative interpretation also obtains if the passive marker is realized on the infinitive (31c) or together on the infinitive and the finite verb (31d).

(31) a. *Mw-aki ũ-r-end-a kũ-hĩhi-a rũkũ.*
3-fire 3-PRS-want-FV 15-burn-FV 11-firewood
'The fire is about to burn the firewood.'
b. *Mw-aki ũ-r-end-wo ũ-hĩhi-e rũkũ.*
3-fire 3-PRS-want-PASS 3-burn-FV 11-firewood
'The fire is expected to burn the firewood.'
c. *Rũkũ rũ-r-end-a rũ-hĩhi-o nĩ mw-aki.*
11.firewood 11-PRS-want-FV 11-burn-PASS by 3-fire
'The firewood is expected to be burned by the fire.'
d. *Rũkũ rũ-r-end-wo rũ-hĩhi-o nĩ mw-aki.*
11.firewood 11-PRS-want-PASS 11-burn-PASS by 3-fire
'The firewood is expected to be burned by the fire.'

The crucial difference between the two constructions is that whereas the prospective *enda*-construction does not allow passivization at all, the phasal *rika*-construction only allows passivization of the embedded infinitive. Thus, the phasal verb itself cannot be passivized. Similar restrictions obtain with respect to causativization and applicativization. In its full verb use, *rika* can, for instance, causativized as shown in (32a). The phasal verb construction, on the other hand, cannot causativized itself. Like passivization, causativization is restricted to the embedded infinitive (32b). The causer is realized by a bound subject marker on *rika* showing that *rika kũinithia* 'start to conduct' has a joint argument structure.

(32) a. *Ma-ma-rik-i-ir-i-e mũ-taro.*
2-2-get.into-CAUS-PFV-DC-FV 3-trench
'They made others to get into the trench.'
b. *Ma-a-rik-ir-e kũ-in-ithi-a rw-imbo.*
2-RMPST-begin-PFV-FV 15-sing-CAUS-FV 11-song
'They began conducting the song.'

The discussion has shown that *rika* and *enda* exhibit very similar restrictions in their semi-auxiliary usage, which we will discuss in detail in the next section.

6 Comparison of the three constructions

The preceding sections focused on a comparison of the three verbs *hũra*, *enda* and *rika* with their respective lexical full verb uses. In this section, we compare the three verbs with respect to their non-full verb uses. We already introduced the label 'semi-auxiliary' for *enda*'s and *rika*'s uses within the prospective aspect, resp. phasal verb construction, whereas we reserved the term 'light verb' for *hũra*'s use in constructions like *hũra thimũ* 'to call'. This distinction between the three verbs is justified on the basis of their shared grammatical properties. The relevant properties can be separated into: (i) properties pertaining to the complex predicate's nominal element and (ii) properties pertaining to the complex predicate's verbal element.

Starting with the nominal elements, the discussion revealed three relevant grammatical properties. First, can the nominal element be pronominalized by a bound argument marker? Second, can the nominal element become subject under passivization? And, third, is the nominal element restricted to a particular nominal class? The results are summarized in Table 1.

Tab. 1: Distribution of the noun-related properties among the three complex predicates.

	***hũra*-LVCs**	***enda*-prospectives**	***rika*-phasals**
(i) pronominalization	yes*	no	no
(ii) passivization	yes*	no	no
(iii) restriction to a particular NC	no	yes	yes

Light *hũra* differs from the two semi-auxiliaries with respect to all three properties. The two first features are related to presence versus absence of direct object properties. As we have shown, the nominal elements of the two semi-auxiliary constructions lack any direct object properties. *Hũra*-LVCs, on the other hand, show a non-uniform behavior. Some LVCs do not allow their nominal element to become subject under passivization (*hũra thimũ*), others do (*hũra baĩni*). Similarly, some LVCs allow their nominal element to be pronominalized by a bound object marker (*hũra thimũ*), others do not (*hũra baĩni*). The interesting question is why the two LVCs show a complementary distribution of these two properties. At the current stage of research, we cannot answer this question but require a larger data set of Gĩkũyũ LVCs to derive sound generalization. However, it is obvious that the nominal elements of the *hũra*-LVC show at least some direct object properties. For this reason, we concluded that the nominal elements can be pronominalized and can become

the subject under passivization but – which we indicate by the asterisk in Table 1 – not necessarily in every single instance.

The three constructions diverge on one further relevant property, namely the restriction of the nominal element to noun class 15. Noun class 15 contains, among some other nouns, infinitives. Infinitives are deverbal nouns which can bear verbal morphology, e.g., a bound object marker as well as passive marking. In difference to other nouns, infinitives do not participate in the nominal number system. Thus, there is no number opposition in case of these nouns. The LVCs under discussion combine with nouns of various noun classes which are still capable of number marking. In case of *thimũ* 'phone', plural is not evident on the noun but is only visible in the agreement system. The quantity expression *-ingĩ* takes the agreement marker for noun class 10 *ny-*, the plural for class 9. The one for noun class 9 – the singular class of *thimũ* – would have been *ĩ-*. The LVCs interpretation is that the woman is not making just one call but many calls (33).

(33) *Mũ-tumia nĩ=a-ra-hũr-a thimũ ny-ingĩ.*
1-woman FOC-1-PRS-beat-FV 10.phone 10-much
'The woman is making many calls.'

There is an additional difference between the nouns occurring in the LVCs and those occurring in the semi-auxiliary constructions. Since the semi-auxiliary constructions take an infinitive, the nominal element is of deverbal origin and therefore denotes an eventuality. The nouns occurring within the light verb constructions are not derived and do not denote any eventuality. In contrast, *thimũ* 'phone' refers to an artifact. The occurrence of artifact nouns within light verb constructions is attested in other languages (e.g., German and Persian) as well (cf. Fleischhauer, 2023b for German and Fleischhauer, 2021 for Persian; but also see the references in the introduction to this volume). Artifacts are objects that have a specific function. This function is a meaning component of the corresponding nouns and allows for the inference of an event in which the artifact, typically as an instrument, is embedded (e.g., Nichols, 2008; Grimm and Levin, 2017; Levin et al., 2019). In the case of *thimũ* 'phone', the inferred event is a calling event, which is the (at least original) primary function of phones.[14] However, due to reasons of space we cannot go into this issue more deeply but we leave the details open for future work.

Turning now to the grammatical properties of the verbs, we identified four relevant properties. First, does the verb license a bound non-subject marker?, Second, does the verb license passive marking? And, third, does the verb license other voice

14 The inferred event in which the artifact is involved as an instrument corresponds to the event associated with Pustejovksky's (1991; 1995) Telic Role.

morphology, i.e., causative and applicative marking? If adding one of the mentioned affixes results in a meaning shift – e.g., from prospective to desiderative – , the verb does not bear the respective property. The results for the verb-related properties are summarized in Table 2:

Tab. 2: Distribution of the verb-related properties among the three complex predicates.

	***hũra*-LVCs**	***enda*-prospectives**	***rika*-phasals**
(i) bound non-subject marker	yes	no	no
(ii) passive morphology	yes	no	no
(iii) other voice morphology	yes	—	no

The light verb *hũra* differs in its grammatical properties from the two semi-auxiliaries. Whereas the semi-auxiliaries are syntactically intransitive, *hũra* is not. This is evidenced by the fact that *hũra* generally licenses bound non-subject markers as well as passive morphology but the semi-auxiliaries do not. In addition, the semi-auxiliaries do not license any argument increasing operations (e.g., causativization, applicativization), whereas *hũra* does. As we observe the same restriction for desiderative *enda*, we cannot attribute this restriction to prospective *enda*'s status as a semi-auxiliary which we indicate by '—' in the respective cell in Table 2.

The mentioned differences might result from the different functions of the constructions. Both the prospective construction as well as the phasal verb construction operate on a (nominalized) verbal concept and can be subsumed under the umbrella term 'periphrastic aspect construction'.[15] Light verb constructions, on the other hand, often fill lexical gaps or extend the lexicon (e.g., Wichmann and Wohlgemuth, 2008, Olejarnik, 2011, 141). Thus, they serve a lexical rather than grammatical function. For instance, Gĩkũyũ does not have a simplex verb meaning 'to call'. The lack of this predicate is compensated by the use of the LVC *hũra thimũ*. The semi-auxiliaries do not function as lexical (complex) predicates but as periphrastic aspectual constructions.

There is a debate whether light verbs are the outcome of a desemantization process or not (e.g., Hopper and Traugott, 1993, 112, Allerton, 2003, 7, Butt and Lahiri, 2013). Sometimes it is even proposed that light verb represent an intermediarey step in the auxiliarization process. Since our analysis is based on limited data, we cannot make a definitive contribution to this debate. However, it seems that we can draw two tentative conclusions from the data. First, both the nominal element as well as the

15 We are using the term 'aspect' here in a very broad sense covering both different types of grammatical aspect (cf. section 4) as well 'phasal aspect'.

verbal element within a semi-auxiliary construction lose some of their independent grammatical properties, i.e., both the verb and the noun show restrictions compared to the corresponding elements in a regular predicate-argument construction. Second, LVCs do not show any general grammatical restrictions compared to corresponding predicate-argument constructions. Whereas semi-auxiliary constructions can be distinguished from corresponding predicate-argument constructions on the basis of their grammatical properties, LVCs differ from corresponding predicate-argument constructions just semantically. The main criterion to distinguish LVCs from regular predicate-argument constructions is: which element –– the noun in the case of the LVC or the finite verb in the case of a regular predicate-argument construction – provides the main predicational content. Thus, we propose that the distinction between LVCs, on the one hand, and regular predicate-argument constructions, on the other hand, is essentially in terms of semantics rather than grammar (a similar claim is presented in Fleischhauer, 2021 for Persian LVCs). Obviously, this is different for the distinction between regular predicate-argument constructions and LVCs, on the one hand, and semi-auxiliary construction on the other. There are indeed grammatical differences allowing to distinguish semi-auxiliary constructions from the mentioned other two predicative construction types.

7 Gĩkũyũ LVCs from a cross-linguistic perspective

In this section, we compare Gĩkũyũ LVCs headed by *hũra* with LVCs in other languages. We compare it to Swahili LVC headed by *piga* 'beat' in section 7.1 and turn to a comparison with LVCs in the Indo-European language Persian in section 7.2.

7.1 Comparison with Swahili LVCs

Similarly to Gĩkũyũ, Swahili possesses LVCs headed by a light verb meaning 'beat, hit' which is *piga*. Olejarnik (2011, 146) states that Swahili LVCs license voice marking. However, concerning the nominal element within such constructions, Krifka (1995, 1401) states that it cannot be a direct object since it can neither be pronominalized nor can it become subject under passivization.[16] Martin (2019, 29) mentions that the

16 Krifka (1995, 1401) proposes that the nominal element of an LVC is (pseudo) incorporated into the verb. Similar claims have been put forward for Persian LVCs (e.g., Saeddi, 2016) but see Fleischhauer (2021) as well as Fleischhauer and Neisani (2020) for arguments against this claim based on Persian language data. These arguments, however, similarly apply to the Bantu languages as well.

nominal element of the LVC cannot be taken up by a bound argument marker. This is shown for the LVC *piga faini* 'to fine' (lit. beat fine) in (34a). However, the noun *raia* 'citizen' – denoting the individual being fined – can be pronominalized by a bound object marker (34b). Crucially, there is no applicative marker on the verb which shows that *raia* is licensed by the complex predicate but not by voice morphology.

(34) a. **Mahakama hiyo i-li-i-pig-a faini raia.*
9.court 9.DEM 9-PST-9-beat-FV 9.fine 1.citizen
'That court fined the citizen.'
b. *Mahakama hiyo i-li-m-pig-a faini raia.*
9.court 9.DEM 9-PST-1-beat-FV 9.fine 1.citizen
'That court fined the citizen.'

(Martin, 2019, 29)

Swahili differs in one relevant syntactic property from Gĩkũyũ. Whereas Gĩkũyũ does not allow the co-occurrence of a bound non-subject marker and coreferential RP within the same sentence, Swahili does not show a similar restriction, as evidenced in (34b). Crucially, *faini* has not to be right-dislocated but can co-occur with a coreferential bound object marker within the same clause. If Martin is right, Swahili differs from Gĩkũyũ concerning the possibility to pronominalize the nominal element of an LVC.

The two languages show similarities when it comes to passivization As the examples in (35) show, the nominal element of the LVC *piga mayowe* 'to shout/scream' can become the subject under passivization.

(35) a. *Mama huyo a-li-pig-a mayowe ku-omba msaada.*
1.mother 1.DEM 1-PST-beat-FV 5.shout 15-beg 3.help
'That mother shouted begging for help.'
b. *Mayowe ya-li-pig-wa kwa sauti kubwa.*
5.shout 5-PST-beat-PASS with 9.voice large
'Shouts were given loudly.'

(Martin, 2019, 30)

Like in Gĩkũyũ, some Swahili LVCs simply reject passivization of their nominal element. This is shown for the LVC *piga mnada* 'to auction' (lit. beat auction) in (36). *Mifugo* 'livestock' which is an argument of the LVC can become subject under passivization (36b), the nominal element *mnada* 'auction' cannot (36c). (For additional data, see, Olejarnik, 2011, 158–159.)

(36) a. *Wakusanya kodi wa-li-pig-a mnada mifugo.*
2.collector 9.tax 2-PST-beat-FV 3.auction 4.livestock
'The tax collectors auctioned off the livestock.'

b. *Mifugo i-li-pig-wa mnada ma wakusanya kodi.*
4.livestock 4-PST-beat-PASS 3.auction by 2.collector 9.tax
'The livestock were auctioned off by the tax collectors.'

c. **Mnada u-li-pig-wa mifungo na wakusanya kodi.*
3.auction 3-PST-beat-PASS 4.livestock by 2.collector 9.tax
'The livestock were auctioned off by the tax collectors.'

(Martin, 2019, 30)

That *mayove* can become subject under passivization in (35b) but not *mnada* in (36c), follows from a general syntactic property of Swahili. The language is, in difference to, for instance, Gĩkũyũ, an asymmetric object language (e.g., Fleischhauer, 2023a, 167). Whereas Gĩkũyũ can make any non-subject argument of a ditransitive predicate the direct object (37), Swahili – and similar asymmetric object languages – restrict direct object properties to the beneficiary argument (38), see, for instance, Krifka (1995, 1400).

(37) a. *A-a-he-ir-e mw-anake i-buku.*
1-RMPST-give-PFV-FV 1-boy 5-book
'S/he gave the boy a book.'

b. *A-a-rĩ-he-ir-e mw-anake.*
1-RMPST-5-give-PFV-FV 1-boy
'S/he gave it to the boy.'

c. *A-a-mũ-he-ir-e i-buku.*
1-RMPST-1-give-PFV-FV 5-book
'S/he gave him/her a book.'

(38) a. *M-sichana a-li-m-pat-i-a ki-jana ki-tabu.*
1-girl 1-PST-1-give-CAUS-FV 1-boy 7-book
'The girl gave the boy a book.'

b. **M-sichana a-li-ki-pat-i-a ki-jana.*
1-girl 1-PST-7-give-CAUS-FV 1-boy
'The girl gave it to the boy.'

Mayove in (35b) can become the subject under passivization since it is the only non-subject argument of *piga*. This is different for the example in (36) as *mnada* 'auction' and *mifugo* 'livestock' are both non-subject arguments. In this case, it is *mifugo* which is realized as the direct object and not *mnada*. As shown in the discussion of the Gĩkũyũ LVC *hũra baini* 'to fine', the two non-subject arguments can become the subject under passivization in accordance with the language's status as a symmetric object language. Thus, the asymmetry in (38) does not represent a specific syntactic property of Swahili LVCs but can be explained on the basis of a general syntactic property of the language. In fact, taking this as a property of Swahili complex predicates –

rather than as a general syntactic constrained – obscures the real nature of complex predicates in that language. We therefore reject the descriptive as well as theoretical analysis proposed by Martin (2019).

The restriction on passivization has been related to the fact that Swahili is an asymmetric object language. This, however, does not account for the fact that the nominal element cannot co-occur with a bound argument marker. Swahili generally licenses the co-occurrence of a bound argument marker with a co-referential RP but – as shown above – prohibits it in the case of an LVC's nominal element. However, this restriction can be accounted for in a different general syntactic property of the language. Swahili shows differential object marking, it restricts object marking – in terms of bound argument markers co-occurring with a coreferential RP – to a specific subset of salient nouns (Aissen, 2003). Iemmolo and Klumpp (2014, 272) state that bound object marking is restricted to arguments having animate referents (cf. the examples in 39). Krifka (1995, 1399) states that the bound object marker occurs "regularly with animate NPs, especially if they are definite, and sometimes with inanimate definite NPs". Thus, bound object marking is restricted to salient argument expressions (animate or definite). Salient argument expressions are such which rank high on the animacy or definiteness scale meaning, i.e., having, for instance, an animate or a definite referent (e.g., Aissen, 2003, 437). The nominal element occurring within LVCs usually rank low both on these scales – they have inanimate and indefinite referents – and therefore do not license bound argument marking in Swahili. In Gĩkũyũ, bound object marking is not subject to differential object marking which means that, for instance, animate as well as inanimate object arguments can be pronominalized by a bound object marker. And in fact, we observed bound object marking of the LVC's nominal element in the LVC *hũra thimũ*. Thus, we can again relate the difference between Swahili and Gĩkũyũ concerning the pronominalization of the LVC's nominal element to a more general syntactic difference between the two languages.

(39) a. *Juma a-li-m-pig-a risasi tembo jana usiku.*
Juma 1-PST-1-beat-FV 9.bullet 1.elephant yesterday night
'Juma shot an/the elephant last night.'

b. *Risasi i-li-pig-a m-ti karibu na sisi.*
9.bullet 9-PST-beat-FV 3-tree near with us
'A bullet struck the tree near us.'
(Vitale, 1981, 123–124; glossing slightly adapted)

Thus, there is an independent grammatical reason for why the nominal element of an LVC cannot be pronominalized by a bound argument marker in Swahili: bound object markers are subject to differential object marking and are restricted to object

argument's which have animate/ definite referents. Concerning Swahili, we reject Krifka's strict statement that the LVC's nominal element cannot be direct object as we can explain the grammatical behavior of the LVC's nominal element to more general syntactic properties related to direct object arguments. In addition, some LVCs even allow their nominal element to become the subject under passivization, as shown above. This clearly falsifies Krifka's strict statement. We have no reason to propose that Swahili LVCs differ concerning their grammatical properties from their Gĩkũyũ correspondents. Instead, we have shown that the language-specific differences between Gĩkũyũ and Swahili LVCs follow from more general syntactic properties with respect to which the two languages differ and which are not specifically restricted to complex predicates.

7.2 Direct object properties: A look into Persian

The preceding sections have shown that the nominal element of LVCs show a non-uniform behavior with respect to direct object diagnostics. For Swahili, we argued that this is a consequence of the language's syntactic architecture. We have not come up with a similar explanation for Gĩkũyũ but regard it as idiosyncratic properties of individual LVCs. Persian is an Indo-European language which makes massive use of light verb constructions. Like the two Bantu languages discussed abobe, Persian shows similar non-uniformity when it comes to direct object properties of the LVCs' nominal element.

Within its history, Persian lost most of its monolexical verbs and compensated this loss by use of light verb constructions (e.g., Mohammad and Karimi, 1992, 195, Samvelian, 2018, 256). Light verb constructions in Persian come in different morphosyntactic types but we only focus on the most frequent type consisting of a noun-verb sequence (cf. Dabir-Moghaddam, 1997, 31 for an overview on other morphosyntactic construction types). Unlike the Bantu languages, Persian is a verb-final language and the nominal element of an LVC precedes its light verb. An illustrative example consisting of the noun *šekast* 'defeat' and the light verb *dâdan* 'give' meaning 'to defeat' (lit. defeat give) is shown in (40).

(40) *Ân-hâ došman-râ šekast dâd-and.*
DEM-PL enemy-ACC defeat give.PST-3PL
'They defeated the enemy.'

(Fleischhauer, 2021, 44)

A further difference between Persian, on the one hand, and the Bantu languages, on the other, is that Persian is a dependent-marking language. Rather than bound argument markers, it uses a combination of case marking and adpositional marking

to indicate the grammatical relations of the argument expressions. Persian has an accusative case marker *-râ* which attaches to direct object arguments. Indirect objects, which do not play a role in the current section, are marked by the preposition *be* 'to'. Similar to Swahili, Persian displays differential object marking. Accusative case marking is restricted to salient direct object arguments either having indefinite specific or definite reference (cf. Bossong, 1985; Lazard, 1992; Ghomeshi, 1997).

In case of the LVC *šekast dâdan*, accusative case marking is not found on the LVC's nominal element but on the noun *došman* 'enemy' which is the one being defeated. However, other LVCs show case marking on the LVCs nominal element. This is, for instance, the case for *latme zadan* 'to damage' (lit. damage hit) in (41).

(41) *Tegarg-e diruz în latme-râ be bâq-e man zad.*
hail-EZ yesterday DEM damage-ACC to garden-EZ 1SG hit.PST
'The yesterday's hail caused this damage to my garden.'
(Karimi-Doostan, 2011, 71; glossing slightly adapted)

The examples in (42) show that the nominal element of the LVC *jade kešidan* 'built a road' (lit. road pull) can receive accusative case marking (42a) and can become subject under passivization (42b). The noun exhibits *jâde* both relevant direct object properties. This is different from the considered Bantu languages, where nouns exhibit only a subset of direct object properties.

(42) a. *In âqâya in jade-râ be forudgâh kešid-and.*
DEM men DEM road-ACC to airport pull.PST-3PL
'These men build the road to the airport.'
b. *Jâde-hâ-i be forudgâh kešide šod-and.*
road-PL-INDEF to airport pull.PTCP become.PST-3PL
'(Some) Roads to the airport were build.'
(Fleischhauer, 2021, 48)

As the discussion reveals, a non-uniform distribution of direct object properties among the LVCs' nominal elements is attested in genetically unrelated languages. Concerning Persian LVCs, there exists no consensus in the literature whether there is any systematicity underlying the distribution of direct object properties.[17] It might be the case that it is an LVCs idiosyncratic property whether its nominal element bears direct object properties. But it could also be the case that whether an LVCs nominal element bears – some or all – direct object properties or not reflects a difference in the degree of the LVC's lexicalization. The nominal elements in stronger lexicalized

17 For a more detailed discussion of the grammatical properties of Persian light verb constructions, see Fleischhauer and Neisani (2020) and Fleischhauer (2020, 2021).

LVCs lacking – some or all – direct object properties, whereas does in – not yet – lexicalized LVCs retain them. We cannot provide even a speculative answer on this question but conceive of it as an interesting probably cross-linguistic project for future work.

8 Conclusion

The paper started with the observation that Gĩkũyũ possesses at least three different predicative constructions consisting of a semantically reduced verb and a nominal element. One of these constructions – represented by the sequence *hũra thimũ* 'to call' – resembles what in other languages, e.g., Swahili, is called a light verb construction. Departing from this observation we asked whether light verb constructions of the type exemplified by *hũra thimũ* can be distinguished on the basis of grammatical properties (i) from regular predicate-argument constructions headed, for instance, by *hũra* as a lexical full verb meaning 'to beat' and (ii) from semi-auxiliary constructions such as prospective '*enda* + INF' and inceptive '*rika* + INF'.

The analysis revealed that *hũra*-LVCs do not show any systematic grammatical differences compared to heavy uses of the same verb. However, we observed that pronominalization as well as passivization of the LVC's nominal element is subject to – as it seems – idiosyncratic restrictions. We also observed the same variance in other languages, indicating that LVCs are generally compatible with its nominal element bearing a grammatical relation – notably direct object – to the light verb.

LVCs differ from semi-auxiliary constructions with respect to their grammatical properties. The analyses in sections 4 and 5 revealed that semi-auxiliary constructions show more grammatical restrictions than LVCs. The restrictions concern voice marking, especially passivization. The severe restrictions on passivization – in line with the fact that the nominal element cannot be pronominalized by a bound argument marker – indicate that the nominal element entirely lost its direct object properties. We consider this as a reflex of the auxiliarization process which the respective verbs undergo. Such grammatical restrictions can be analyzed as 'decategorization', i.e., the loss of morphosyntactic properties of a lexical unit. Decategorization represents a central parameter in grammaticalization theory (e.g., Narrog and Heine, 2021, 72–78), and the results suggest that semi-auxiliaries show signs of grammaticalization – but are not yet fully grammaticalized – while the light verb does not exhibit this.

As the overall result, the case study shows that *hũra*-LVCs are grammatically more similar to regular predicate-argument constructions than to semi-auxiliary constructions. In fact, there are not any relevant grammatical differences between LVCs and regular predicate-argument constructions. The same is true of the related Bantu

language Swahili, as we attributed any restrictions on passivization and pronominalization to the existence of more general syntactic properties (i.e., asymmetry in the realization of the two non-subject arguments and differential object marking).

The results on the comparison of *hũra*-LVCs to corresponding full verb uses of *hũra* as well as to the semi-auxiliary constructions discussed in the paper should, in a next step, validated on a data set contain a larger number of *hũra*-LVCs. In addition, the criteria should be tested against a different data set containing other types of light verb constructions as well as semi-auxiliary constructions within the same but also other language.

Acknowledgments

The work on this study was partially funded by a grant from the German Research Foundation (DFG) and is part of the DFG-funded project 'Funktionsverbgefüge: Familien & Komposition' (HE 8721/1-3).

Bibliography

Aissen, Judith. 2003. Differential Object Marking: Iconicity vs. Economy. *Natural Language & Linguistic Theory* 21. 435–483. https://doi.org/10.1023/A:1024109008573.

Allerton, David. 2003. *Streched verb constructions in English*. London: Routledge.

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

Barlow, A. Ruffell. 1960. *Studies in Kikuyu grammar and idiom*. Edinburgh: William Blackwood.

Benson, T. Godfrey. 1964. *Kikuyu-English dictionary*. London: Clarendon Press.

Bergvall, Victoria L. 1987. *Focus in Kikuyu and Universal Grammar*. Doctoral dissertation: Harvard University.

Bossong, Georg. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.

Bresnan, Joan and Lioba Moshi. 1990. Object Asymmetries in Comparative Bantu Syntax. *Linguistic Inquiry* 21(2). 147–185.

Butt, Miriam and Aditi Lahiri. 2013. Diachronic pertinacity of light verbs. *Lingua* 135. 7–29. https://doi.org/10.1016/j.lingua.2012.11.006.

Cable, Seth. 2013. Beyond the past, present, and future: towards the semantics of 'graded tense' in Gĩkũyũ. *Natural Language Semantics* 21. 219–276. https://doi.org/10.1007/s11050-012-9092-3.

Clements, George N. 1984. Binding domains in Kikuyu. *Studies in the Linguistic Sciences* 14. 37–56.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Croft, William. 2022. *Morphosyntax*. Cambridge: Cambridge University Press.

Dabir-Moghaddam, Mohammed. 1997. Compound verbs in Persian. *Studies in the Linguistic Sciences* 27(2). 25–59.

Dik, Simon. 1997. *The Theory of Functional Grammar. Part 1: The structure of the clause*. Berlin: De Gruyter Mouton.

Fleischhauer, Jens. 2020. Predicative multi-word expressions in Persian. In *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, 552–561. Hanoi: Association for Computational Linguistics. https://aclanthology.org/2020.paclic-1.63.

Fleischhauer, Jens. 2021. Simplex and Complex Predicates in Persian: An RRG Analysis. In Robert Van Valin (ed.), *Challenges at the Syntax-Semantics-Pragmatics Interface*, 31–62. Newcastle upon Tyne: Cambridge Scholars Publishing.

Fleischhauer, Jens. 2023a. Argument doubling and right-dislocation – An RRG analysis of head-marking in isiZulu. In Jens Fleischhauer and Claudius Patrick Kihara (eds.), *African languages from a Role and Reference Grammar perspective*, 163–191. Berlin: De Gruyter. https://doi.org/10.1515/9783110795295-007.

Fleischhauer, Jens. 2023b. German light verb constructions with non-eventive nouns. In Jean-Pierre Colson (ed.), *Phraseology, constructions and translation. Corpus-based, computational and cultural aspects*, 387–398. Louvain: Presses Universitaires de Louvain.

Fleischhauer, Jens. 2023c. Prospective aspect and current relevance – A case study of the German prospective '*stehen vor* NP'-construction. *Journal of Germanic Linguistics* 35(4). 371–408. https://doi.org/10.1017/S1470542722000198.

Fleischhauer, Jens and Thomas Gamerschlag. 2019. Deriving the meaning of light verb constructions – a frame account of German *stehen* 'stand'. In Constanze Juchem-Grundmann, Michael Pleyer and Monika Pleyer (eds.), *Yearbook of the German Cognitive Linguistics Association, vol. 7*, 137–156. Berlin: De Gruyter Mouton. https://doi.org/10.1515/gcla-2019-0009.

Fleischhauer, Jens and Mozhgan Neisani. 2020. Adverbial and attributive modification of Persian separable light verb constructions. *Journal of Linguistics* 56(1). 45–85. https://doi.org/10.1017/S0022226718000646.

Ghomeshi, Jila. 1997. Topics in Persian VPs. *Lingua* 102(2–3). 133–167. https://doi.org/10.1016/S0024-3841(97)00005-3.

Gibson, Hannah and Lutz Marten. 2016. Variation and Grammaticalisation in Bantu Complex Verbal Constructions: The Dynamics of Information Growth in Swahili, Rangi and SiSwati. In Léa nash and Pollet Samvelian (eds.), *Approaches to Complex Predicates*, 70–109. Leiden: Brill.

Good, Jeff. 2005. Reconstructing morpheme order in Bantu. The case of causativization and applicativization. *Diachronica* 22(1). 3–57. https://doi.org/10.1075/dia.22.1.02goo?locatt=mode:legacy.

Good, Jeff. 2007. Slouching towards deponency: A family of mismatches in the Bantu verb stem. In Matthew Baerman, Greville G. Corbett, Dunstan Brown and Andrew Hippisley (eds.), *Deponency and morphological mismatches*, 203–230. Oxford: Oxford University Press.

Grimm, Scott and Beth Levin. 2017. Artifact Nouns: Reference and Countability. In Andrew Lamont and Katerina Tetzloff (eds.), *Proceedings of the 47th Annual Meeting of the North East Linguistic Society (NELS 47)*, 45–64. Amherst, MA: University of Massachusetts.

Guthrie, Malcolm. 1967. *The classification of Bantu languages*. London: Dawson.

Heine, Bernd. 1994. On the genesis of aspect in African languages: The proximative. *Proceedings of the 20th Annual Meeting of the Berkeley Linguistics Society* 35–46.

Hopper, Paul and Elizabeth Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165525.

Hyman, Larry and Alessandro Duranti. 1982. On the object relation in Bantu. In Paul Hopper and Sandra Thompson (eds.), *Studies in Transitivity*, 217–239. Leiden: Brill.

Iemmolo, Giorgio and Gerson Klumpp. 2014. Introduction to the special issue 'Differential Object Marking: theoretical and empirical issues'. *Linguistics* 52(2). 271–279. https://doi.org/10.1515/ling-2013-0062.

Karimi-Doostan, Gholamhossein. 2011. Separability of light verb constructions in Persian. *Studia Linguistica* 65(1). 70–95. https://doi.org/10.1111/j.1467-9582.2011.01178.x.

Kihara, Claudius Patrick. 2016. *Aspects of Gĩkũyũ (Kikuyu) Complex Sentences: A Role and Reference Grammar Analysis*. Doctoral dissertation: Heinrich-Heine University Düsseldorf.

Krifka, Manfred. 1995. Swahili. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann (eds.), *Syntax: An international handbook of contemporary research*, 1397–1418. Berlin: De Gruyter.

Kuteva, Tania. 2001. *Auxiliation: An enquiry into the nature of grammaticalization*. Oxford: Oxford University Press.

Kuteva, Tania, Bas Aarts, Gergana Popova and Anvita Abbi. 2019. The grammar of 'non-realization'. *Studies in Language* 43(4). 850–895. https://doi.org/10.1075/sl.18044.kut.

Lambrecht, Knud. 2001. Dislocation. In Martin Haspelmath, Ekkehard König, Wolfgang Raible and Wulf Oesterreicher (eds.), *Language typology and language universals: An international handbook. Vol. 2*, 1050–1078. Berlin: De Gruyter.

Lazard, Gilbert. 1992. *A Grammar of Contemporary Persian*. Costa Mesa, CA: Mazda Publishers.

Levin, Beth, Lelia Glass and Dan Jurafsky. 2019. Systematicity in the semantics of noun compounds: The role of artifacts vs. natural kinds. *Linguistics* 57 (3). 429–471.

Longacre, R. E. 1976. *An Anatomy of Speech Notions*. Lisse: The Peter de Ridder Press.

Maho, Jouni F. 2003. A classification of the Bantu languages: an update on Guthrie's referential system. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*, 639–651. London: Routledge.

Martin, Joshua. 2019. Analyzing Complex Predicates in Swahili. In *Proceedings of the Florida Linguistics Yearly Meeting (FLYM) 5*, vol. 6, 25–36.

Mohammad, Jan and Simin Karimi. 1992. Light verbs are taking over: Complex verbs in Persian. In J. A. Nevis and V. Samiian (eds.), *Proceedings of the Western Conference on Linguistics. Vol. 5*, 195–212. Fresno: California State University.

Mugane, John. 2003. Hybrid Constructions in Gĩkũyũ: Agentive Nominalizations and Infinitive-gerund constructions. In Miriam Butt and Tracy Holloway King (eds.), *Nominals: Inside and Out*, 235–266. Stanford: CSLI Publications.

Narrog, Heiko and Bernd Heine. 2021. *Grammaticalization*. Oxford: Oxford University Press.

Nichols, Johanna. 1986. Head-Marking and Dependent-Marking Grammar. *Language* 62(1). 59–119. https://doi.org/10.2307/415601.

Nichols, Lynn. 2008. Lexical semantic constraints on noun roots and noun borrowability. *Studies in Language* 32 (3). 683–700.

Nicolle, Steve. 2023. *A Grammar of Digo*. Dallas: SLI Institute.

Noonan, Michael. 2007. Complementation. In Timothy Shopen (ed.), *Language Typology and Syntactic Description: Vol. 2, Complex Constructions*, 52–150. Cambridge: Cambridge University Press.

Olejarnik, Magdalena. 2009. *Complex Predicates in Swahili: An LFG Approach*. Doctoral dissertation: Universytet Warszawski.

Olejarnik, Magdalena. 2011. An LFG Approach to Complex Predicates in Swahili. In Piotr Banski, Beata Lukaszewicz, Monika Opalinska and Joanna Zaleska (eds.), *Generative Investigations: Syntax, Morphology, and Phonology*, 139–165. Newcastle upon Tyne: Cambridge Scholars Publishing.

Pustejovski, James. 1991. The Generative Lexicon. *Computational linguistics* 17. 409–441.

Pustejovski, James. 1995. *The Generative Lexicon*. Boston: The MIT Press.

Saeddi, Zari. 2016. *Complex predicates in modern Persian*. Sheffield: Equinox.

Samvelian, Pollet. 2018. Specific features of Persian syntax. In Anousha Sedighi and Pouneh Shabani-Jadidi (eds.), *The Oxford Handbook of Persian Linguistics*, 226–269. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198736745.013.9.

Schwarz, Florian. 2003. Focus Marking in Kikuyu. *ZAS Papers in Linguistics* 30. 41–118.

Van Valin, Robert. 2008. RPs and the nature of lexical and syntactic categories in Role and Reference Grammar. In Robert Van Valin (ed.), *Investigations of the Syntax-Semantics-Pragmatics Interface*, 161–178. Amsterdam: John Benjamins.

Van Valin, Robert. 2023. Principles of Role and Reference Grammar. In Delia Bentley, Wataru Nakamura Ricardo Mairal Usón and Robert Van Valin (eds.), *The Cambridge Handbook of Role and Reference Grammar*, 17–180. Cambridge: Cambridge University Press.

Vitale, Anthony J. 1981. *Swahili syntax*. Dordrecht: Foris.

Wichmann, Søren and Jan Wohlgemuth. 2008. Loan verbs in a typological perspective. In Thomas Stolz, Dik Bakker and Rosa Salas Palomo (eds.), *Aspects of Language Contact*, 89–122. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110206043.89.

Wittke, Jonas. 2015. Verbs. In Robert Englebretson (ed.), *A Basic Sketch Grammar of Gĩkũyũ*, 52–81. Houston: Rice Working Papers in Linguistics.

Ryan Ka Yau Lai

Beyond bidirectional association: Distinguishing light verb constructions from other conventionalised noun-verb combinations in modern Tibetan

1 Introduction

Light verb constructions (LVCs), consisting of a semantically light verb with the predicational information primarily expressed by a semantically heavy accompanying expression (usually a nominal), have long been studied in corpus linguistics within collocation analysis, which examines co-occurrence of linguistic forms in a corpus. Such analyses measure the strength of attraction between two forms with statistical association measures (AMs; Evert, 2005, 75): A pair of forms with a higher AM is likelier to be a conventionalised combination.

Traditionally, linguists focus on 'bidirectional' or 'symmetric' AMs, which quantify mutual attraction: the degree to which the two components of the construction tend to co-occur. They do not reflect whether one part of the collocation is more attracted to the other than vice versa. For example, in *wishful thinking*, *wishful* attracts *thinking* more than the other way around, but bidirectional AMs cannot represent this information. LVCs are particularly likely to exhibit asymmetric association: For example, in the English LVC *take a look*, the light verb nominal *look* likely attracts the verb *take* far more than vice versa. This may motivate the use of 'unidirectional' or 'asymmetric' measures of attraction (Michelbacher et al., 2007; Gries, 2013, 141-155; Gries, 2022, Sec. 2.1.5.), which have hitherto gained little attention.

In this paper, I explore how unidirectional measures can be added to the traditional toolkit of statistical measures for LVC detection in the context of modern Tibetan, focusing on Modern Literary Tibetan and the Central Tibetan variety spoken by the Dharamsala diaspora community. The paper's focus is not to directly propose a detection method, but to explore how the statistical distributions of light verb constructions distinguish them from other frequent noun-verb combinations, providing a clearer conceptual basis for improving feature engineering and model structure in future LVC detection work. Tibetan LVCs present challenges distinct from English

Ryan Ka Yau Lai, University of Georgia, Athens, USA

https://doi.org/10.1515/9783112229330-005

(Vincze et al., 2011). They are much more ubiquitous, including high-frequency items like the following:

(1) a. སློབ་སྦྱོང་བྱེད་
slob.sbyong byed
study do
'to study'

b. བསམ་བློ་གཏོང་
bsam.blo gtong
thought send
'to think (vol.)'

c. ཆང་སა་རྒྱག་
chang.sa rgyag
marriage strike
'to marry'

LVCs are so important in modern Tibetan that they outnumber simple verbs in spoken Lhasa Tibetan by 2:1 (Randall, 2016, 4). In these constructions, the verb is semantically light, but still encodes some semantic contrasts like volitionality and honorificity (see (4) and (5) for examples), and as Randall (2016, 122–177) shows, the nominal exhibits syntactic properties that suggest it is not as referential as the object of a 'regular' noun-verb combination, such as the inability to be modified by adjectives and demonstratives or replaced by interrogative words.

Compared to the vast resources on monomorphemic verb roots (e.g. Hackett, 2003; Hill, 2010; Hoshi, 2003), LVCs remain understudied: Information is scattered across dictionaries, with scant descriptions of argument structure and other usage information (though Bailey and Walker, 2004 includes argument structure information for a non-specialist audience). Statistical detection of LVCs from bodies of naturally occurring texts is instrumental to creating better datasets for teaching and researching Tibetan, for which the statistical distributional properties of LVCs must be better understood.

Computational work (Zhào et al., 2016) has previously detected Tibetan LVCs from corpora, employing statistical measures including entropy of surrounding tokens, measuring the diversity of contexts in which the combination appears, and pointwise mutual information (PMI), measuring bidirectional association. Yet these measures may also pick up other types of conventionalised noun-verb combinations in Tibetan. Consider the examples in (2). In (2a) the noun is semantically lighter than the verb: ངོ་ *ngo* 'face' only tells us that the target of the action has an appearance, a very general class consistent with many types of situations, whereas ཤེས་ *shes* 'know' tells us that the LVC is tells us that something is within the subject's knowledge. In (2b), the noun and verb convey similar information: both convey that the action is done for the purpose of enjoyment. In (2c), the noun and verb have distinct semantic contributions: 'alcohol' is a concrete object and 'drink' a concrete activity, but one cannot straightforwardly guess one from the other. These alternatives pri-

marily differ from LVCs with respect to the relationship between the noun and the verb: while LVCs have a semantically specific noun carrying most of the semantics plus a semantically general verb encoding relatively little information, meaning is distributed differently in these other types of constructions. Thus, bidirectional association and context diversity are likely unable to fully capture these differences. Instead, 'unidirectional' measures, which examine both the verb's relationship to the noun and vice versa instead of collapsing them into a single measure, may be more appropriate.

(2) a. དོ་ཤེས་
ngo shes
face know
'to recognise'

b. རྩེད་མོ་རྩེ་
rtsed.mo rtse
game play
'play a game'

c. ཆང་འཐུང་
chang 'thung
alcohol drink
'to drink alcohol'

To explore the possibility of using unidirectional measures, I first extract absolutive (i.e. unmarked case) noun-verb combinations from the Nanhai corpus (Schmidt, 2019), code each retrieved combination above a frequency threshold as an LVC or one of the above three categories, and calculate a number of measures of conventionalisation, including measures of association (including traditional bidirectional ones like PMI and G^2 but also unidirectional ones like ΔP and G^2rank), context entropy, and productivity. I conduct exploratory analyses of the results by plotting the measures using various graphs and performing unplanned (exploratory) hypothesis tests, focusing how these measures may potentially distinguish between different types of conventionalised N-V combinations. I also examine how similar the information encoded by different measures is.[1]

In the following, Section 2 reviews background information on LVCs, statistical association measures, and (a)symmetry in collocations. Section 3 describes the problem tackled by this paper. Section 4 describes the methodological approach for extracting potential conventionalised noun-verb combinations, and Section 5 presents the results of the study. Section 6 presents a discussion of the data.

1 Data and code are available at https://zenodo.org/doi/10.5281/zenodo.10895756, last access 12/9/2025.

2 Background

This section reviews linguistic properties of LVCs in Modern Tibetan most relevant to the motivations and methodology of this paper (Section 2.1), statistical measures of LVC strength from in previous studies (Section 2.2), and collocational (a)symmetry and asymmetric measures (Section 2.3).

2.1 Linguistic properties of LVCs in Modern Tibetan

Tibetan syntax is generally verb-final. Most verb roots are monosyllabic with four stem forms – present, past, future, and imperative – often followed by auxiliaries and agglutinative morphology, especially in speech. Tibetan has a pragmatic case system containing, among others, an ergative གིས་=*gis* and an oblique ལ་=*la* (and their allomorphs), mostly non-obligatory (see Tournadre, 1996 for discussion on the ergative in Central Tibetan). 'Object' arguments are generally in the unmarked absolutive. As in most languages, the non-verbal element in an LVC resembles either an object (3a) or an intransitive subject (3b) even if they've lost subject/ object properties (such as the ability to be modified or replaced by an interrogative pronoun) synchronically (Lessan Pezechki and Tournadre, 2014, 9), so the LVC typically consists of a nominal with no overt case followed by an inflected verb,[2] though there are some expressions which also seem to have a semantically light verb, coupled with a semantically heavy case-marked nominal (3c; see also Randall, 2016, 71–91):

(3) a. གོམ་པ་རྒྱག་
gom.pa rgyag
walk strike
'to walk'

b. འབྲུག་སྐད་རྒྱག་
'brug.skad rgyag
thunder strike
'thunder strikes'

c. སིལ་བུར་འགྲོ་
sil.bu=r 'gro
piece=OBL go
'to fall into pieces'

The term 'light verb' is not standard in work on Tibetan and related varieties. LVC-like constructions, consisting of a highly conventionalised combination of a nominal (or adjective) plus a verb acting as a lexical unit, are also known as 'compound

2 Sometimes, there is an additional absolutive argument, so the light verb nominal may be better described as caseless, not absolutive (Simon, 2011, 61).

verbs/ verbal compounds', 'phrasal verbs' (Agha, 1993, Denwood, 1999, etc.), 'complex predicates' (Randall, 2016), 'conjunct verbs' (Matthew & Sumi, 2005 as cited in Randall, 2016) or 'verbalised constructions' (Bartee, 2007), and the verbal component may be known as 'verbalisers' (Tournadre and Dorje, 1998), 'support verbs' (Simon, 2011), or 'secondary verbs' (see Randall, 2016, 8–12 for a terminological overview). Additionally, since disyllabic nominals with a monosyllabic verb is the most common combination, 'trisyllabic verb' is also common (Zhào et al., 2016).[3] These traditional terms cover constructions not typically considered LVCs in general linguistics. The term 'light verb' is also used by some authors (Zemp, 2018, e.g.) for semantically light auxiliary or serial verbs like ཚར་ *tshar* 'finish' or ཡོང་ *yong* 'come' which have distinct syntactic properties: They generally follow a tensed verb root rather than a deverbal nominal, and some exhibit properties like restricted inflectional possibilities and phonological reduction (see DeLancey, 1991 for examples from Central Tibetan).

The three most common Tibetan light verbs are བྱེད་ *byed*, གཏོང་ *gtong*, and རྒྱག་ *rgyag*, typically encoding volitional action verbs, as in (1). Some examples with non-volitional light verbs are as follows:

(4) a. ཁོང་ཁྲོ་ལང་
khong.khro lang
anger arise
'to get angry'

b. མགོ་སྐོར་ཐེབས་
mgo.skor thebs
fraud arrive
'to be cheated'

c. གད་མོ་ཤོར་
gad.mo shor
laughter let.loose
'to let out laughter'

Light verbs also encode honorific/ humilific status, as in the following:

(5) a. དགའ་བསུ་ཞུ་
dga'.bsu zhu
welcome do.HUM
'to welcome (hum.)'

b. དོ་སྤྲོད་གནང་
ngo.sprod gnang
introduction do.HON
'to introduce (hon.)'

c. འདྲ་པར་སྐྱོན་
'dra.par skyon
photograph do.HON
'to photograph (hon.)'

3 The terminological diversity reflects uncertainty as to whether the verb and nominal constitute one 'word'; although terms like 'verbaliser' and 'trisyllabic verb' strongly imply that they are a single word, there are reasons to consider them separate words, e.g. the fact that the noun and verb may be separated by forms like adverbs (discussed below). I will be agnostic on the issue.

There is no straightforward mapping from plain to honorific/ humilific forms. Though *skyon* is generally used for *rgyag* 'strike', *zhu* and *gnang* are far more flexible; for example, while the non-honorific version of (3) would use *byed*, *zhu* can also map to གཤོད་ *shod* 'say'.

The non-verbal element accompanying a light verb is most commonly, but not always, a deverbal nominal (e.g. *'dra.par* in (5c) is not deverbal). Moreover, most deverbal light-verb nominals are composed of two verbal roots, which do not necessarily appear frequently as main verbs in the modern language (e.g. *skad* in (3c) is uncommon as a verb). This is unlike English, where most light-verb nominals are derived from productive verbs, e.g. *look* in *take a look*. Thus, measures based on properties of the verbalised nominal in English corpus linguistics (e.g. using properties of the verb look when investigating *take a look*) cannot be straightforwardly applied to modern Tibetan.

As in English, a single nominal can take on multiple light verbs with different semantic nuances, and where the nominal is deverbal, the original verb(s) is sometimes also usable as a verb alone (6). Here, (6a) and (6b) differ in volitionality, and (6c) perhaps in aspect from (6a).

(6) a. བསམ་བློ་གཏོང་
bsam.blo gtong
thought send
'to think' (= 1b)

b. བསམ་བློ་འཁོར་
bsam.blo 'khor
thought turn
'to think of'

c. བསམ་
bsam
think
'to think'

Light verbs can also accompany adjectives. These constructions are harder to detect, since adjectives can also modify the verb – and in such cases, they may in turn have a light verb nominal before them. Light verbs can also accompany light noun-adjective combinations, e.g. སེམས་པ་བཟང་པོ་བྱེད་ *sems.pa bzang.po byed* 'have a good heart'. I thus leave the complex topic of adjective-light verb combinations to future research. Similarly, I put aside constructions where an honorific light verb accompanies a plain verb.

Syntactically, Tibetan light verbs are separable from their accompanying non-verbal element by verb-modifying forms, including adverbs, adjectives, the indefinite marker *gcig*, and the similative demonstratives *'di.'dra*/ *de.'dra* (Randall, 2016, 47–63). In this paper, I call these forms interveners. Arguments cannot intervene between the light verb and the non-verbal element. Unlike in English, accompanying nominals are usually not modifiable (Randall, 2016, 43–71), so determiner-based light verb

detection methods in English (Stevenson et al., 2004; Tu and Roth, 2011) will also fail for Tibetan.

2.2 Statistical properties of LVCs across languages

In this section, I first explain some notational conventions, then use them to introduce statistical measures for LVC detection proposed in the literature, focusing only on measures applicable to Tibetan.

Notation and definitions. Measures of statistical properties of LVCs, and collocation more generally, are generally based on the following contingency table 1, adapted to the case of light verbs.

Tab. 1: Contingency table for computing statistical measures of word association.

	n	¬*n*	Totals
v	*f*(*n*,*v*)	*f*(¬*n*,*v*)	*f*(*v*)
¬*v*	*f*(*n*,¬*v*)	*f*(¬*n*,¬*v*)	*f*(¬*v*)
Totals	*f*(*n*)	*f*(¬*n*)	*N*

Here, *f* refers to frequency, *n* stands for nominals, *v* stands for verbs, ¬ means 'not', and *N* is the total sample of candidate noun-verb combinations currently considered. For example, if *n* is *look* and *v* is *take*, then *f*(*n*,*v*) is the frequency of *take* + *look*, *f*(¬*n*,*v*) is the frequency of *take* + nouns other than *look*, and *f*(¬*n*,¬*v*) refers to noun-verb combinations with neither *take* nor *look*. The bottom row and rightmost column represent column and row totals respectively: *f*(*n*) refers to the frequency of *look* in noun-verb combinations overall, *f*(¬*v*) refers to the frequency of verbs other than *take*, etc.

In addition to the above frequencies, I denote estimated probabilities of a noun-verb combination belonging to a particular cell in the table, estimated using frequencies from the table, as follows:

$$p(n, v) = f(n, v)/N, p(n) = f(n)/N, etc.$$

Here, *p* stands for probability. (Technically, these are not actual probabilities, but relative frequencies used to estimate probabilities). Thus, in our case of *take a look*, *p*(*n*,*v*) is the estimated probability that the verb is *take* and the noun is *look*, calculated by dividing the frequency of *take* + *look* by the total number of noun-verb combinations considered.

A final piece of notation is 'conditional probability'. It measures the probability that a nominal will be used given the verb, or vice versa:

$$p(n|v) = f(n,v)/f(v), p(\neg v|n) = f(\neg v,n)/f(n), etc.$$

Here, the pipe | means 'given', so $p(n|v)$ in our example means the probability that the noun is *look* given that the verb is *take*. This is calculated by dividing the number of times *take* + *look* appeared by the number of times *take* appeared.

Measures of association strength. $f(n,v)$ itself has been used as a statistical correlate of LVC status (Tan et al., 2006, 51). Since a common assumption in the corpus-linguistic literature is that light verbs are conventionalised combinations, they can be expected to occur more frequently than more incidental noun-verb combinations.

Perhaps the most common measure for detecting LVCs is pointwise mutual information (PMI) (Stevenson et al., 2004; Tan et al., 2006):

$$PMI = \log_2 \left(\frac{p(n,v)}{p(n)p(v)} \right)$$

Thus, PMI considers the how likely the noun and verb are to co-occur, normalised by how often the noun and verb occur individually. Because the verb and nominal components of LVCs are generally strongly associated with each other, previous studies have frequently used high PMI as an indication that something is an LVC. Zhào et al. (2015, 2016) the PMI approach to Tibetan.

Significance measures. Significance measures are based on significant tests that quantify how unlikely the data would be if the two forms were not associated with each other, such as χ^2 and log-likelihood ratio (G^2, Dunning, 1994). This paper will adopt G^2, given its attractive properties over χ^2 especially for skewed samples (Dunning, 1994). G^2 is based on the difference between observed and expected counts. Expected counts are the number that we would, on average, expect to see in a cell of the contingency table if there were no association between the noun and the verb. It is calculated as follows:

$$E(n,v) = Np(n)p(v), E(\neg n,\neg v) = Np(\neg n)p(\neg v), etc.$$

The G^2 statistic quantifies how much the actual table deviates from these expected values. Technically, it is the log-likelihood ratio between the null hypothesis of independence and the alternative hypothesis of dependence between the noun and the verb. It takes the log of the observed divided by the expected frequency in each cell of the contingency table, weighted by twice the observed frequency:

$$G^2 = 2\left[f(n,v)\log\frac{f(n,v)}{E(n,v)} + \cdots + f(\neg n,\neg v)\log\frac{f(\neg n,\neg v)}{E(\neg n,\neg v)} \right]$$

Though, to my knowledge, significance measures have never been applied directly to LVCs, they are widespread in collocation analysis.

Contextual measures. Apart from the association measures based on contingency table values, Zhào et al.'s (2015; 2016) studies on Tibetan also considered the entropies of the previous and next tokens. These values quantify how diverse the words preceding and following the LVC are, respectively. A more conventionalised LVC would be expected to appear in a wider range of contexts:

$$H_{prev}(n,v) = -\sum_{w_{prev}} \left(p(w_{prev}|n,v)\log_2 p(w_{prev}|n,v\right)$$

$$H_{next}(n,v) = -\sum_{w_{next}} \left(p(w_{next}|n,v)\log_2 p(w_{next}|n,v\right)$$

Here, w_{prev} and w_{next} represent distinct previous and next word types, and $\sum$ is the summation symbol. Entropy is obtained by calculating the probability of each next word type multiplied by the log of itself, and summing the results. A noun-verb combination appearing in predictable contexts has low context entropy, and one appearing in diverse contexts has high context entropy.

2.3 (A)symmetry and direction in collocation analysis

A popular notion of asymmetry or directionality in English corpus linguistics comes from Kjellmer (2014, 112–115). Kjellmer distinguished between 'left-predictive' and 'right-predictive' collocations. In left-predictive collocations like *wishful thinking*, the first form strongly predicts the second one but not vice versa; in right-predictive collocations like *from afar*, the second element strongly predicts the first one. To enhance clarity and use terminology that applies to languages with lexicalised N-V combinations regardless of word order (and writing direction), I will refer instead to 'noun-to-verb (N2V)-predicting' (where the noun strongly predicts the verb but not vice versa) vs 'verb-to-noun (V2N)-predicting' (where the verb strongly predicts the noun but not vice versa) noun-verb combinations. Thus, to use English examples, *take* + *nap* is N2V, while *spite* + *face* is V2N. Both notions will be considered when investigating statistical properties in this paper. Generally, we expect LVCs to be N2V-predicting, but not V2N-predicting.

There is relatively little work on asymmetric measures for LVC detection. One exception, inspired by Dras and Johnson's (1996) related work, is from Tan et al. (2006, 50), who propose the following measure:

$$DJ = f(n,v)f(v)$$

Thus, this measure also considers how often the verb appears alone; for two sequences with each co-occurrence frequency between the noun and the verb, the LVC with the more frequent verb would have a higher DJ value. In a similar vein, Nagy and Vincze (2011, 4–6) and Vincze et al. (2011, 118–120) see improvements in some performance measures after adding a binary feature on whether a verb belongs to the 15 most common verbs in the corpus.

Other than these purely frequency-based measures, there are two other types of asymmetric measures. Firstly, in asymmetric/ unidirectional association measures, there would be separate measures for the nominal's attraction to the verb (which would be higher for V2N-predicting combinations) and the verb's attraction to the nominal (which would be higher for N2V-predicting combinations).

Michelbacher et al. (2007, 368–369) propose two asymmetric association measures. The first is conditional probabilities, i.e. $p(v \mid n)$ and $p(n \mid v)$. The second method is rank-based: For each word, they list its collocates in terms of χ^2-statistics, and use the rank of the collocate in the list as a measure of how much the collocate is attracted to the word. The higher the value of the ranking (i.e. the lower the ranking), the less attracted the collocate is to the node. Michelbacher et al. (2011, 254–257) generalise this measure to apply to arbitrary bidirectional measures, including G^2; this paper adopts this measure, denoting the rank statistic as rank(G^2).

Gries (2022, Sec. 2.1.2) offers unidirectional association measures based on the normalised Kullback-Leibler Divergence (KLD), which measures how different a distribution is from a baseline distribution. When examining how much a verb is attracted to a nominal, we compare the distribution of verbs given the noun to the distribution of verbs in the corpus overall using the following KLD formula:

$$KLD_{v \to n} = p(v \mid n) \log_2 \frac{p(v \mid n)}{p(v)} + p(\neg v \mid n) \log_2 \frac{p(\neg v \mid n)}{p(\neg v)}$$

If there is no association between the verb and the nominal, there would be little difference between $p(v \mid n)$ and $p(v)$ and between $p(\neg v \mid n)$ and $p(\neg v)$, pushing the KLD close to 0. KLD is normalised with the following formula so that it lies between 0 and 1:

$$KLD^{norm}_{v \to n} = 1 - e^{-KLD_{v \to n}}$$

The nominal's attraction to the verb is calculated identically, just swapping the v and n around:

$$KLD_{v \to n} = p(n \mid v) \log_2 \frac{p(n \mid v)}{p(n)} + p(\neg n \mid v) \log_2 \frac{p(\neg n \mid v)}{p(\neg n)},$$
$$KLD^{norm}_{v \to n} = 1 - e^{-KLD_{v \to n}}$$

Finally, *ΔP* (Gries, 2013, 143–155; Desagulier, 2016, 189–195) takes the difference between the probability of the verb given the noun and the probability of the verb given all other nouns (for measuring the attraction of the verb to the noun), and the difference between the probability of the noun given the verb and the probability of the noun given other verbs (for measuring the attraction of the noun to the verb).

$$\Delta P_{n|v} = p(n \mid v) - p(n \mid \neg v), \Delta P_{v|n} = p(v \mid n) - p(v|\neg n)$$

The second type of measures examines how flexible one slot in a collocation is, given another slot, which I refer to as 'slot productivity measures'. For LVCs, this means how flexible is the verb given that we know the nominal, and vice versa? Gries (2022, Sec. 2.1.3- 2.1.4), for example, proposes two such measures. Firstly, he offers 'type frequency' (TF). In the LVC case, this means the number of distinct nouns that can appear with the verb (*TF*(*v*)) and vice versa (*TF*(*n*)). Secondly, he offers 'normalised entropy', which measures how predictable the distribution of nominals is given the verb, and vice versa:

$$H_{norm}(v \mid n) = \frac{-\sum_v p(v \mid n) \log_2 p(v \mid n)}{\log_2 f(n)},$$
$$H_{norm}(n \mid v) = \frac{-\sum_n p(n \mid v) \log_2 p(n \mid v)}{\log_2 f(v)}$$

Thus, if a verb is hard to guess from the accompanying nominal, then $H_{norm}(v|n)$ would be high; if the nominal is easy to guess from the accompanying verb, $H_{norm}(n|v)$ would be low. A summary of measures is given in Table 2.

3 Problem and predicted patterns

In this section, I will first introduce N-V combinations other than LVCs in Tibetan, along with general expectations about distributional properties from an intuitive, rather than quantitative, perspective. I then make predictions about how these constructions differ distributionally from LVCs with respect to particular measures.

3.1 N-V combinations other than LVCs in Modern Tibetan

As mentioned in Section 2.1, the term 'light verb construction' is uncommon in Tibetan linguistics. Some common alternatives like 'compound verb' or 'complex predicate' do not invoke a semantically lighter verb – perhaps precisely because

Tab. 2: List of measures reviewed in Section 2.

Measure	Symbol	Symmetry	Concept operationalised
Raw co-occurrence frequency	$f(n,v)$	Symmetric	Co-occurrence frequency
Pointwise mutual information	PMI	Symmetric	Association strength
G-squared statistic	G^2	Symmetric	
Context entropy	$H_{\text{prev}}(n,v)$	Symmetric	Context diversity
	$H_{\text{next}}(n,v)$		
Dras-Johnson measure	DJ	Asymmetric	Verb frequency, co-occurrence frequency
Conditional probability	$p(n \mid v)$, $p(v \mid n)$	Asymmetric	Association strength
G-squared rank	$\text{rank}_{(v \to n)}$ (G^2)	Asymmetric	
	$\text{rank}_{(n \to v)}$ (G^2)		
Normalised Kullback-Leibler divergence	$KLD_{(v \to n)}{}^{\text{norm}}$	Asymmetric	
	$KLD_{(n \to v)}{}^{\text{norm}}$		
ΔP	$\Delta P_{(v \mid n)}$, $\Delta P_{(n \mid v)}$	Asymmetric	
Type frequency	$TF(n)$, $TF(v)$	Asymmetric	Flexibility of one slot given the other
Normalised entropy	$H_{\text{norm}}(v \mid n)$	Asymmetric	
	$H_{\text{norm}}(n \mid v)$		

there are constructions placed in this traditional category where the verb is not actually lighter.

Firstly, sometimes the nominal is highly predictable from the verb, rather than vice versa, such as (2a). Another example is as follows – Randall (2016, 37) considers all these complex predicates:

(7) a. གྲོད་ཁོག་རྒྱགས་
grod.khog rgyags
stomach full
'to be full'

b. གྲོད་ཁོག་ལྟོགས་
grod.khog ltogs
stomach hungry
'to be hungry'

c. གྲོད་ཁོག་བཤལ་
grod.khog bshal
stomach cleanse
'to have diarrhea'

The verb in each case, especially (7a & 7b), is strongly associated with stomachs semantically (in (7c), *bshal* seems to be most commonly used in the 'diarrhea' sense). I call these 'light noun constructions' (abbreviated LN). The nominals generally do not seem as 'light' as light verbs: They are predictable from the more specific verbs but still carry significant semantic weight on their own. Thus, these constructions are likely highly V2N-predictive but also somewhat N2V-predictive.

Secondly, sometimes the verb and noun provide similar semantic information.

(8) a. རྐུན་མ་རྐུ་
rkun.ma rku
thief steal
'to steal'

b. ཟས་ཟ་
zas za
food eat
'to eat'

c. སེམས་ཐག་གཅོད་
sems.thag gcod
resolve cut
'to make up one's mind'

When the verb and noun are etymologically related, as in (8a) & (8b), this is called 'lexical reduplication' by Randall (2016, 113–115), or a 'cognate object construction' in other languages. While the two components are not exactly equivalent in (8c), *gcod* 'cut' has several metaphorical meanings like 'decide, solve', and hence also has relatively large overlap with the nominal. To include examples like (8c) and exclude examples like རྒྱག་གཏམ་རྒྱག་ *rgyag.gtam rgyag* 'to make harsh remarks' (Randall, 2016, 114), where the nominal contains a root in the verb but is not predictable from the verb, I call these 'mutually predictable constructions' (abbreviated MUTUAL), which are most likely both N2V-predictive and V2N-predictive.

Finally, there exist conventionalised collocations in Tibetan where the verb and noun have clearly distinct semantic contributions, and yet remain highly associated. While some work has put these constructions in the same category as LVCs and the two other categories above, most previous work that do explicitly delineate the boundaries of complex predicates, phrasal verbs, etc. (e.g. Agha, 1993; Bartee, 2007; Randall, 2016), have focused on distinguishing them from constructions like these, which I call DISTINCT. Compared to the other construction types, I expected these to be less N2V-predictive and V2N-predictive (though of course, I still expect there to be some predictive power).

(9) a. ཡི་གེ་གཏོང་
yi.ge gtong
letter send
'to send a letter'

b. དེབ་ཀློག་ས་
deb klogs
book read
'to read a book'

c. མིང་འདོགས་
ming 'dogs
name hang
'to give a nickname'

3.2 Metrics examined and predictions

In this study, I will test most of the measures discussed in Section 2 on their ability to distinguish LVCs from other constructions. Since I am artificially restricting discussion to a subset of the full range of frequency values (i.e. taking the most frequent ones), measures based only on raw frequency will be dropped. For G^2 ranks, I use ascending ranks to keep the intuition that higher value means higher association. Ranks are defined to be maximal when G^2 is undefined, a phenomenon caused by the noun appearing only with the verb or vice versa, indicating very high attraction. Since distributions with many low probabilities are often difficult to visualise raw, the conditional probability measure is replaced by negative log-probabilities, equivalent to 'surprisal' in information theory.

For bidirectional association and context entropy measures used in previous studies, there is no predicted difference between LVCs and the other three categories, since they primarily differ from LVCs in terms of the differential roles of the noun and verbs.

Since it is generally easy to guess what kind of LVCs will follow a nominal, but very difficult to predict in the other direction, I predict LVCs to have the highest conditional surprisal for nouns given the verb (i.e. lowest V2N-predictability) and lowest surprisal for verbs given the noun (i.e. highest N2V-predictability), except perhaps for MUTUAL constructions, where the verb is also highly predictable from the noun. By the same token, I predict LVCs to have higher G^2 ranks (since there are many nouns for each given verb) and lower KLD and ΔP for nouns given verbs, and lower G^2 ranks and higher KLD and ΔP for verbs given nouns.

For slot flexibility measures, I predict that LVCs will have the highest type frequency and normalised entropy for verbs, since they tend to be highly flexible and compatible with a large number of event-denoting nouns, but make no predictions for nouns.

Table 3 shows the measures examined in this study, along with my predictions.

4 Methodology

Zhào et al.'s approach detects light verbs based on a large proprietary dataset of raw texts. Their error analyses reveal a fair number of false positives due to incorrect tokenisation, such as inappropriately selecting the last two syllables of longer light verb nominals, and examination of their results suggests that they may have been biased towards political content.

Tab. 3: Table of expectations about measures explored in this paper.

Measure type	Measure	Symbol	Compared to LVCs …		
			LN	DISTINCT	MUTUAL
Bidirectional association	Pointwise mutual information	PMI	No expected difference		
	G-squared statistic	G^2	No expected difference		
Context diversity	Context entropy	$H_{prev}(n,v)$ $H_{next}(n,v)$	No expected difference		
Unidirectional association	Conditional surprisal	$-\log_2 p(n \mid v)$ $-\log_2 p(v \mid n)$	Lower Higher	Lower Higher	Lower /
	Chi-squared rank	$rank_{(n\rightarrow v)}(G^2)$ $rank_{(v\rightarrow n)}(G^2)$	Lower Higher	Lower Higher	Lower /
	Normalised KLD	$KLD_{(n\rightarrow v)}{}^{norm}$ $KLD_{(v\rightarrow n)}{}^{norm}$	Higher Lower	Higher Lower	Higher /
	ΔP	$\Delta P_{(n \mid v)}$ $\Delta P_{(v \mid n)}$	Higher Lower	Higher Lower	Higher /
Slot flexibility	Type frequency	$TF(v)$ $TF(n)$	Lower	Lower No expected difference	Lower
	Normalised entropy	$H_{norm}(n \mid v)$ $H_{norm}(v \mid n)$	Lower	Lower No expected difference	Lower

Unfortunately, large corpora of modern Tibetan are not currently publicly available. Moreover, because of advances in computational resources for Tibetan computational linguistics since then – including open-source tools like *botok* (Esukhia, 2023) for tokenisation – it may be more promising to detect light verbs from a tokenised corpus, minimising false positives due to inappropriate tokenisation. I thus base my study instead on a smaller, open-source corpus (~1.3M word tokens) which is balanced and tokenised, the Nanhai Corpus (Schmidt, 2019), then process it using a variety of lexical resources.

My first step is to extract noun-verb combinations, where the noun takes no case or relator noun, and acts as an argument or argument-like complex predicate nominal (i.e. not part of an adverbial phrase modifying the verb – to use an English analogy, if one were to say *I took a walk in the park yesterday*, *took* would be paired with *walk*, not *in the park* or *yesterday*). The calculations were then based on these combinations. Thus my methodology bears similarities with covarying collexeme analysis (Stefanowitsch and Gries, 2005), though it is unclear that all the combinations extracted indeed belong to a single construction (see Section 6 for theoretical discussion).

The steps I followed were as follows. The final steps as described here are the result of many iterations of trial and error, manually noting sources of false positives and false negatives and modifying the automatic process to fix them using filters and exceptions:

– Step 1: Get candidate verbs by stemming and identifying all verbs in the corpus.
– Step 2: Get candidate nominals using existing dictionary resources.
– Step 3: For each candidate verb, if there is a candidate noun before it, and there is no intervening argument, case marker, copula or quotative marker, then classify that noun-verb combination as a candidate noun-verb combination.
– Step 4: Lemmatise the verbs and create a frequency list of candidate noun-verb combinations.

Since large POS-tagged corpora of modern Tibetan are not, at the time of writing, publicly available (though this may soon change), I used a simple rule-based method for the preprocessing steps, using dictionary resources to identify potential lexicalised noun-verb combinations.

For Step 1, I began by creating a stemmed version of the corpus by removing verbal agglutinative morphology tokenised from all words, including subordinators དུས་ *dus*, ནས་ *nas*, ན་ *na*, བཞིན་ *bzhin*, nominalisers and light nouns ཡ་ *ya*, བ *ba*, པ་ *pa*, མཁན་ *mkhan*, རྒྱུ་ *rgyu*, སྟངས་ *stangs*, སྲོལ་ *srol*, ཐབས་ *thabs*, བཞིན་ *bzhin*, ས་ *sa*, ཡུལ་ *yul*, རྩིས་ *rtsis*, auxiliary འདོད་ *'dod*, and negators མ་ *ma* and མི་ *mi*. Exceptions were made for common false positives (e.g. མ་འོངས་པ་ *ma.ongs.pa* which usually means 'future', not 'not having come'). I then retrieved the POS of the stemmed (but unlemmatised) word from the Monlam dictionary (Monlam, 2016), since its JSON version contains POS information, and also determined whether it was considered a verb in Hill (2010). A word satisfying both criteria, not homonymous with a case form (other than ན་ *na*, more commonly a verb 'to be ill'), the general extender སོགས་ *sogs* or the indefinite article forms ཞིག་ *zhig*/ཤིག་ *shig* was considered a candidate verb.

For Step 2, I first identified words that were most likely nouns, i.e. words that either had 'noun' as their sole POS listed in the Monlam dictionary, or had NOUN as their most common POS tag in ACTib corpus of Classical Tibetan (Hill and Meelen, 2017). (Both resources were important, since the ACTib list lacks modern words, and Monlam is relatively small.) Some exceptions were again filtered out, and false negatives common in the Nanhai corpus were brought back into the list. These false negatives were found by searching headwords in Steinert (2023), a resource combining multiple existing lexicographic sources. For each headword, I looked for other headwords consisting of the current headword plus a verb identified in Step 1 (possibly followed by the infinitive པ་ *pa*/བ་ *ba*). Headwords that satisfied this but

were not previously identified as nouns were examined and, if I identified them as nouns, added back to the noun list.

For Step 3, I separated each text into smaller text segments with the Tibetan punctuation mark *shad* ། , and within each segment, considered all the candidate verbs. For each candidate verb, I identified an accompanying caseless noun as follows: If there is a noun occurring between the verb and the previous verb (or the start of the text segment if there is no previous verb), and there are no case markers, quotative markers, copulas, or words found in neither Monlam nor ACTib intervening between the noun and the candidate verb, then I extracted the noun + candidate verb combination. A small handful of nouns, e.g. time and measure words usually appearing in adverbials, were skipped over.

Of the noun-verb combinations extracted, I excluded the ones where the noun was a relational noun; these nouns act like postpositions to the more substantive nouns they follow, and thus do not form a single unit with the light verb. I then manually went through examples where the noun and verb are not adjacent. Interveners that appeared more than twice in the corpus were examined and the entries with interveners indicative of problems were removed – this mostly included nouns not recognised by the above method (including relator nouns), phrases indicating that the noun was not directly dependent on the verb (such as subordinators), and adjectives commonly participating in LVCs.

For Step 4, for each noun-verb combination, I used Hill (2010) to determine the alternative forms of each verb stem. I then take the most common form of the verb stem in the corpus as the lemmatised form.

I went through the resulting candidate N-V combinations and marked entries appearing more than 10 times as light verb, light noun, mutually predictive or distinct, resulting in n = 155 noun-verb combinations whose statistical properties will be examined below. If a false positive was found, I examined other LVCs with the same noun or verb to determine whether they were false positives too.

5 Results

In this section, I will examine the results of each statistical measure from Section 3. For association and context diversity measures, I compare the measures across the four categories of noun-verb constructions using Wilcoxo's rank-sum tests, with three contrasts per measure: between LVCs and each of LN, DISTINCT and MUTUAL. Significance levels were at .05 with Holm-Bonferroni corrections among the three contrasts. Undefined values (caused by zero counts in the case of G^2) were removed. For the slot flexibility measures, since each value is associated with multiple N-V

combinations with the same noun or verb, I instead fit a Dirichlet regression with the location-scale parametrisation, the proportion of each N-V combination type as the response variable, and LVC as the baseline category, and examined the *p*-values for whether the fixed effects of the slot flexibility measures are significantly different from 0, with the same correction. A significant coefficient for a non-LVC category indicates that the odds of a construction belonging to that non-LVC category over the LVC category varies according to the slot flexibility measure in question.

5.1 Bidirectional measures of association

Examining the PMIs of different collocations reveals that high-frequency LVCs (Figure 1), while having highly positive PMI values, actually tend to have lower PMI values than high-frequency LN, DISTINCT and MUTUAL constructions. In particular, LN and LV constructions have barely any overlap in distribution. For G^2 values, LVCs are similar to DISTINCT constructions, but still much lower than LN and MUTUAL constructions. Wilcoxon tests show that PMIs' distribution is significantly different than the other three types except for DISTINCT (p = 0.00280 for LN, 0.37182 for DISTINCT, 0.00030 for MUTUAL), though not G^2 values (p = 0.0467 for LN, 0.0189 for DISTINCT, 0.4013 for MUTUAL). These results heavily suggest that traditional LVC detection methods that use bidirectional association will, if anything, end up favouring types of frequent N-V combinations other than LVCs.

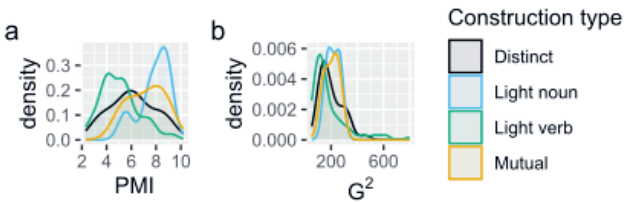

Fig. 1: Kernel density estimates with Gaussian kernels with standard deviation bandwidths for (a) PMIs and (b) χ^2.

5.2 Context diversity measures

Context entropy, as discussed by Zhào et al. (2016, 140–141), does not seem to distinguish between the different types clearly, as there is considerable overlap in the

data from all the construction types, as shown in Figure 2. Wilcoxon tests show a significant difference between LVCs and DISTINCT constructions, but not the other two (preceding tokens: p = 0.1653 for LN, 0.0045 for DISTINCT, 0.9247 for MUTUAL; following tokens: p = 0.55 for LN, 0.000014 for DISTINCT, 0.90 for MUTUAL), most likely simply because DISTINCT constructions are more common, leading to a larger sample size.

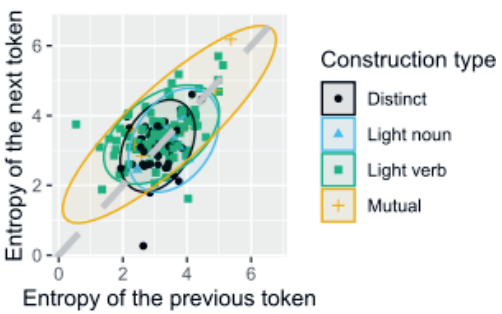

Fig. 2: Context entropy with estimated 95% bivariate normal ellipses.

5.3 Unidirectional association measures

The unidirectional association measures show very clear separation between LVCs and other types of constructions (see Figure 3). In fact, all four show the same pattern: In LVCs, the verb is attracted to the noun far more than the noun to the verb. Again, LNCs are especially divergent from LVCs.

Other than rank(χ^2), the unidirectional association measures encode virtually identical information, as shown by the scatterplots and correlations in Figure 4 on page 137.

This pattern can be easily explained. The $p(n \mid \neg v)$ value (in ΔP) and $\log_2 \frac{p(n \mid v)}{p(n)}$ value (in the KLD) are virtually negligible in this case, where a very large number of nouns corresponds to a very small number of verbs. For ΔP, this means the value is virtually equivalent to $p(n \mid v)$. For KLD, since each noun is very rare in the corpus, this means $p(\neg n) \approx 1$. Hence,

$$KLD^{norm}_{n \to v} \approx 1 - \exp(-p(\neg n \mid v) \log_2 p(\neg n \mid v))$$

which depends only on $p(\neg n \mid v) = 1 - p(n \mid v)$. Wilcoxon tests were thus excluded for ΔP and KLD. Remaining tests, shown in Table 4, found support for all differences except the ones for the verb's attraction to the noun in the cases of MUTUAL and LN.

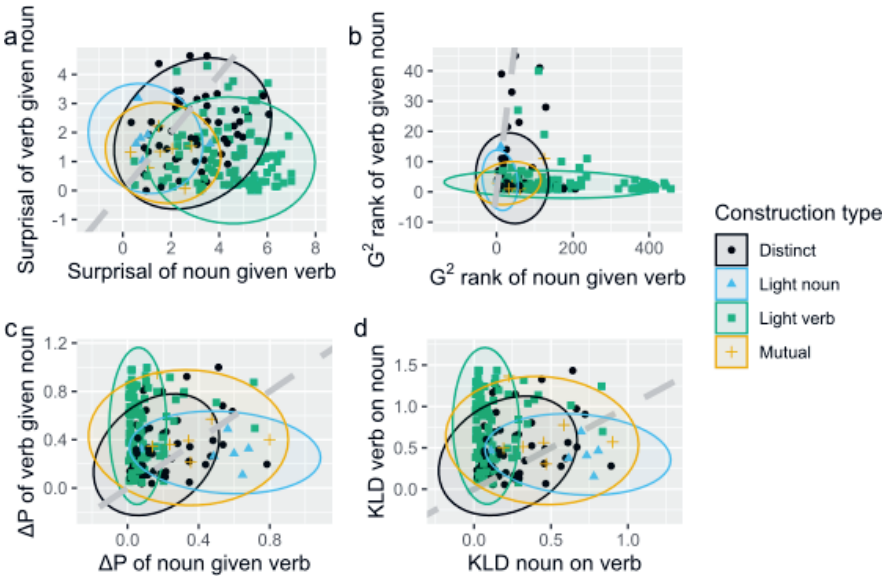

Fig. 3: (a) Surprisal, (b) G^2 ranks (with undefined values defined to have highest rank), (c) ΔP and (d) KLDs with estimated 95% bivariate normal ellipses. The dashed line indicates positions where the unidirectional measure is the same in both directions.

5.4 Slot flexibility measures

There were also clear relationships between slot flexibility measures and type of expression. Figure 5 shows the normalised entropy and type frequency values of verbs, along with a corresponding barplot showing the proportion of each category. LVCs are most common at higher values, while LNCs and to some extent MUTUAL are more common at the lower end, and DISTINCT constructions appear throughout. The opposite situation is shown in Figure 6, where LVCs are concentrated at the low end of the spectrum, and the other three categories are concentrated on the high end. Figure 7 shows that the type frequency and entropy measures are also correlated, albeit not as much as between the unidirectional association measures.

The Dirichlet regression model revealed that type frequency of nouns and normalised entropy of verbs given nouns are significantly associated with the odds of getting LNCs over LVCs and DISTINCT constructions, but clearly not for MUTUAL ones (Table 5). Likely due to the low power, there was no comparable result for verbs.

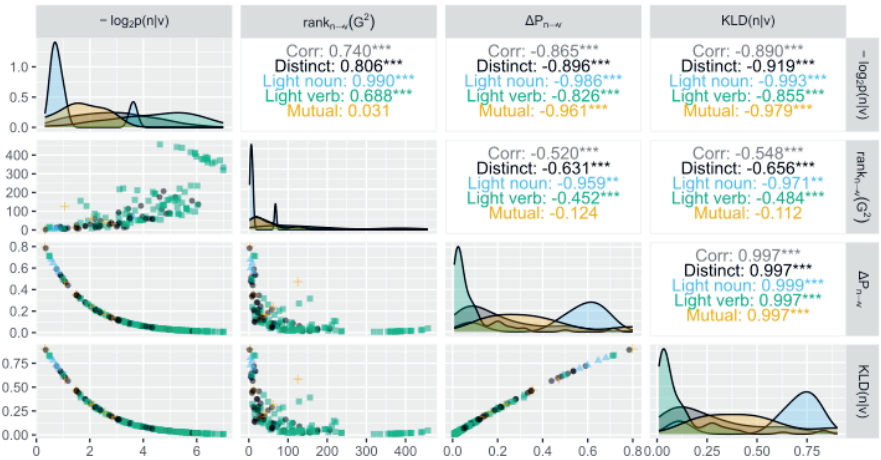

Fig. 4: Correlations between surprisal, G^2 ranks, ΔP and KLDs, for measurements of the noun's attraction to the verb. Surprisal, ΔP and KLDs are virtually just rescalings of each other. A similar pattern holds for the verb's attraction to the noun but is not shown here.

Tab. 4: *P*-values of Wilcoxon tests between LVCs and the three other types, with significant results (at .05 significance level and Bonferroni correction with $g = 3$) shaded.

LVCs ...	with LN	with DISTINCT	with MUTUAL
$-\log_2 p(n \mid v)$	0.00040	6.0×10^{-7}	6.3×10^{-5}
$-\log_2 p(v \mid n)$	0.081	1.4×10^{-7}	0.44
$\text{rank}_{(n \to v)}(G^2)$	0.00024	1.3×10^{-11}	0.00023
$\text{rank}_{(v \to n)}(G^2)$	0.2623	0.0016	0.5598

5.5 Combining measures

Summing up the results above, we get Table 6. Differences that are significant but small are written as 'slightly', and differences that are insignificant but visually clear and have relatively small *p*-values are written as 'possibly' higher/ lower.

A question that arises from these results is whether the various measures that show differences between LVCs and the three other construction types provide overlapping or redundant information. A principal components analysis of the above measures (other than the context diversity measures, replacing G^2 with χ^2, which has the same patterns, because of the presence of undefined values for G^2) finds that 87.5% of variation in the data can be represented in three dimensions. The variables are plotted in Figure 8 and the individuals in Figure 9.

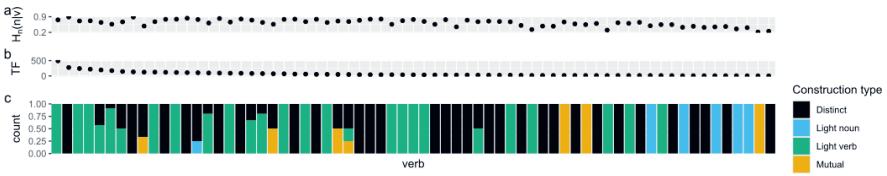

Fig. 5: (a) Normalised entropies of the noun given the verb, (b) type frequencies for each verb, and (c) proportion of the four constructions for each verb.

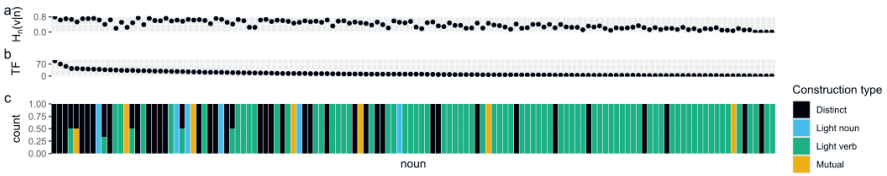

Fig. 6: (a) Normalised entropies of the verb given the noun, (b) type frequencies for each verb, and (c) proportion of the four constructions for each verb.

The first dimension (42.6% of the variance) is primarily correlated positively with high PMI and χ^2 (i.e. bidirectional association) and negatively with measures associated with lighter verbs/heavier nouns (i.e. verbs are more attracted to the noun, and nouns are more flexible given the verb); thus, LVCs tend to have very low values and LNCs high values, with MUTUAL/DISTINCT in between. The second dimension (29.9%) is mostly associated with high bidirectional association and low values of measures associated with lighter nouns/heavier verbs; hence the MUTUAL and LN constructions, where nouns are more predictable from the verb tend to have the highest values. The third dimension (8.8%) is dominated by high noun entropy, and weakly separates LNCs and MUTUAL (lower values) from some of the LVCs and DISTINCT constructions (higher values). The large overlaps in arrows in the diagrams indicate considerable overlap in the information encoded by different measures, with three main 'clusters' of measures indicating light nouns/heavy verbs, heavy verbs/light nouns and bidirectional association respectively, with the entropy of the noun given the verb being independent from all of these.

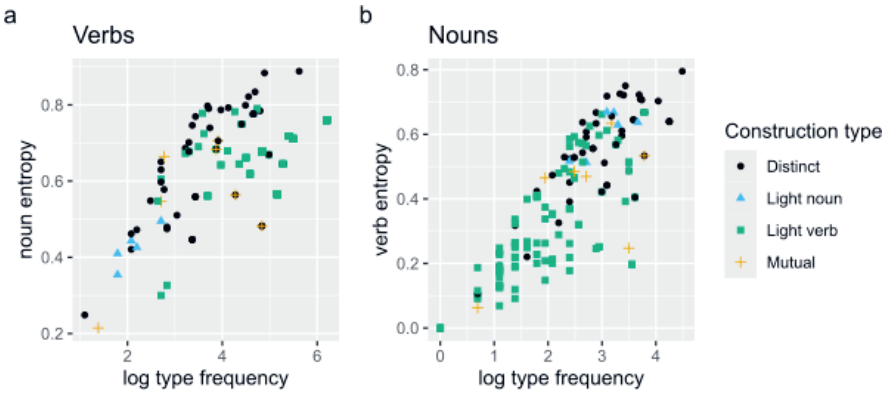

Fig. 7: Scatterplots of the relationship between logged type frequency and entropy for verbs (a) and nouns (b).

Tab. 5: *P*-values of Wald tests for the significance of Dirichlet regression coefficients, with significant results (at .05 significance level and Bonferroni correction with $g = 3$) shaded.

LVCs …	with LN	with DISTINCT	with MUTUAL
$TF(v)$	.133	.166	.0502
$TF(n)$	2.19×10^{-7}	.0147	.956
$H_{norm}(n \mid v)$	.158	.868	.697
$H_{norm}(v \mid n)$	1.82×10^{-12}	.00427	.314

6 Discussion

6.1 Asymmetry in measuring and detecting co-occurrence

The results of this paper show that, while linguists have traditionally assumed that LVCs can be detected using high bidirectional association, LVCs actually have weaker bidirectional association than other conventionalised N-V pairings at high frequencies. Thus, the corpus linguist looking to detect LVCs may need to consider a nonlinear relationship between bidirectional association and LVC status to tease LVCs apart from other types of constructions.

More generally, collocation and covarying collexeme analysis would benefit from looking beyond bidirectional measures of association, and start combining them with asymmetric measures, including both unidirectional association measures and slot flexibility measures (Gries, 2013, 2022). In this case I found substantial redundancy between different measures – including some variables that, because

Tab. 6: How different measures distinguish LVCs from other types of conventionalised noun-verb combinations. Significant results are shaded in grey.

Measures type	Measure	Compared to LVs ...		
		LN	DISTINCT	MUTUAL
Bidirectional association	PMI	Much higher	/	Higher
	G^2	Possibly higher	/	Possibly higher
Context diversity	$H_{prev}(n,v)$	/	Slightly higher	/
	$H_{next}(n,v)$	/	Slightly lower	/
Unidirectional association	$-\log_2 p(n \mid v)$	Lower	Lower	Lower
	$-\log_2 p(v \mid n)$	Possibly higher	Higher	/
	$\text{rank}_{(n \to v)}(G^2)$	Lower	Lower	Lower
	$\text{rank}_{(v \to n)}(G^2)$	Possibly higher	Higher	/
Slot flexibility	$TF(v)$	Possibly lower	Possibly lower	Possibly lower
	$TF(n)$	Higher	Higher	/
	$H_{norm}(n \mid v)$	Possibly lower	/	/
	$H_{norm}(v \mid n)$	Higher	Higher	/

of the properties of the current dataset, are virtually rescalings of each other, but there remain at least three major clusters of variables, plus an additional variable (noun entropy) independent of the rest; they all seem to play different roles in distinguishing the four types of constructions.

6.2 Slot asymmetry and the construction hierarchy

As mentioned in section 2, Tibetan linguistics literature typically uses not the term LVC, but other terms like 'compound verb' and 'complex predicate' that consist mostly of LVCs, but also contain constructions of the other three categories. For example, Randall's (2016) complex predicate includes many LN and MUTUAL constructions, and even at least one DISTINCT construction ཁོག་པ་སྙོག་ *khog.pa sngog* 'inside' + 'dig' = 'ferret out information'.

While LVCs and LNCs both typically have one relatively productive and one relatively unproductive slot, DISTINCT constructions have two productive slots and MUTUAL ones are lexicalised with two unproductive slots. Borrowing terminology from the literature on serial verb constructions (a different type of complex predicate; Aikhenvald and Dixon, 2006), the first two are asymmetric and the second two are symmetric. Randall's grammatical tests show that there are constructions from the four categories which can all be seen as belonging to one overarching complex predicate construction, with LVCs being particularly asymmetric and hence

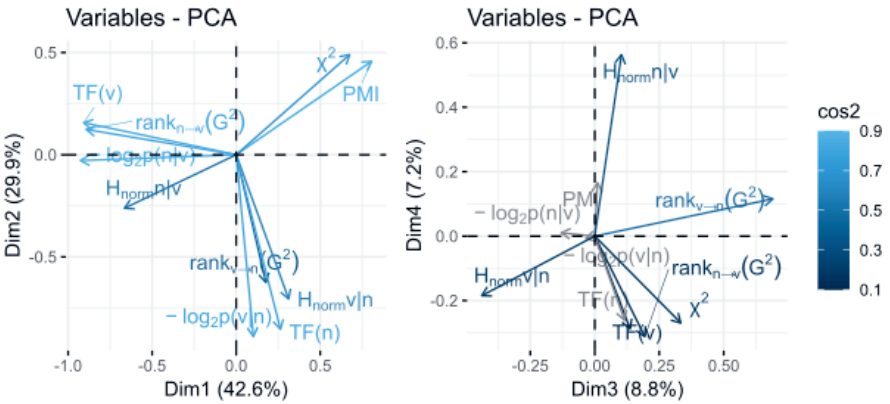

Fig. 8: Map of the different variables on the first four principal dimensions. The colours indicate squared cosine values: The larger the number, the greater the amount of variation inside that variable is represented by the dimensions in the graph.

the fixed verbal slot (light verbs) acts like a grammatical verbaliser, and the rest having less grammaticalised properties. This situation resembles what Lai and Pang (2023) describe for the Cantonese causative-resultative construction: this serial verb superconstruction also subsumes asymmetric subconstructions with the first slot highly productive and second one unproductive or vice versa, and relatively symmetric subconstructions where both slots are highly (un)productive, contra established assumptions that serial verb constructions in a specific language can be enumerated and classified simply into symmetric vs asymmetric. The Tibetan data suggests that a similar situation obtains for complex predicates derived from the dereferentialisation of the argument in argument-verb constructions.

7 Conclusion

In this paper, I explore several measures of conventionalisation as they apply to high-frequency noun-verb combinations in Tibetan to see how well they distinguish light verb constructions from other constructions, including light noun constructions, mutually predictive constructions, and constructions whose components have distinctive semantics. I find context diversity does not clearly distinguish between them. Bidirectional association measures, whose high values are typically taken to be strong indicators of LVC status, turn out to be lower for LVCs than the other three. By contrast, unidirectional association and slot flexibility measures better distinguish between LVCs from other construction types; while there is substantial redundancy

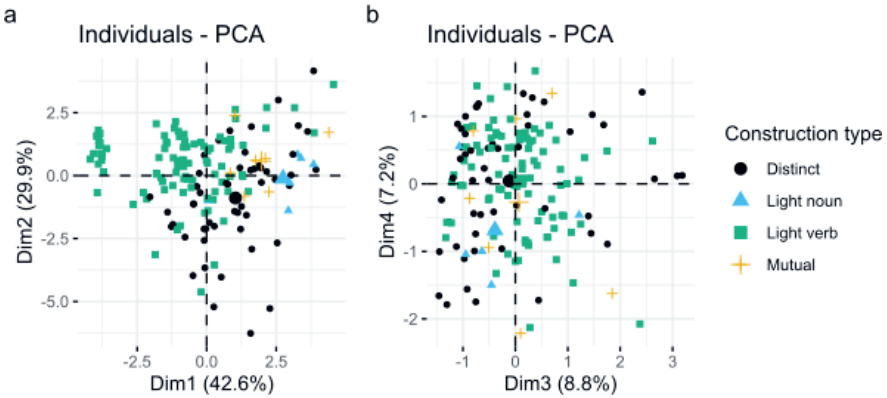

Fig. 9: Map of the different N-V combinations on the first four principal dimensions.

between these measures, they do have contributions independent of bidirectional association, and noun entropy may have a contribution independent of all other factors. I conclude that to characterise the statistical distributional properties of LVCs, it is important to look beyond classic association measures, and examine measures that zero in on the differences between the nominal and verbal slots. It is hoped that future LVC detection methods will add unidirectional association and slot flexibility measures.

Although the points raised in this paper are limited to Tibetan, they may also apply to light verb detection more generally. For example, English also has N-V collocations where the verb predicts the noun (*part ways*), the noun and the verb have similar semantic contributions (*sing a song, fire a shot*) or where the two elements' contributions are relatively distinct (*slice carrots*). It is also hoped that even more potentially useful characteristics of LVCs, such as semantic domains (Taslimipoor et al., 2012; Vaidya et al., 2016; Singh et al., 2016), distribution of interveners, argument structure (Simon, 2011; Singh et al., 2016), and register (Tibetan LVCs are considered more informal than MUTUAL and DISTINCT constructions with similar meanings; Geissler, 2018, 11) may be considered for the Tibetan LVC detection toolkit in the future. Finally, as this paper focuses on high-frequency combinations, it rests on the general assumption that LVCs are conventionalised combinations; however, it is possible that there may be LVCs that are created on an ad hoc basis. I leave it to future work to investigate this issue in more detail, which may require larger datasets and more precise measurements of productivity than have been explored in this paper.

Acknowledgments

This chapter was written as part of my PhD thesis at UC Santa Barbara. I would like to thank my committee, Stefan Th Gries, John W DuBois, Wendy Meiring and Satoko Shirai as well as members of UCSB's CEILing group for their valuable comments.

Bibliography

Agha, Asif. 1993. *Structural form and utterance context in Lhasa Tibetan: grammar and indexicality in a non-configurational language*. New York: Peter Lang.

Aikhenvald, Alexandra and Robert M. W. Dixon. 2006. *Serial verb constructions: a cross-linguistic typology*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199279159.001.0001.

Bailey, Geoff and Christopher E. Walker. 2004. *Lhasa verbs: a practical introduction*. Lhasa: Tibetan Academy of Social Science.

Bartee, Ellen Lynn. 2007. *A grammar of Dongwang Tibetan*. University of California, Santa Barbara.

DeLancey, Scott. 1991. The origins of verb serialization in modern Tibetan. *Studies in Language* 15(1). 1–23. https://doi.org/10.1075/sl.15.1.02del.

Denwood, Philip. 1999. *Tibetan*. Amsterdam: John Benjamins.

Desagulier, Guillaume. 2016. A lesson from associative learning: asymmetry and productivity in multiple-slot constructions. *Corpus Linguistics and Linguistic Theory* 12(2). 173–219. https://doi.org/10.1515/CLLT-2015-0012.

Dras, Mark and Michael Johnson. 1996. Death and lightness: using a demographic model to find support verbs. In *International Conference on the Cognitive Science of Natural Language Processing (5th: 1996)*, Dublin: Dublin City University Natural Language Group.

Dunning, Ted. 1994. Accurate methods for the statistics of surprise and coincidence. *Computational linguistics* 19(1). 61–74.

Esukhia. 2023. *botok*. https://github.com/OpenPecha, last access 27/9/2024.

Evert, Stephanie. 2005. *The statistics of word cooccurrences: word pairs and collocations*. Doctoral dissertation: Universität Stuttgart.

Geissler, Christopher. 2018. Phonological koinéization in Kathmandu Tibetan. In Gillian Gallagher, Maria Gouskova and Sora Yin (eds.), *Proceedings of the Annual Meetings on Phonology*, 1–12. Washinton D.C.: Linguistic Society of America. https://doi.org/10.3765/amp.v5i0.4242.

Gries, Stefan Th. 2013. 50-something years of work on collocations: What is or should be next …. *International Journal of Corpus Linguistics* 18(1). 137–166. https://doi.org/10.1075/ijcl.18.1.09gri.

Gries, Stefan Th. 2022. Multi-word units (and tokenization more generally): a multi-dimensional and largely information-theoretic approach. *Lexis: Journal in English Lexicology* 19(3). 1–23. https://doi.org/10.4000/lexis.6231.

Hackett, Paul. 2003. *A Tibetan verb lexicon: verbs, classes, and syntactic frames*. Ithaca: Snow Lion Publications.

Hill, Nathan W. 2010. *A lexicon of Tibetan verb stems as reported by the grammatical tradition*. Munich: Bayerische Akademie der Wissenschaften.

Hill, Nathan W. and Marieke Meelen. 2017. Segmenting and POS tagging Classical Tibetan using a memory-based tagger. *Himalayan Linguistics* 16(2). 64–86. https://doi.org/10.5070/H916234501.

Hoshi, Izumi. 2003. *Gendai Chibettogo Doushi Jiten [A verb dictionary of the modern spoken Tibetan of Lhasa]*. Tokyo: Tokyo University of Foreign Studies.

Kjellmer, Goran. 2014. A mint of phrases. In Karin Aijmer and Bengt Altenberg (eds.), *English Corpus Linguistics*, 123–139. London: Routledge.

Lai, Ryan Ka Yau and Michelle Man-Long Pang. 2023. Rethinking the description and typology of Cantonese causative–resultative constructions: a dynamic constructionist lens. *Languages* 8(2). 1–48. https://doi.org/10.3390/languages8020151.

Lessan Pezechki, Homa and Nicolas Tournadre. 2014. La question du sujet et les verbes support: Les cas du persan et du tibétain. Paper presented at Du sujet et de son absence. Le Mans, France.

Michelbacher, Lukas, Stefan Evert and Hinrich Schütze. 2011. Asymmetry in corpus-derived and human word associations. *Corpus Linguistics and Linguistic Theory* 7(2). 245–276. https://doi.org/10.1515/cllt.2011.012.

Michelbacher, Lukas, Stephanie Evert and Hinrich Schütze. 2007. Asymmetric association measures. In Ruslan Mitkov and Galia Angelova (eds.), *Proceedings of the 6th International Conference on Recent Advances in Natural Language Processing (ranlp)*, 367–372. Amsterdam: John Benjamins.

Monlam, Lobsang. 2016. *Monlam Tibetan-English Dictionary*. https://github.com/iamironrabbit/monlam-dictionary, last access 6/3/2023.

Nagy, István and Veronika Vincze. 2011. Identifying verbal collocations in Wikipedia articles. In Ivan Habernal and Václav Matoušek (eds.), *Text, Speech and Dialogue*, 179–186. Heidelberg: Springer. https://doi.org/10.1007/978-3-642-23538-2_23.

Randall, Michael Gordon. 2016. *The properties of Lhasa Tibetan verbalizers*. Master's thesis: Payap University, Chiang Mai.

Schmidt, Dirk. 2019. A speech corpus of Dharamsala Tibetan. 10.17613/r9x9f-a3174. Paper presented at HLS25: 25th Himalayan Languages Symposium. Sydney, Australia.

Simon, Camille. 2011. *Derivation causative en tibétain (Lhasa)*. Master's thesis: Université de Provence, Aix-en-Provence.

Singh, Dhirendra, Sudha Bhingardive and Pushpak Bhattacharyyaa. 2016. Detection of compound nouns and light verb constructions using IndoWordNet. In Verginica Barbu Mititelu, Corina Forascu, Christiane Fellbaum and Piek Vossen (eds.), *Proceedings of the 8th global WordNet conference (GWC)*, 404–410. Weesp.

Stefanowitsch, Anatol and Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1). 1–43. https://doi.org/10.1515/cllt.2005.1.1.1.

Steinert, Christian. 2023. *Tibetan-English dictionary application*. https://github.com/christiansteinert/tibetan-dictionary, last access 6/3/2023.

Stevenson, Suzanne, Afsaneh Fazly and Ryan North. 2004. Statistical measures of the semi-productivity of light verb constructions. In Takaaki Tanaka, Aline Villavicencio, Francis Bond and Anna Korhonen (eds.), *Proceedings of the Workshop on Multiword Expressions: Integrating Processing*, 1–8. Barcelona: Association for Computational Linguistics.

Tan, Yee Fan, Min-Yen Kan and Hang Cui. 2006. Extending corpus-based identification of light verb constructions using a supervised learning framework. In Paul Rayson, Serge Sharoff and Svenja Adolphs (eds.), *Proceedings of the Workshop on Multi-word-expressions in a multilingual context*, 49–56. Trento: Association for Computational Linguistics.

Taslimipoor, Shiva, Afsaneh Fazly and Ali Hamzeh. 2012. Using noun similarity to adapt an acceptability measure for Persian light verb constructions. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis (eds.), *Proceedings of the eighth international conference on language resources and evaluation (lrec'12)*, 670–673. Istanbul: European Language Resources Association (ELRA).

Tournadre, Nicolas. 1996. *L'ergativité en tibétain: approche morphosyntaxique de la langue parlée*. Louvain: Peeters.

Tournadre, Nicolas and Sangda Dorje. 1998. *Manuel de tibétain standard: langue et civilisation: introduction au tibétain standard (parlé et écrit) suivie d'un appendice consacré au tibétain littéraire classique*. Paris: L'Asiathèque-Maison des langues du monde.

Tu, Yuancheng and Dan Roth. 2011. Learning English light verb constructions: contextual or statistical. In Valia Kordoni, Carlos Ramisch and Aline Villavicencio (eds.), *Proceedings of the Workshop on Multiword Expressions: From Parsing and Generation to the Real World*, 31–39. Portland: Association for Computational Linguistics.

Vaidya, Ashwini, Sumeet Agarwal and Martha Palmer. 2016. Linguistic features for Hindi light verb construction identification. In Yuji Matsumoto and Rashmi Prasad (eds.), *Proceedings of Coling 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 1320–1329. Osaka: Coling.

Vincze, Veronika, István Nagy and Gábor Berend. 2011. Detecting noun compounds and light verb constructions: a contrastive study. In Valia Kordoni, Carlos Ramisch and Aline Villavicencio (eds.), *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, 116–121. Portland: Association for Computational Linguistics.

Zemp, Marius. 2018. *A grammar of Purik Tibetan*. Leiden: Brill.

Zhào, Wéinà, Lin Li, Huìdān Liú and Jiàn Wú. 2016. Tibetan trisyllabic light verb construction recognition. *Himalayan Linguistics* 15(1). 137–148. https://doi.org/10.5070/H915130102.

Zhào, Wéinà, Lín Lǐ, Huìdān Liú, Pǔbùdùnzhū and Jiàn Wú. 2015. Automatic extraction of trisyllabic verb phrases in Tibetan. *Journal of Chinese Information Processing* 29(3). 196–200.

Timur Maisak

Complex verbs in Kina Rutul (Nakh-Daghestanian)

1 Introduction: Complex verbs in Nakh-Daghestanian

Nakh-Daghestanian, also known as East Caucasian, is one of the three indigenous language families of the Caucasus together with Abkhaz-Adyghe (West Caucasian) and Kartvelian. Nakh-Daghestanian languages are spoken in the eastern part of the Caucasus, mainly in the Russian republics of Daghestan, Chechnya and Ingushetia, as well as in northern Azerbaijan and eastern Georgia. More than forty languages belong to the family, many of which are minority languages and remain unwritten and understudied; the precise number of languages remains unclear due to the problematicity of the language/ dialect distinction for quite a few idioms spoken in Daghestan. The main branches of the family, going from north-west to south-east, are Nakh, Avar-Andic, Tsezic, Lak, Dargwa, Lezgic and Khinalug; for a recent general overview of the family, see (Ganenkov and Maisak, 2020).

Although, to varying degrees, the contrast between simplex and complex verbs is relevant to all branches of Nakh-Daghestanian, it is especially well-pronounced in the Lezgic languages, which possess large amounts of complex verbs. In many of these languages, verbal compounding is also the only available means of creating new verbal lexemes: there is no other way to make borrowed verb stems inflected. The Lezgic branch, together with a single-language Khinalug branch, is southernmost: it includes nine languages, which are spoken in central and southern Daghestan (Archi, Agul, Tabasaran), or both in Daghestan and northern Azerbaijan (Rutul, Tsakhur, Lezgian), or exclusively on the Azerbaijani side of the current political border (Budugh, Kryz, Udi). Given long-standing contact with Azerbaijani, as well as, in earlier times, with Iranian languages, the Turkic and Iranian influence on the Lezgic languages is quite significant. The spread of complex verbs, which are very typical of both Turkic and Iranian, may be one of the manifestations of this contact influence.

In the Lezgic languages, complex verbs, also known as verbal compounds or light-verb constructions, consist of a lexical component and a postposed light verb. Lexical

Timur Maisak, Institute of Linguistics, Russian Academy of Sciences & HSE University, Moscow, Russia

https://doi.org/10.1515/9783112229330-006

components, sometimes labeled 'nominal parts' or 'coverbs',[1] can be represented by nouns, adjectives, adverbs, ideophones, and also acategorical bound stems which cannot function as autonomous words. Light verbs, which usually host all verbal inflectional marking (tense, aspect, mood, polarity, etc.), typically include such high-frequency verbs with a generalized meaning, as 'be, become', 'do, make', 'give', 'say', 'hit, beat', 'go', 'come' and some others. The most productive means of expanding verbal vocabulary is to borrow verbs from a dominant language (e.g. Russian or Azerbaijani) in combination with a light verb, 'be, become' for intransitives and 'do, make' for transitives. For brief overviews of complex verbs in Nakh-Daghestanian, see Klimov and Alekseev (1980, 182–183), Ganenkov and Maisak (2020, 109–111) and Maisak (2020a, 346–352), as well as papers in the volume on word-formation (Müller et al., 2016). In-depth studies of complex verbs in individual languages remain rare. In her book on the Udi morphosyntax, Harris (2002, ch.4) dedicates a special chapter to the analysis of the Udi complex verbs with respect to the wordhood criteria. In a recent study, Kerimova (2023) scrutinizes the morphosyntatic properties of coverbs in complex verbs, mainly focusing on the Lezgian data. A detailed analysis of complex verbs with the light verb 'say' in Archi is given by Authier (2024), who suggests a semantic classification of verbs belonging to this very numerous class.

An important problem in the description of complex verbs in Nakh-Daghestanian is the identification of the boundaries of this class. As a rule, complex verbs are not a homogeneous group, as coverbs can be phonologically, morphologically and syntactically 'fused' with light verbs to varying degrees. There may be no strict border between what is usually described as complex verbs and free syntactic combinations (or idioms, in case of non-compositional meaning). For example, in the literature it usually remains without any discussion, to what extent verbs like 'to help' which are expressed by a combination "help (noun) + do"[2], are different from ordinary patient–verb combinations like "house + build" or "letter + write".

The present paper aims at a description of complex verbs in Kina Rutul, a dialect of Rutul, a Nakh-Daghestanian language of the Lezgic branch spoken in southern Daghestan, Russia. The data for the study were collected during the author's fieldwork in 2018–2023. Apart from the introduction and the conclusion, there are three main sections. In Section 2, I present a brief linguistic characteristics of Kina Rutul in terms of major peculiarities of its phonological, morphological and syntactic structure. Section 3 gives an overview of structural types of complex verbs, including various types of "adjective + light verb", "noun + light verb", "acategorical coverb + light

1 The term 'coverb' (or 'co-verb') should not be confused with 'converb' which refers to a non-finite verb form that serves to express adverbial subordination.

2 Hereafter, I give the literal meaning of verbal compounds/ complex verbs in double quotes, while their proper meaning is cited in single quotes according to a general convention.

verb", "verb stem + light verb" and "locative adverb + verb" combinations. In Section 4, I look at the morphosyntactic properties of complex verbs, focusing on whether they can be treated as (compound) words.

2 Kina Rutul: The typological profile

Rutul is a relatively small Nakh-Daghestanian language of the Lezgic branch, mainly spoken in a dozen of villages of southern Daghestan (mostly Rutulsky District of the Republic of Daghestan, Russia) and a few settlements in the northern part of Azerbaijan. The total number of speakers is usually estimated as 30,000 or slightly more. Kina Rutul is a single-village variety of Rutul spoken in the village of Kina (Rutulsky District), which is located on the western periphery of the Rutul-speaking area and close to the area where Tsakhur, a sister language of Rutul, is spoken.

According to a dialectal classification of Rutul presented by Ibragimov (1978, 2004), the main dialects of the language are Mukhad, Shinaz, Myukhrek, Ikhrek and Borch-Knnov. Under this approach, Kina Rutul (as well as a few other single-village idioms) is an 'intermediate' or 'mixed' variety, which does not belong to any of these major dialects but rather accumulates linguistic features of Mukhad, Shinaz and Ikhrek dialects. An alternative may be to treat Kina Rutul as a small dialect on its own.

Rutul remains understudied, although there exist a number of grammar sketches of the language published since the early 20th century, including Dirr (1912); Ibragimov (1978, 2004); Alekseyev (1994) and Makhmudova (2001), all of which mostly treat phonological and morphological issues. None of these has the data on the Kina variety, which became a subject of dedicated research only in the last decade. Since 2016, a project aiming at the documentation of Kina Rutul has been running at the HSE University (Moscow, Russia), and the present contribution stems from the author's fieldwork fulfilled within this project. A collective descriptive volume on Kina Rutul is currently being prepared (Daniel et al., to appear), and in the present chapter, I largely follow the transcription and glossing conventions of the volume. In the description of Kina Rutul light verb constructions, I rely both on elicitation and the data from a spoken text corpus collected by the project members: the corpus mainly includes personal narratives and folk tales, and at present comprises ca. 20,000 word tokens. In the examples cited below, those taken from texts are accompanied by the indication 'Text Corpus'; no indication is given for those examples that were elicited.

Like the majority of Nakh-Daghestanian languages, Kina Rutul possesses a rich phonological inventory, with a three-way contrast between voiceless (aspirated), voiced (non-aspirated) and ejective stops (e.g., /k/ ∼ /g/ ∼ /k'/), a number of uvulars

(e.g., /χ/, /q/, /ɢ/), an opposition between short and long vowels (e.g., /a/ ∼ /a:/), and a prosodic pharyngealization feature acting as a secondary articulation with a peak on vowels and/ or uvulars (e.g., *raˤq* 'road', *qʼʷaˤd* 'two', *jɨˤχɨs* 'hit, beat').[3]

In the nominal paradigm, there is an opposition between an unmarked singular and a suffixally marked plural (e.g., *χal* 'house' ∼ *χal-bɨr* 'house-PL'). Each noun is assigned to one of the four genders, also known as '(noun) classes': Gender 1 is masculine, Gender 2 is feminine, the two other genders include non-human referents (most non-human animates belong to Gender 3 and most inanimates belong to Gender 4, but there are exceptions). The category of gender displays itself in agreement, mainly on verb stems, but also on cardinal numerals and the reflexive pronoun. The exponents of both series, which can be prefixes, infixes or suffixes, are single-consonant markers *r, b, d, w, j, l* and also zero. Gender agreement in noun phrases is controlled by the head, whereas on the clause level the controller is the absolutive noun phrase, be it the patient of a transitive verb or the single core argument of an intransitive verb.

Also mirroring a general Nakh-Daghestanian profile, Kina Rutul has a rich case inventory, with the unmarked absolutive case and all the other (oblique) cases being derived from a marked oblique stem. Rutul is a morphologically ergative language, thus agents of transitive verbs are encoded by the ergative case, patients of transitive verbs and core arguments of intransitive verbs are encoded by the absolutive case. A number of experiential, or affective, verbs ('see', 'know', 'want', 'love', among some others) encode their subject in the dative instead of the ergative. The genitive case in Rutul is a nominal instantiation of a more general 'attributive' category, with one and the same attributive suffix being used on various types of nominal modifiers including adjectives, demonstrative, participles, etc.

Sentences (1)–(3) from the text corpus of Kina Rutul illustrate the use of the ergative for marking the agents of transitive verbs (cf. *wɨylira* 'husband', *rɨšera* 'girl', *zad* 'I'), the absolutive for marking the patients (cf. *gʷalaχ* 'job' and *ʁil* 'leg'), and the genitive, or 'attributive', form of a noun in the modifying function (cf. *gedijed* 'boy's'). In (2), we can also see the postpositional phrase *urusašdɨ bejda* 'near the Russians', in which the complement takes the genitive form. As for the gender agreement on verbs, in (1) it is with the Gender 3 absolutive *gʷalaχ* 'work, job', in (2) it is with the elided absolutive plural subject 'we' (hence, the 'human plural' gender marker is used), and in (3) it is with the Gender 3 absolutive *ʁil* 'leg'. Note that in (1)–(2), *gʷalaχ waʔas* 'to work' and *ješemiš hɨkɨs* 'to live' represents two of many types of compound verbs, to be discussed in Section 3.

3 By convention, pharyngealization is marked only on vowels of the first syllable, although it usually 'spreads' to the second syllable as well (e.g., [raˤqˤ] 'road', [qʼʷˤaˤd] 'two', [jɨˤχˤɨˤs] 'hit, beat').

(1) *wɨɣl-ir-a=xa* *gʷalaχ* *w-aʔa-r-i-j*, *za-d=xa* *gʷalaχ*
husband-OBL-ERG=ADD job.ABS 3-do.IPFV-CVB-COP-PST I-ERG=ADD job.ABS
w-aʔa-r-i-j.
3-do.IPFV-CVB-COP-PST
{Many years ago we lived in Volgograd.} 'Both my husband worked and I worked.' (Text Corpus)

(2) *ješemiš* *d-iši-r* *urus-aš-dɨ* *bejda*
live HPL-become.PFV-CVB(AOR) Russian-OBL.PL-ATTR beside
'We lived near the Russians.' (Text Corpus)

(3) *mi* *ged-ije-d* *ʁil-i=xʷa* *mi* [...] *riš-e-ra* *ʁil*
this boy-OBL-ATTR leg-OBL(SUP)=REP this girl-OBL-ERG leg.ABS
s-i<w>χɨ-r=xʷa
DOWN-<3>push.PFV-CVB(AOR)=REP
{Once in a bus ...} 'The girl stepped on the boy's foot.' (Text Corpus)

The case inventory includes a large array of locative cases, which express both localization with respect to the landmark (e.g., 'inside' or 'on the upper surface') and direction of motion ('towards' or 'from') or the lack thereof. Thus, the nominal form *ʁili* 'on a leg' in (3) is the form of the case called 'superessive', as it describes location at or motion towards the upper surface. Localization and direction is also expressed on verbs by means of locative prefixes (or 'preverbs'), some of which are cognate to locative case markers. For example, the verb *s-ɨwχɨr* 'pushed down' (4) contains the prefix *s-* expressing downward motion, *l-iwxir* 'put on' (4) has a prefix *l-* 'on', whereas *kʲ-awar* 'add, mix' (4) includes a prefix *kʲ-* with the meaning 'in contact with the surface, in a substance'.

(4) *hejeg* *l-i<w>xi-r*, *ara* *sa* *q'ʷaˤ-b* *dur* *q'el-dɨ*
cauldron.ABS UP-<3>put.PFV-CVB inside one two-3 spoon.ABS salt-ATTR
kʲ-a<w>a-r=a
CONT-<3>do.IPFV-CVB=be
'After we have put the cauldron, we add a couple (lit. one-two) of spoons of salt.' (Text Corpus)

In (4), the numeral *q'ʷaˤ-b* 'two' agrees in gender with the Gender 3 head 'spoon' it modifies; this is not the case with adjectives which lack gender agreement. Note that adjective stems combine with the attributive suffix only in the noun-modifying function, e.g., *hɨχ-dɨ insan* [good-ATTR person] 'a good person'. They take adverbial suffixes when they modify predicates, as in *hɨχ-ana hac'ara* [good-ADV knows] 'knows well'. There is also a third form of adjectives, labeled 'co-verbal' form, which is found in complex predicates which the adjectives is part of, e.g. *hɨχ-a jiʔi* [good-ADJ is] 'is

well, feels well'. Compound verbs with the co-verbal adjectival form will be treated below in Section 3.

As for the verbal system of Kina Rutul, it is fairly rich and includes quite a few synthetic and periphrastic forms (Maisak, 2020b). The overwhelming majority of verb stems agree in gender and include a morphological slot (prefixal or infixal) for gender agreement. The presence and the type of this slot is a lexical property of a verb stem and has nothing to do with the meaning of the verb or any of its morphosyntactic features. Thus, the verb stem *hiši-*/ *w-iši-*/ *r-iši-* etc. [1.become.PFV/ 2-become.PFV/ 3-become.PFV] has a prefixal gender slot, the Gender 1 being zero-marked, the verb stem *ji<r>q'i-*/ *ji<b>q'i-*/ *ji<d>q'i-* etc. [<1/2>become.PFV/ <3>become.PFV/ <4>become.PFV] has an infixal gender slot, and the verb *gič'e-* 'be afraid' lacks an agreement slot altogether. A group of stative verbs comprising the copula *ji?i*/ *i* and a few existential verbs with locative semantics ('be in', 'be under' etc.), have a reduced paradigm and possess only one stem, from which all forms are derived. To the contrary, 'canonical' verbs possess several inflectional stems, on which the synthetic forms are based: these are the perfective, imperfective and infinitive stems, none of which is generally predictable from the other two. The relation between the stems can also be suppletive, e.g. the perfective stem of a very frequent verb 'become' is *hiši-* (Gender 1), whereas its imperfective stem is *ru?u-* and its infinitive stem is *hiki-*.

While the stative verbs have two finite synthetic tenses (present and past), the synthetic forms of canonical verbs are mostly non-finite forms including converbs, participles, infinitive and verbal noun, but also a few non-indicative forms (imperative, optative). The finite indicative system consists of periphrastic forms, which include two components, a lexical verb and a postposed auxiliary, the latter usually being the short copula *i* or the stative verb *a* 'be (inside)'. In the periphrastic forms, lexical verbs occur in one of the two converbs (perfective or imperfective), one of the three participles (perfective, imperfective and prospective) and the infinitive. Thus, the form *wa?ar-ij* in (1) is a periphrastic past habitual consisting of the imperfective converb *wa?ar-* 'doing' and the past tense copula *ij* 'was', and the form *kʲawar=a* in (4) is a periphrastic present tense consisting of the imperfective converb *kʲawar-* 'adding' and the present tense auxiliary *a* 'is (inside)'. The aorist, which is a high-frequent perfective past tense, is special in that the affirmative copula is always dropped, so that finite forms like *liwxir* in (4) are syncretic with perfective converbs: *liwxir* may mean both 'put on' (past) and 'having put on'.

Traditionally, verbal lexemes in Rutul (and the Lezgic languages in general) are divided into simple, or underived, and two groups of derived verbs, namely prefixal verbs containing one or more locative prefixes, and complex verbs consisting of a coverb and a light verb. The number of simple verbs is very restricted: thus, Alisultanov and Sulejmanova (2019, 480) call the number of simple verbs in the Mukhad dialect 'insignificant', while Makhmudova (2001, 163) suggests that the

underived verb stems account for just 7 or 8 percent of the verbal lexicon. For Kina Rutul, 55 simple verbs have been found (e.g., *ɬikɬis* 'become', *haʔas* 'do', *ɬiwɬis* 'give', *jiqʼas* 'come', *jaχas* 'run', *iles* 'eat', *raʿʁas* 'drink', *jetas* 'kick', *wezas* 'milk', etc.). Prefixal verbs are a more numerous group, while the number of verbal compounds as the only open class of verbal lexemes can be estimated in hundreds.[4] One should take into account, however, that the boundaries of the class of verbal compounds are fuzzy, and in the remainder of the paper I will try to touch upon the issue of relation between verbal compounds in a more general sense and complex verbs in a narrower sense of multi-word verbal lexemes.

The present paper suggests the first comprehensive treatment of complex verbs in Rutul. Starting with Dirr (1912, 97), the existing descriptions of the language always mention the existence of compound (or complex, 'periphrastic' etc.) verbs, but usually only present a few examples (as in Alekseyev, 1994, 226 and Alekseyev, 2016, 3538 for the Luchek dialect). Makhmudova's (2001, 167–168) grammar sketch of the Mukhad dialect of Rutul is an exception is that it contains longer lists of complex verbs, although without much comments on their morphosyntactic behaviour.

3 Structural types of light-verb constructions in Kina Rutul

There are quite a few compound verbal lexemes in Kina Rutul which are translational equivalents of simple verbs in English or Russian. Thus, 'to help' is *kumag haʔas* "help do" or *kumag ɬiwɬis* "help give", 'to search' is *aramiš waʔas* "search do", 'to marry (about a man)' is *ɢari raʔas* "wife do", 'to smoke' is *pʼapʼrɬis deʔes* "cigarette pull", and so on. The number of such combinations amounts to hundreds, at the very least. Some of these combinations may turn out to be just periphrastic ways of expressing certain meanings, without displaying any special morphosyntactic

4 The proportion of simple, prefixal and complex verbs is subject to variation even among the languages of the Lezgic group (also, the methodology of counting may differ across the language experts). Thus, in Udi there are about 50–60 simple verbs, and at least 50 prefixal verbs (Maisak, 2008, 98); according to Schulze (2016, 3570), complex verbs account for more than 75% of all verbal lexemes in the language. In Agul, there are at least 120 simple verbs and more that 350 prefixal verbs (Maisak and Ganenkov, 2016, 3585). In Archi, the number of simple verbs is about 170, and the rest of the lexicon consists of complex verbs, as there is no locative prefixation (Chumakina, 2016, 3599). To the contrary, in Tsakhur there seem to be less than ten underived verbs without preverbs (Kibrik and Testelets, 1999, 67).

behaviour, and some of them may turn out to show at least some properties of single words (see further Section 4).

Although various types of verbal compounds exits in many languages, including English (cf. *take a rest* or *look forward*) or Russian (cf. *okazat' pomošč* 'to deliver help' or *soveršit' samoubijstvo* 'to commit suicide'), they are especially prominent in languages like Rutul, where compounds are found among the lexemes expressing very 'basic' concepts. In the present study, I focus on a set of complex verbs that appear in the (unpublished) Kina Rutul dictionary, as well as in elicitation tasks and the available text corpus. Thus, in the Kina Rutul dictionary, collected by the fieldwork team I was part of, almost half (48 per cent) of about 300 verbs included in the dictionary are complex verbs, the rest being simple and prefixal ones.[5]

In this section, I present an overview of compound verbs in Kina Rutul, briefly commenting on their composition and basic properties. The types of verbal compounds surveyed below, ordered by the form of the lexical part, are:

– ADJECTIVE (co-verbal form) + copula
– ADJECTIVE (co-verbal form) + light verb 'become'
– ADJECTIVE (co-verbal form) + light verb 'do'
– NOUN (absolutive case) + light verb 'become'
– NOUN (absolutive case) + light verb 'do'
– NOUN (absolutive case) + other light verbs
– various 'NOUN (oblique case) + verb' combinations
– ACATEGORICAL BOUND COVERB + light verbs 'become', 'do' or 'give'
– VERB STEM + light verb 'do' (= causatives)
– BORROWED VERB STEM + light verbs 'become' and 'do'
– LOCATIVE ADVERB + verb

3.1 Verbal compounds with adjectives

There are three subtypes of verbal compounds which include an adjective as the lexical part. In all three subtypes, adjectives occur in the co-verbal form in *-a*/ *-e* (some adjectives have an unmarked co-verbal form identical to the bare adjective stem).

5 The Kina Rutul dictionary is based on a word list accepted within the international project 'LexCauc – A lexical database for the languages of the Caucasus' (2017–2020), based at the MPI-SHH Jena and led by Diana Forker and Oleg Belyaev. The LexCauc word list includes 1132 entries, of which about 300 represent verbal meanings. The Kina Rutul data were collected by Konstantin Filatov, supervised by Michael Daniel.

3.1.1 Adjective (co-verbal form) + copula

In the first subtype, the light verb is the copula. Verbal compounds of this type describe states. Some of the compound predicates belong to the experiential syntactic class and have a dative subject, e.g. *hɨχa i* 'be good; feel good' (5), while the majority of them belong to the intransitive class, e.g. *hazɨr i* 'be ready; be prepared' (6).

(5) *za-s hɨχ-a jiʔi.*
I-DAT good-ADJ 1.COP
'I'm fine; I feel good.'

(6) *iz-dɨ sin hazɨr i*
I-ATTR all.ABS ready COP
'Everything on my side (lit. my all) is ready.'

3.1.2 Adjective (co-verbal form) + light verb 'become'

In the second subtype, the light verb is *hɨkɨs* 'become'. As the copula has a very reduced paradigm, *hɨkɨs* is used in its place in all the syntactic and tense-aspect contexts that are not available to the copula. Thus, in (7) the aorist form *č'irč'ima višira* expresses perfective past, in (8) the present tense *ɢuru ruʔura* is used with the present habitual meaning, and in (9) the imperative form *jawaš hiš* is employed.

(7) *nek č'irč'im-a w-iši-r=a*
milk.ABS sour-ADJ 3-become.PFV-CVB=be
'The milk turned sour.'

(8) *wirʁ-ɨ-da walɨg-mar k'ɨb ɢuru ruʔu-r=a*
sun-OBL-APUD clothes-PL.ABS quickly dry.ADJ NPL.become.IPFV-CVB=be
'The clothes get dry quickly in the sun.'

(9) *huʁal jawaš hiš, je-s wirɨʁ w-ɨga-r=a*
rain.ABS slow.ADJ 4.become.IMP we-DAT sun.ABS 3-want-CVB=be
'The rain, stop! we need the sun.' (Text Corpus)

Any 'stative' compound with the copula has a 'dynamic' counterpart with the light verb 'become'. Still, compounds with the same lexical part, like *hazɨr i* 'be ready; be prepared' ~*hazɨr hɨkɨs* 'become ready; become prepared', cannot be treated as belonging to one lexeme, given that the copula and the verb 'become' are two distinct lexical items.

3.1.3 Adjective (co-verbal form) + light verb ‘do’

The third subtype of verbal compounds with adjectives comprises the light verb *haʔas* ‘do, make’. Semantically, such compounds are causative equivalents of compounds with adjectives and the copula or the verb ‘become’. There are numerous examples of compound inchoative/ causative pairs like *jawaš hɨkɨs* ‘come to a stop, become quiet’ *~jawaš haʔas* ‘stop (tr.), make quiet’ or *qʼɨxe hɨkɨs* ‘grow, become older’ *~qʼɨxe haʔas* ‘grow (tr.), raise’, etc.

(10) a. *mij-aː haˤ<t>χiˤ-r ha-bɨr sa-ʔ, mašin jawaš*
here-ELAT <HPL>go.PFV-CVB(AOR) that-PL.ABS down-LAT car.ABS stop.ADJ
ha<w>ɨ-r-diš.
<3>do.PFV-CVB-COP.NEG
‘From here, they went further down, they didn’t stop the car.’ (Text Corpus)

b. *ha-now-a (…) qʼɨx-e d-ɨʔɨ-r ha-d*
that-OBL.H-ERG senior-ADJ HPL-do.PFV-CVB(AOR) that-ATTR
χɨnime-r
child-PL.ABS
‘She raised those children.’ (Text Corpus)

c. *čʼabal-er raˤ<t>χaˤ-r-ɨ-j, čʼabal-eš-dɨ dɨn*
sheep-PL.ABS <4>shear.IPFV-CVB-COP-PST sheep-OBL.PL-ATTR wool.ABS
ji<d>ʁɨ-r tamɨz hiʔɨ-r
4.wash.PFV-CVB clean 4.do.PFV-CVB(AOR)
‘They sheared the sheep, washed the wool and made it clean.’ (Text Corpus)

3.2 Verbal compounds with nouns in the absolutive case

3.2.1 Noun (absolutive) + light verb ‘become’

The combinations of a noun in the absolutive case and the verb ‘become’ are not numerous among verbal compounds, as normally such combinations simply represent subject–verb constructions, as in (11a)–(11b), and not complex verb lexemes.

(11) a. *nu haː-sa-d žizni w-iši-r*
PART[R] that-MNR-ATTR life[R].ABS 3-become.PFV-CVB(AOR)
‘Well, life was like that.’ (Text Corpus)

b. *wɨɣɨl-dɨ kulfat w-iši-r-diš=x^{w}a, riši-jmar*
male-ATTR child.ABS 3-become.PFV-CVB-COP.NEG=REP sister-PL.ABS
d-iši-r […]
HPL-become.PFV-CVB(AOR)
'No boys were born, sisters were born …' (Text Corpus)

However, one can subsume under verbal compounds such combinations as *sur hikɨs* and *q'at' hikɨs* 'to break, to be divided', which include the light verb 'become' and the nouns *sur* 'side, half' and *q'at'* 'part, piece'. Both verbs describe the process of division or separation, and both are intransitive verbs taking the absolutive subject, cf. *k'azit* 'paper; letter' in (12a) and *berad ǯibɨr* 'spade handle' in (13a). Note that it is the absolutive subject that controls gender agreement: while in (12a) and (13a) this is not obvious, this is clear from (12b) and (13b) where the nouns *badu* 'trousers' and *rub* 'needle' both belong to gender 3. The nouns *sur* and *q'at'*, as independent nouns, belong to gender 4, hence in (12b) and (13b) with gender 3 agreement on the verb, it is definitely not them that are agreement controllers.

(12) a. *k'azit sur hiši-r=a*
paper.ABS side 4.become.PFV-CVB=be
'The paper tore apart.'
b. *badu sur w-iši-r=a*
trousers.ABS side 3-become.PFV-CVB=be
'The trousers tore apart.'

(13) a. *ber-a-d ǯibɨr q'at' hiši-r*
spade-OBL-ATTR tip.ABS piece 4.become.PFV-CVB(AOR)
'The spade handle broke off.'
b. *rub q'at' w-iši-r*
needle.ABS piece 3-become.PFV-CVB(AOR)
'The needle broke.'

3.2.2 Noun (absolutive) + light verb 'do'

As already mentioned above, combinations of nouns in the absolutive case with the verb 'do' are very numerous in Kina Rutul and often correspond to single-word verbs in languages like Russian or English: thus, 'to work' is *g^{w}alaχ waʔas* ("work do"), 'to moo' is *maʿh waʔas* ("moo do"),[6] 'to dance' is *muk' waʔas* ("dance do"), 'to

6 Note that I subsume ideophones as parts of verbal compounds under nouns, as they do not display a behaviour noticeably different from nouns in their inflection or syntactic distribution.

think' is *fikɨr waʔas* ("thought do"), and so on. As a rule, such combinations describe physical activity, social activity or sound production (rarer, also mental activity). As examples below show, compounds of this type take an agent in the ergative.

(14) a. *zir-ɨ-ra guʒli-na maʿh w-aʔa-r=a*
cow-OBL-ERG strong-ADV IDEOPH 3-do.IPFV-CVB=be
'The cow is mooing loudly.'

b. *iz-dɨ χinχ-ɨ-ra hiχ-ana xed haʔa-r=a*
I-ATTR child-OBL-ERG good-ADV water.ABS 4.do.IPFV-CVB=be
'My son swims well.'

c. *iz-dɨ riši-ra bat'r-ana muk' w-aʔa-r=a*
I-ATTR sister-ERG beautiful-ADV dance.ABS 3-do.IPFV-CVB=be
'My sister dances gracefully.'

d. *wa-d hiji-d fikɨr w-aʔa-r=aʔ*
you-ERG what.OBL-ATTR thought.ABS 3-do.IPFV-CVB=be
'What are you thinking about?'

Probably, the most interesting in this class are those compounds which assign another absolutive argument besides the one that occurs in the 'noun + light verb' combination. Thus, in the following examples *sur* 'side, half', *q'at'* 'part, piece' and *peškeš* 'gift' are nouns in the absolutive case, but there are also other absolutive noun phrases ('trousers', 'bread', 'knife') which are patients of transitive verbs 'tear', 'break' and 'give as a gift', respectively. It is these 'external' absolutive NPs that control gender agreement on the light verb (for example, *badu* 'trousers', a gender 3 noun, controls agreement on *wɨʔɨra* 'did', whereas *sur* 'side' cannot be the controller here, as it belongs to gender 4). See Section 4 for the discussion of whether such nouns can really be treated as absolutive NPs.

(15) a. *za-d badu sur w-ɨʔɨ-r=a*
I-ERG trousers.ABS side 3-do.PFV-CVB=be
'I tore the trousers.'

b. *za-d xɨw q'at' hiʔɨ-r*
I-ERG bread.ABS piece 4.do.PFV-CVB(AOR)
'I broke the bread (in two).'

c. *za-d ha-nowu-s kant' peškeš hiʔɨ-r*
I-ERG that-OBL.H-DAT knife.ABS gift 4.do.PFV-CVB(AOR)
'I gave him the knife as a present.'

3.2.3 Noun (absolutive) + other light verbs

There is a number of more or less idiomatic combinations with a noun in the absolutive and a transitive verb other that 'do', e.g. 'give' or 'pull'. In such combinations, the noun seems to simply fill the position of the patient argument. Note that for some nouns, there exist semantically equivalent combinations with 'do' or a different verb, e.g. *kef deʔes*/ *kef haʔas* 'to enjoy' ("pleasure pull"/ "pleasure do"), *kumag ɬiwis*/ *kumag waʔas* 'to help' ("help give"/ "help do").

(16) a. *ha-now dawat-mɨ-kʲ kef d-eʔe-r=a*
that-OBL.H(ERG) wedding-OBL.PL-CONT pleasure.ABS 4-pull.IPFV-CVB=be
'He is chilling at the wedding.'

b. *iz-dɨ šu-ra bala d-eʔe-r=a p'ap'rɨs*
I-ATTR brother-ERG much 4-pull.IPFV-CVB=be cigarette.ABS
'My brother smokes much.'

c. *je-d wa-s kumag wɨlc'a-r=a*
we-ERG you-DAT help.ABS 3.give.IPFV-CVB=be
'We are helping you!' (Text Corpus)

d. *χɨnχ-ɨ-ra heh raˤʁa-r=a*
child-OBL-ERG IDEOPH 4.drink.IPFV-CVB=be
'The boy is yawning (lit. drinks a yawn).'

e. *tɨl-ije-ra za-s miz jiˤ<w>χɨ-r*
dog-OBL-ERG I-DAT tongue.ABS <3>hit.PFV-CVB(AOR)
'The dog licked me (lit. hit tongue at me).'

3.3 Verbal compounds with nouns in oblique cases

A number of compounds include nouns in oblique cases, which usually refer to locations or instruments. As a rule, such combinations are parts of standard syntactic constructions and neither semantically, not morphosyntactically seem to represent complex verb lexemes. For example, the dative noun *wɨɣlis* as part of *wɨɣlis rurus* 'to marry (about a woman)' is the beneficiary or goal NP 'to the husband' accompanying the motion verb (17a). The ergative noun *šinara* as part of *šinara hɨkɨs* 'to sweat' is the instrumental noun phrase 'with sweat' combined with the verb 'become' (17b). In the combination *neqʲ'a: kiʁipxus* 'to wake up', literally "to fall out from a dream" (17c), *neq'a:* is a locative case (inelative) expressing the source of motion.

(17) a. *wɨ-dɨ riši wɨɣl-ɨ-s r-ɨxɨ-r=a-m*?
you-ATTR sister.ABS husband-OBL-DAT 2-go_to.PFV-CVB=be-Q
'Have your sister married?'

b. *siɣin-e-ra hejwamar šin-a-ra hiši-r=a*
heat-OBL-ERG horse.PL.ABS sweat-OBL-ERG NPL.become.PFV-CVB=be
'Horses sweated because of the heat.'

c. *wi ʁijʁa mis neqʲ-a: ki-ʁ-i<r>xu-r*?
you.ABS today when sleep-IN.ELAT PV-PV-<1>wake_up.PFV-CVB(AOR)
'When did you wake up today?'

However, in some compounds oblique nouns seem to be 'incorporated' elements, forming a closer bond with the verb than ordinary arguments. Thus, in *xije ewč'us* 'to wash oneself' ("to get into water") and *xiji liwxis* 'to drown' ("to appear upon water"), the location of drowning is expressed separately from the noun phrase *xije*/*xiji* 'in/ on water', which hence looks as a syntactically 'incorporated' lexical part.

(18) a. *zi lec'-ur-a xij-e eč'u-r*
I.ABS river-OBL-IN water.OBL-IN 1.bathe.PFV-CVB(AOR)
'I bathed in the river.'

b. *armi-j-a: majit q-i<d>q'i-r*,
army[R]-OBL-IN.ELAT corpse.ABS RE-<HPL>come.PFV-CVB(AOR)
mori-j-a xiji l-i<r>xu-r
sea[R]-OBL-IN water.OBL(SUP) PV-<2>appear.PFV-CVB(AOR)
'They brought the body from the Army, (he) drowned in the sea.' (Text Corpus)

3.4 Verbal compounds with acategorical bound coverbs

An important type of verbal compounds is based on coverbs that do not belong to any major lexical class. The lexical part is morphosyntactically bound in that it does not occur elsewhere besides a compound it is part of, and it is acategorical in that it is neither noun, nor adjective, adverb or verb. Unlike nouns used as coverbs, acategorical coverbs do not control gender agreement. When combined with 'become', they do not fill the subject slot and when combined with 'do', they do not fill the patient slot. Some of the acategorical coverbs (e.g. *χaˁr* 'to learn; learning', *dagul* 'to steal; stealing') can combine with both light verbs 'become' and 'do', deriving inchoative/causative pairs, see (19).

(19) a. *nin-e χinɨχ kitab q'ale haʔa-s χaˁr hɨʔi-r*
mother-ERG child.ABS book.ABS read 4.do-INF learn 1.do.PFV-CVB(AOR)
'The mother taught the child to read books.'

b. *rasul juž žu-s χaʿr hiši-r haʿrf-bir*
Rasul.ABS 1.self.ABS 1.self-DAT learn 1.become.PFV-CVB(AOR) letter-PL.ABS
q'ale haʔas
read NPL.do-INF
'Rasul learned to read letters by himself.'

c. *rasul-a iz-dɨ t'exʲ dagul ha<w>ɨ-r*
Rasul-ERG I-ATTR sheep.ABS disappear <3>do.PFV-CVB(AOR)
'Rasul stole my sheep.'

d. *χɨniχ iškab-a dagul hiši-r*
child.ABS wardrobe-IN disappear 1.become.PFV-CVB(AOR)
'The little boy hid himself in the wardrobe.'

Others acategorical coverbs are restricted to combinations with just one light verb: e.g., *q'ale*/ *q'ɨle* 'to read; reading' only combines with 'do' (20a). An acategorical component *masa* 'to sell; sale' combines with the light verb 'give' (20b).

(20) a. *q'urʔan q'ɨle w-aʔa-r-i-j man, jasin*
Quran.ABS read 3-do.IPFV-CVB-COP-PST PART Yasin.ABS
w-aʔa-r-i-j wɨyle-š-e
3-do.IPFV-CVB-COP-PST husband-OBL.PL-ERG
'The men used to read the Quran, they read Surah Yasin.' (Text Corpus)

b. *za-d č'abɨl-er masa hɨwɨ-r, artɨχ-na*
I-ERG sheep-PL.ABS sell NPL.give.PFV-CVB(AOR) excessive-ADV
s-atɨ-r-diš
PV-NPL.leave.PFV-CVB-COP.NEG
'I sold the sheep, did not keep any extra ones.' (Text Corpus)

3.5 Causatives with the light verb 'do' as a type of verbal compounds

There are no morphological (affixal) causative derivations in Kina Rutul. Causatives are light verb constructions with a special verb form and the light verb *haʔas* 'do'. Only intransitive and experiential verbs can derive this type of causatives, but not transitive verbs with an ergative subject. Still, the group of compound 'do'-causatives is very numerous.

The lexical part of 'do'-causatives is a special 'periphrasis form' of the verb. With most verbs, it is formally identical with the 'full form' of the imperative, i.e. the form with a vocalic suffix. Thus, the imperative 'die!' is *jiq'-e* [1.die-IMP], although a truncated form *jiq'* is also available. In a compound 'do'-causative which means 'to kill', namely *jiq'-e haʔas* [die-IMP do.INF], only the form *jiq'-e*, not *jiq'* can be employed

(21a). The verb *lat'us* 'be finished, come to an end' (21b) is also intransitive, and its causative is a means to express intentional completion ('finish, bring to an end'). The verb *ʁag^{w}as* 'see' (21c) belongs to a class of bivalent experiential verbs with a dative subject; in the possibility of deriving causatives experiential verbs pattern with intransitives.

(21) a. *mi si-ri-ra mi edemi jiq'-e ha?a-r=a*
this bear-OBL-ERG this man.ABS 1.die-PER 1.do.IPFV-CVB=be
'And the bear kills ("makes die") the man.' (Text Corpus)

b. *armi-j-e služba l-a<p>t'-a ha<w>i-r*
army[R]-OBL-IN service[R].ABS PV-<3>end-PER <3>do.PFV-CVB
q-i<r>q'i-r
RE-<1>come.PFV-CVB(AOR)
'He finished ("made end") his service and came back.' (Text Corpus)

c. *xela-la qu-? za-d ha-biši-s ʁ-ag^{w}-a*
then-ELAT back-LAT I-ERG that-OBL.HPL-DAT PV-4.see-PER
q-i?i-r-diš
RE-4.do.PFV-CVB-COP.NEG
'After that, I didn't show ("make see") this to them anymore.' (Text Corpus)

3.6 Verbal compounds with borrowed verb stems

Lexical parts of verbal compounds can be represented by a borrowed Azerbaijani or Russian verb. Both types of borrowed verbs are bound components, which do not occur as independent words outside complex verbs.

This type of verbal compounds is especially important, because it represents the only way of creating new verbal lexemes. Compounds based on Azerbaijani verbs belong to an older layer. Azerbaijani used to be an important contact language in southern Daghestan for centuries, and also functioned as a language of school education in early Soviet times, until Russian was finally introduced in this role in 1950s. Nowadays, the command of Azerbaijani among the Rutul people is low.[7] To the contrary, the Rutul–Russian bilingualism is almost complete, as elsewhere in Daghestan: only very small children or very old people (especially women, who do not often leave their own villages) have little to no command of Russian. Thus, the role of Russian in the extension of the lexicon, be it the borrowing of nouns or

7 See Chechuro et al. (2021) for a detailed account of Azerbaijani bilingualism in southern Daghestan.

verbs (as part of verbal compounds, in the latter case) has become higher in the last decades.

The borrowed Azerbaijani verbs occur in the form in *-miš*, which is originally the perfect participle in Azerbaijani. As a borrowed item, the form in *-miš* is acategorical and simply represents the lexical verb without any inflections. The choice of this particular form as the lexical part of complex verbs is typical both in Nakh-Daghestanian languages and elsewhere, see Bağrıaçık et al. (2015, 110) on verbs borrowed from Oghuz Turkic. The light verbs is either 'become', for intransitives, or 'do', for transitives.

Some of the lexical parts are attested with just one light verb: thus, *aramiš* is only found in the transitive compound *aramiš haʔas* 'to look for, search' (22a), while *ješemiš* only occurs within the intransitive compound *ješemiš hikis* 'to live' (22b).

(22) a. *gad-ije-ra ǯu-s lišanči aramiš r-aʔa-r=a.*
boy-OBL-ERG 1.REFL-DAT bride.ABS search 2-do.IPFV-CVB=be
'The boy is looking for a bride.'

b. *je mij ješemiš du-ruʔu-r=a.*
we.ABS here live HPL-become.IPFV-CVB=be
'We live here.'

To the contrary, the lexical part *jaʁmiš* is 'labile' in the sense that it can co-occur with both intransitive and transitive light verbs, cf. *jaʁmiš hikis* 'to gather together (intr.)' vs. *jaʁmiš haʔas* 'to gather, collect (tr.)'. There are other coverbs with similar behaviour, e.g., *bezmiš* which occurs both in a transitive compound *bezmiš haʔas* 'to annoy, make bored' (23c) and an intransitive one *bezmiš hikis* 'to get bored'; with the latter, the source/ stimulus is encoded by a locative (superelative) case form (23d).

(23) a. *za-d bala uq' jaʁmiš hiʔi-r=a.*
I-ERG much hay.ABS gather 4.do.PFV-CVB=be
'I have collected much hay.'

b. *je gim-a jaʁmiš d-iši-r=a-j.*
we.ABS godekan-IN gather HPL-become.PFV-CVB=be-PST
'We gathered at the godekan (a gathering place for men in a village).'

c. *ha-now zi bezmiš hiʔi-r=a.*
that-OBL.H(ERG) I.ABS annoy 1.do.PFV-CVB=be
'He annoyed me.'

d. *zi ha-now-la bezmiš hiši-r=a.*
I.ABS that-OBL.H(SUP)-ELAT annoy 1.become.PFV-CVB=be
'I got sick and tired of him.'

The borrowed Russian verbs occur in the form of the Russian infinitive (in *-at*/ *-at^{j}*),[8] which is another pattern typical for Nakh-Daghestanian languages and many other languages of Russia (Forker and Grenoble, 2021, 261–263). Like the Azerbaijani *miš*-form, the borrowed Russian infinitives are acategorical and do not bear any other inflections. The light verbs are 'become' for intransitives and 'do' for transitives. The following examples are taken from the spoken text corpus.

(24) a. *ti xu-d-di mu^{ʕ}ɢ-u-la za-d perwij mesto*
that five-4-ATTR village-OBL(SUP)-ELAT I-ERG first[R] place[R].ABS
zanimat^{j} ha<w>i-r, sowχoz-a hixi-r
take[R] <3>do.PFV-CVB(AOR) sovkhoz[R]-IN 1.go_to.PFV-CVB
'Among the five villages, I took the first place (in a competition), when I went to sovkhoz (a collective farm in Soviet times).' (Text Corpus)

b. *xele ma^{ʕ}ʔalim-a-ra ha-bir nakazat^{j} d-aʔa-r-i-j*
after teacher-OBL-ERG that-PL.ABS punish[R] HPL-do.IPFV-CVB-COP-PST
'Then the teacher punished them.' (Text Corpus)

As both (24a) and (24b) show, the lexical parts of complex verbs *zanimat^{j} haʔas* and *nakazat^{j} haʔas* are not assigned any thematic role: both verbs have their own patient NPs, namely *perwij mesto* 'the first place' in (24a) and *habir* 'they' in (24b).

Unlike the compounds with Azerbaijani forms in *-miš*, which are relatively numerous and common in the text corpus, the compounds with Russian infinitives are relatively rare. Interestingly, only compounds with Azerbaijani forms in *-miš* can be found in the dictionary of Standard Rutul based on the Mukhad dialect (Alisultanov and Sulejmanova, 2019), whereas those with Russian infinitives are absent in this source. This may point at the perception of Russian verbs in Rutul compounds as instances of code-mixing and/ or imperfect command of one's native language. Indeed, at least some of such verbs can be regarded as occasional combinations used whenever a Rutul speaker finds appropriate to combine a Russian lexical part with a light verb instead of using a Rutul lexeme. Such compound verbs can be often seen in those contexts where the description of a situation includes a whole range of Russian borrowings, like in (24a): here, *perwij mesto* 'the first place' and *sowχoz* 'sovkhoz, collective farm' are also Russian loans, and the situation described is a workplace competition among the collective farms in late Soviet or post-Soviet period.

8 In Russian, the infinitive suffix *-t'* (-ть) is a palatalized /t^{j}/; in Rutul, however, borrowed Russian infinitives do not always keep the palatalization on the final consonant.

3.7 Verbal compounds with locative adverbs

Compounds that represent combinations of a preposed locative adverb and a verb stand apart from the rest of the group for two reasons. First, only in this type do we see the use of locative adverbs in verbal compounds. The adverbs used in compounds are mostly those with essive (locational) semantics, e.g. *u:* 'on top of', *a:* 'below', *ara* 'inside', or those with lative (directional) semantics, e.g. *la?* 'up', *sa?* 'down', *a?* 'inside', *ʁa?* 'outside', *xu?* 'in front'.

Second, the verbs used with locative adverbs are not the semantically general light verbs like 'do' or 'become' which we see in the other types of compounds. In the "adverb + verb" combinations, the verbs are mainly verbs with locative prefixes which already express a locational meaning by themselves. Thus, in all three compounds *la? luzas* ("up + stand") 'to get up; stand up', *a: lukas* ("below + lie") 'to lie down' and *xu? la?as* ("in front + throw") 'to vomit' the verb starts with *l-*, which is a locative prefix with the meaning 'up, on'. In a sense, in "adverb + verb" compounds, adverbs function as an additional layer of spatial marking specifying the spatial configuration of the event. In some cases, this yields semantically transparent combinations like *la? luzas* 'to get up; stand up' or *sa? sirxus* ("down + get down") 'to fall down', in which the adverb reinforces and specifies the semantics of the verb. Some combinations, like *xu? la?as* ("in front + throw on") 'to vomit' or *ara jiˤχas* ("inside + hit, kick") 'to collide, to intermix', are idiomatic. In any case, it appears that quite often the most natural way to describe a situation is by using an "adverb + verb" combination, and not the verb alone.[9]

The following examples illustrate the use of "adverb + verb" compounds *a: lukas* 'to lie down', *la? luzas* 'to get up; stand up' and *ara jiˤχas* 'to collide, to intermix'.

(25) a. *ha<w>i-r=xʷa* *kašir*, *mij-a:*
<3>do.PFV-CVB(AOR)=REP porridge.ABS here-ELAT
ha<w>i-r=xʷa, *mi-bir* *a:* *l-ü<t>kü-r*
<3>do.PFV-CVB(AOR)=REP this-PL.ABS below PV-<HPL>lie.PFV-CVB
s-e<t>χi-r=xʷa
PV-<HPL>sleep.PFV-CVB(AOR)=REP
'They made the porridge, and after they made it, they lied down to sleep.'
(Text Corpus)

9 Nasledskova and Netkačev (2020, 812) mention that in the combination *sa? ?irxus* 'to fall down' (the verb *?irxus* 'to get into, to appear' includes the locative prefix *?-* 'inside'), the use of the adverb *sa?* 'down' is optional in some cases, but is obligatory for the description of those situations when the starting point of falling is mentioned explicitly and it is higher than the final point.

b. *haj-a:* *dibir* *la-ʔ* *l-uzu-r* *ruxʷa-r=a=xʷa*
DEM-ELAT mullah.ABS up-LAT PV-1.stand.PFV-CVB 4.say.IPFV-CVB=be=REP
'Then the mullah stood up and said …' (Text Corpus)

c. *qʼʷaˤ-d čabɨl-eš-dɨ* *süri* *ara* *jiˤ<l>χi-r=a-j*
two-4 sheep-OBL.PL-ATTR flock.ABS inside <APL>hit.PFV-CVB=be-PST
'The two flocks of sheep mixed up.'

Note that combinations of one and the same adverb with two different verbs are not uncommon. Apart from the pair *saʔ sirxus* ("down + get down") ~*saʔ ʔirxus* ("down + get into") 'to fall down', cf. the pair with a difference in meaning in (26a)–(26b). Whereas *u: liʔir* ("on top + put on") 'to put on' is used to describe putting on of clothes (26a), its counterpart with a different verb bearing the same locative prefix, namely *u: luʁur* ("on top + lift on") 'to place onto' describes lifting something on an upper landmark, as in (26b).

(26) a. *za-d* *ʁijʁa* *bɨtʼrɨ-d* *uχun* *u:* *l-a<w>a-s-ɨ*
I-ERG today beautiful-ATTR dress.ABS on_top PV-<3>put_on-INF-COP
'I shall put on a beautiful dress today.'

b. *hejwan-a* *χinɨχ* *u:* *l-aʁʷ*
horse-OBL(SUP) child.ABS on_top PV-1.lift.IMP
'Mount the child on the horse!'

4 Properties of light-verb constructions in Kina Rutul

As I mentioned in the Introduction, it may be not so easy to draw a boundary between ordinary syntactic combinations (especially, 'patient NP + transitive verb') and complex verbs in the sense of compound (bi-componental) verbal lexemes, whose behaviour is different from ordinary syntactic combinations and which should be lexically listed. Thus, among the numerous combinations like "work do", "big become", "die do", "live become" or "down fall" there may be both just periphrastic ways of expressing certain verbal meanings, without special morphosyntactic behaviour, or compound lexical items, displaying at least some properties of single words.

In order to scrutinize this distinction, quite a few parameters of wordhood may potentially be taken into account, although not all of them may be applicable to every language. For example, Harris (2002, 76–87) in her study of complex verbs in Udi, argues that complex verbs in this language are single verbs with an 'incorporated' lexical part, by addressing the following criteria: complex verbs have the stress of a

single verb; complex verbs are written as single words in texts and dictionaries; all of the deverbal derivational processes (e.g., derivation of various non-finite forms and causatives) take complex verbs as input; the negative marker occurs before the whole complex verb; complex verbs do not incorporate phrases with modifiers; conjoining parts of complex verbs is impossible; gapping parts of complex verbs is impossible; parts of complex verbs are anaphoric islands (as they cannot be replaced with proforms). Although for Kina Rutul, a dedicated in-depth study involving the application of these (and, potentially, also other) criteria to all the groups of verbal compounds remains a task for the future, in the present section I will partially approach the problem of wordhood of Kina Rutul compounds, focusing on the syntactic and morphological fusion of their parts. In particular, I will look at whether the order of the components is fixed (by default, the lexical part is followed by the light verb) and whether the insertion of other lexical material between the two parts is allowed, whether the 'nominal' lexical component may be assigned a thematic role (e.g. the patient) or be modified, and whether lexical parts of compounds or light verbs can become morphologically bound and 'incorporated', turning former bi-componental structures into simple verb stems.

4.1 Word order

Describing complex verbs in Archi, also a Lezgic language, Chumakina (2016, 3600) states that "[s]yntactically, all types of complex verbs demonstrate the characteristics of a single word: the order of the parts is fixed (the lexical part is followed by the light verb) and the insertion of other lexical material between these parts is not, as a rule, allowed". Kerimova (2023) make a similar conclusion about the complex verbs in Lezgian, although showing at the same time that in Agul, a closely related language, it is possible both to change the order of components and to insert other material between them. Kina Rutul displays the same behaviour of complex verbs as described for Agul (see also Maisak and Ganenkov, 2016, 3580–3583): although, for all type of complex verbs, the default order is 'lexical part + light verb', the reverse order is judged acceptable by native speakers. Likewise, the insertion of other material between the parts did not make the respective examples ungrammatical:

(27) a. *za-d q'ile w-i?i-r=a-j maskow-di*
I-ERG read 3-do.PERF-CVB=be-PST Moscow-OBL(SUP)
'I studied in Moscow.'

b. *za-d maskow-di w-i?i-r=a-j q'ile*
I-I-ERG Moscow-OBL(SUP) 3-do.PERF-CVB=be-PST read

c. *za-d q'ile maskow-di w-iʔi-r=a-j*
I-ERG read Moscow-OBL(SUP) 3-do.PERF-CVB=be-PST

4.2 Thematic roles of coverbs

In the 'noun (absolutive) + light verb' combinations like *gʷalaχ waʔas* ("work do") 'to work' or *kumag hɬwɨs* ("help give") 'to help', the light verbs are transitive verbs that also function as independent lexical items. In their turn, the coverbs like *gʷalaχ* 'work, job' or *kumag* 'help' also function as independent nouns. As parts of the combinations mentioned above, these nouns seem to occupy the position of patient noun phrases (in the absolutive case), and they also control gender agreement on the verb. In sentences like (1), (14a)–(14d) or (16a)–(16e), there are no other nouns that might be alternative candidates for the patient role.

The situation is different for compound verbs like "side do", "piece do", "present do" mentioned in Section 3.2: although their coverbs *sur* 'side, half', *q'at'* 'part, piece' and *peškeš* 'gift' are also 'normal' nouns, it is other absolutive noun phrases that occupy the patient slot in (15a)–(15c). The nominal parts of the compounds *sur haʔas* 'to break, to tear (tr.)' and the like thus seem to represent syntactically 'incorporated' components, i.e. indispensable parts of complex verbs rather than free absolutive NPs. But let us have a look at the intransitive combinations with the same coverbs, e.g., *sur hɬkɨs* and *q'at' hɬkɨs* 'to break, to tear (intr.)', with an absolutive subject distinct from the nouns *sur* or *q'at'* (12a)–(13a). One may wonder, whether the lexical parts of such verbs actually fill the Theme slot of the intransitive verb 'become': "[Subject$_{ABS}$] becomes [Theme$_{ABS}$]", a construction that can be illustrated with (28). Then, in (12a) it is the trousers that become a 'half', and in (13a), it is the spade handle that becomes a 'piece'. Under such approach, the construction would be morphosyntactically canonical, like (28), and does not represent a complex verb as such.

(28) *riši=xa šu=xa duχtur-ar d-iši-r=a*
sister.ABS=ADD brother.ABS=ADD doctor-PL.ABS HPL-become.PERF-CVB=be
'Both the sister and the brother became doctors.'

The same logic can be applied to combinations with the verb 'do': constructions like *sur haʔas* and *q'at' haʔas* (15a)–(15c) can be seen as merely causative counterparts of "[Subject$_{ABS}$] becomes [Theme$_{ABS}$]" clauses. That is, 'I broke the bread (in two)' can be possibly represented as an instance of "[Agent$_{ERG}$] makes [Patient$_{ABS}$] a [Theme$_{ABS}$]", namely "I (Agent) made the bread (Patient) a piece (Theme)". This, again, gives us an opportunity to treat 'noun + verb' compounds as regular syntactic combinations and not as complex verbs in the narrow sense of bi-componental lexemes.

4.3 Modification of coverbs

We might expect that for those light verb constructions whose lexical parts are already 'incorporated' these lexical parts cannot be modified or bear any inflections. Although this it true for acategorical bound components, this is not so for 'noun + light verb' combinations. We find examples, where nouns in such compounds can be inflected for number (i.e. bear the plural suffix), as in (29a) and (29b), or be modified. Thus, in (29b) the noun 'work, job', as part of the "work + do" combination, is preceded by an adjective meaning 'hard', which modifies the noun. In such instances, the lexical component can again only be analyzed as the absolutive patient noun phrase and not as an incorporated part of a complex lexical item.[10]

(29) a. *muˁgʷ-a-d q'ix-dɨ q'aˁs-dɨ insan-aši-s za-d*
village-OBL-ATTR big-ATTR old-ATTR person-OBL.PL-DAT I-ERG
hɨxɨ-r kumag-bɨr haʔa-r-i-j
1.go_to.PFV-CVB help-PL NPL.do.IPFV-CVB-COP-PST
'I used to go and help the old people of the village, the elders.' (Text Corpus)

b. *gal-a d-ɨxɨ-r gʷalaχ-bɨr haʔa-r=a* […]
Gal-IN HPL-go.PFV-CVB job-PL.ABS NPL.do.IPFV-CVB=be
'They go to Gal (a valley in Azerbaijan) and do various kinds of work …' (Text Corpus)

c. *haj-a: qu-ʔ za-da: gʷalaχ* […] *juˁq'-dɨ gʷalaχ*
there-IN.ELAT back-LAT I-APUD.ELAT job.ABS hard-ATTR job.ABS
qa-w-aʔa-s ruʔu-r-diš
RE-3-do-INF 4.become.IPFV-CVB-COP.NEG
'After that I cannot do any hard work.' (Text Corpus)

4.4 Morphologically bound coverbs?

Discussing complex verbs in Agul, another Lezgic language, Maisak and Ganenkov (2016, 3582) claim that the class of complex verbs in Agul is not uniform in that "some of them are close to free syntactic combinations of verbs and object noun phrases, while others are lexicalized to a considerable degree and approach simplex verb stems". In particular, as regards the latter group, lexical parts of some compound

10 In (29c), the light verb 'do' bears a repetitive prefix *q-* 'again' (here, in the context of negation, it adds the meaning 'not anymore'). However, this prefix is so regular that we can probably speak here about a 'repetitive form' of the same verb 'do' and not of a different light verb. See also Section 4.4 on repetitive prefixation.

verbs can become so tightly fused, that it is not the light verb, but the whole 'coverb + light verb' complex that can serve as an input for derivational processes. For example, prefixes in Agul, including the productive repetitive prefix (meaning 'again' or 'backwards'), only attach to verb stems and, in case of compound verbs, to light verbs, as in (30a). In just a few compound verbs, however, it is possible to prefix the repetitive marker to the lexical part, as in (30b).[11] In both examples, the complex verbs *gunt' xas* 'to gather (intr.)' and *gunt' aq'as* 'to gather (tr.)' include the lexical part *gunt'*, which is a bound element, probably related to the noun *k'unt'* 'heap'. The light verbs are 'become' and 'do', which correspond to independent verbs of the language. The behaviour of *gunt'* in (30b) shows, that, as part of a complex verb, it became not only syntactically, but also morphologically fused with the light verb: former bi-componental combinations *gunt' aq'*- "gather do" and *gunt' x*- "gather become" have been reanalyzed (even if optionally) as single stems *gunt'aq'*-, *gunt'x*-, to which a verbal prefix can be added.

(30) a. *saje-wur-i-f-as gunt' q-aq'a-je ħupː-ar e.*
other-PL-OBL-APUD-ELAT gather RE-do.IPFV-PTCP:PRS sheep-PL.ABS COP
'These were the sheep which they used to collect from the others.'

b. *aχpːaj bagajmi qa-gunt'-xa-s-e wari.*
then in_the_morning RE-gather-become.IPFV-INF-COP all.ABS
'Then in the morning, everyone is getting together again.'

For Kina Rutul, we are not aware of similar cases of complex verbs becoming simplex verb stems. (The only exception might be the verb 'to sneeze', which will be treated in the next section.) All complex verbs known to us still represent two components, without morphological fusion. Even bound lexical parts like the Azerbaijani and Russian roots do not become reanalyzed as simply parts of verb stems. An obvious exception, of course, are periphrasis forms of Rutul verbs as parts of causative compounds (Section 3.5), which are naturally verbal and retain the inflectional potential (like gender agreement) they originally possessed as verb stems.

4.5 Morphologically bound or 'zero' light verbs?

The light verbs occurring in verbal compounds can potentially include morphologically bound elements that are not used elsewhere as autonomous verbal predicates. This situation is observed in Udi, where, apart from the verbs 'do', 'become' and

11 Examples (30a) and (30b) are taken from the unpublished spoken corpus of Huppuq' Agul, collected by Dmitry Ganenkov, Timur Maisak and Solmaz Merdanova in the 2000s.

'say', which have counterparts among independent lexical items, a handful of other light verbs can be identified. The light verbs of this latter group are not found as free verbs in the modern language, and only tentative etymologies can be suggested for them. For example, the light verb *-ec-* found in intransitive/ decausative compounds, belongs to the same inflectional type as motion verbs, and might be related to a motion verb 'go, come' historically, and the light verb *-d-* found in transitive/ causative compounds, is very probably related to the root 'give', which in modern Udi is only attested in a prefixed stem *tad-* 'give'; see Maisak (2008, 99–102) and Schulze (2016, 3569) for discussion.

In Kina Rutul, the verb 'to sneeze' may at first glance look as an instance of a complex verb with a bound light verb *urχas* combined with an ideophonic lexical part *čiʔ*. The putative light verb *urχas*, which possess an infixal gender agreement slot (31), does not occur elsewhere in the language. And for the Lezgic languages in general, it is not uncommon to express the meaning 'to sneeze' by means of a compound predicate with a phonetically similar ideophonic part associated with the sound of sneezing, see e.g. *ʕamči aq'as* "sneezing do" in Agul, *inči ap'uz* "sneezing do" in Tabasaran or *aʿnšaʿ bos* "sneezing say" in Archi. However, as it turns out, nothing can intervene between the (putative) components of the Kina Rutul verb, nor their linear order can be shifted. Given that both components are restricted to just this combination, one has to conclude that *čiʔurχas* in fact represents a simplex verb and not a compound.[12]

(31) *χɨnɨχ čiʔurχa-r=a/ rɨš čiʔu<r>urχa-r=a*
child.ABS 1.sneeze.IPFV-CVB=be girl.ABS <2>sneeze.IPFV-CVB=be
'The boy is sneezing.'/ 'The girl is sneezing.'

In some languages possessing complex verbs, it is even possible that the light verb 'disappears' as a result of phonological reduction, so that former compounds cannot be formally identified as bipartite at all. Thus, in Lezgian complex verbs including the light verb *awun* 'do' (e.g., *k'ʷalaχ awun* 'to work', lit. "work do") can occur in their full or reduced forms. In their full forms, the verb is present, cf. the perfective past tense *k'ʷalaχ awuna* 'worked'. In the reduced form, the root 'do' is not visible, cf. simply *k'ʷalaχ-na*, with the tense-aspect suffix *-na* added directly to the coverb (Haspelmath, 1993, 178). Similarly, the Udi light verb *b-* 'do' undergoes devoicing and can even drop if occurs in the middle of a consonant cluster, cf. *äš-e-b-sa* [work-3SG-do-PRS] 's/he

12 Interestingly, in Mukhad Rutul, which is a dialect relatively close to Kina Rutul, the verb 'to sneeze' is much shorter and does not even superficially look like a compound, cf. *čirχas* (Gender 1), *čirirχas* (Gender 2), *čibχas* (Gender 3), etc. (Alisultanov and Sulejmanova, 2019, 429).

works', but *äš-pː-sa* [work-do-PRS] or even simply *äš-sa* in case the personal marker does not occupy the position between the lexical part and the light verb.

In Kina Rutul, no similar instances of light verb reduction or loss have been found: both the coverb and the light verb are always 'full' and prosodically autonomous words. In the available dictionaries of the Mukhad dialect (Alisultanov and Sulejmanova, 2019) and the Ikhrek dialect (Dzhamalov and Semedov, 2006), both employing the Cyrillic script, parts of verbal compounds are written separately, as two independent words.

5 Conclusion

As I have tried to demonstrate in the present contribution, Kina Rutul, just like its closest relatives in the Nakh-Daghestanian family, possesses a rich number of verbal compounds, or light verb constructions, which represent combinations of light verbs, mostly 'do' and 'become', with adjectives, nouns (in the absolutive case, but also some other cases), locative adverbs, verb stems (in the causative construction only) and coverbs with the acategorical status. Among the light verbs, there are no verbs whose use is restricted to compounds, all the verbs can be also used as independent lexical items.

It is highly probable that the majority of verbal compounds, namely all combinations including nouns, adjectives and adverbs, can be analyzed simply as 'noun + verb', 'adjective + verb' and 'adverb + verb' combinations, respectively. They represent a periphrastic way of expressing certain verbal meanings, which in some languages can be expressed by single words. Thus, such compounds are only 'complex verbs' from the semantic point of view. It cannot be excluded, however, that a more fine-tuned analysis of the morphosyntactic behaviour of such combinations can reveal certain restrictions that would allow us to qualify verbal compounds of the respective types as 'complex verbs' in a more narrow sense, namely bi-componental single words. Likewise, for the causative 'verb stem + do' combinations, additional evidence is needed to see whether such combinations represent monoclausal or biclausal structures, i.e. whether they are indeed single words or syntactic constructions.

Still, there are genuine complex verbs in Kina Rutul, whose coverbs cannot be assigned a thematic role and are syntactically 'incorporated' elements of verb lexemes, only responsible for the expression of the lexical meaning. Such complex verbs include coverbs that are neither visibly nominal nor verbal, and thus should be best treated as acategorical. Diachronically, these converbs can represents former nouns or verbs which only survived in compounds (cf. the converbs in *χaʕr hɨkɨs*

'to learn', *q'ɨle haʔas* 'to read; study' or *masa hɨwɨs* 'to sell'), but also can be direct borrowings, especially of the Azerbaijani and Russian verbs (cf. *ješemiš hɨkɨs* 'to live' or *nakazatʲ haʔas* 'to punish'). It is the latter group of complex verbs with borrowed coverbs that can still be supplemented by new members. Whether other types of verbal compounds in Kina Rutul, or at least some representatives of these types, will be proved to represent complex verbs as well, compounds with acategorical coverbs will probably still remain at the core of the complex verb class.

Bibliography

Alekseyev, Mikhail E. 1994. Rutul. In Rieks Smeets (ed.), *The Indigenous languages of the Caucasus. Vol. 3. North East Caucasian languages. Part 2*, 213–258. Delmar: Caravam.

Alekseyev, Mikhail E. 2016. Rutul. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen and Franz Rainer (eds.), *Word-Formation: An international handbook of the languages of Europe. Vol. 5*, 3536–3545. Berlin: De Gruyter. https://doi.org/10.1515/9783110424942-027.

Alisultanov, A. S. and T. A. Sulejmanova. 2019. *Rutul´sko-russkij slovar´ [Rutul–Russian dictionary]*. Makhachkala: Alef.

Authier, Gilles. 2024. Ideophonic verb compounds in Archi. *Linguistic Typology at the Crossroads* 4(1). 325–380. https://doi.org/10.6092/issn.2785-0943/16518.

Bağrıaçık, Metin, Angela Ralli and Dimitra Melissaropoulou. 2015. Borrowing verbs from Oghuz Turkic: two linguistic areas. In Francesco Gardani, Peter Arkadiev and Nino Amiridze (eds.), *Borrowed morphology*, 109–136. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781614513209.109.

Chechuro, Ilia, Michael Daniel and Samira Verhees. 2021. Small-scale multilingualism through the prism of lexical borrowing. *International Journal of Bilingualism* 25(4). 1019–1039. https://doi.org/10.1177/13670069211023141.

Chumakina, Marina. 2016. Archi. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen and Franz Rainer (eds.), *Word-Formation: An international handbook of the languages of Europe. Vol. 5*, 3595–3604. Berlin: De Gruyter. https://doi.org/10.1515/9783110424942-031.

Daniel, Michael, Nina Dobrushina and Timur Maisak. to appear. *Studies in Kina Rutul*.

Dirr, Adolf M. 1912. Rutul'skij jazyk [The Rutul language]. *Sbornik materialov dlja opisanija mestnostej i plemën Kavkaza* 42(3). 1–204.

Dzhamalov, Kamal È. and Semed A. Semedov. 2006. *Rutul´sko-russkij slovar´ (ixrekskij dialekt) [Rutul-Russian dictionary (Ikhrek dialect)]*. Moscow: ÈKON-Inform.

Forker, Diana and Lenore A. Grenoble. 2021. Some structural similarities in the outcomes of language contact with Russian. In Diana Forker and Lenore A. Grenoble (eds.), *Language contact in the territory of the former Soviet Union*, 259–287. Amsterdam: John Benjamins. https://doi.org/10.1075/impact.50.09for.

Ganenkov, Dmitry and Timur Maisak. 2020. Nakh-Dagestanian languages. In Maria Polinsky (ed.), *The Oxford handbook of the languages of the Caucasus*, 87–145. Oxford: Oxford University Press.

Harris, Alice C. 2002. *Endoclitics and the origins of Udi morphosyntax*. Oxford: Oxford University Press.

Haspelmath, Martin. 1993. *A grammar of Lezgian*. Berlin: De Gruyter Mouton.

Ibragimov, Garun Kh. 1978. *Rutul´skij jazyk [The Rutul language]*. Moscow: Nauka.

Ibragimov, Garun Kh. 2004. *Rutul´skij jazyk. Sinxronija i diaxronija [The Rutul language: synchrony and diachrony]*. Makhachkala: DNC RAN.

Kerimova, Zarina R. 2023. *Aspekty grammatičeskoj struktury složnogo glagola lezginskogo i agul´skogo jazykov [Aspects of the grammatical structure of complex verbs in Lezgian and Agul]*. Master Thesis: RGGU, Moscow.

Kibrik, Aleksandr E. and Yakov G. Testelets. 1999. *Èlementy caxurskogo jazyka v tipologičeskom osveščenii [Elements of Tsakhur in a typological perspective]*. Moscow: Nasledie.

Klimov, Georgij A. and Mikhail E. Alekseev. 1980. *Tipologija kavkazskix jazykov [Typology of Caucasian languages]*. Moscow: Nauka.

Maisak, Timur. 2008. Glagol´naja paradigma udinskogo jazyka [The verbal paradigm of Udi]. In M. E. Alekseyev, T. A. Maisak, D. S. Ganenkov and Yu. A. Lander (eds.), *Udinskij sbornik. Grammatika. Leksika. Istorija jazyka*, 96–161. Moscow: Academia.

Maisak, Timur. 2020a. Grammaticalization in Lezgic (East Caucasian). In Walter Bisang and Andrej Malchukov (eds.), *Grammaticalization Scenarios: Cross-linguistic Variation and Universal Tendencies. Vol. 1: Grammaticalization Scenarios from Europe and Asia*, 309–360. Berlin: de Gruyter. https://doi.org/10.1515/9783110563146-006.

Maisak, Timur. 2020b. The verb paradigm in Kina Rutul. HSE Working papers. National Research University Higher School of Economics, Moscow. https://wp.hse.ru/data/2020/10/27/1359252160/95LNG2020.pdf, last access 27/9/2024.

Maisak, Timur and Dmitry Ganenkov. 2016. Aghul. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen and Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe, Vol. 5*, 3579–3594. Berlin: de Gruyter. https://doi.org/10.1515/9783110424942-030.

Makhmudova, Svetlana M. 2001. *Morfologija rutul´skogo jazyka [Morphology of Rutul]*. Moscow: Sovetskij pisatel'.

Müller, Peter O., Ingeborg Ohnheiser, Susan Olsen and Franz Rainer. 2016. *Word-Formation: An International Handbook of the Languages of Europe, Vol. 5*. Berlin: de Gruyter. https://doi.org/10.1515/9783110424942.

Nasledskova, Polina L. and Ivan V. Netkačev. 2020. K opisaniju semantičeskogo polja padenija v rutul'skom jazyke [Towards a description of the semantic field of falling in Rutul]. *Acta Linguistica Petropolitana* 16(1). 801–820.

Schulze, Wolfgang. 2016. Udi. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen and Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe, Vol. 5*, 3564–3578. Berlin: de Gruyter. https://doi.org/10.1515/9783110424942-029.

Anna Riccio

Italian light verb constructions and definiteness

1 Introduction

This paper aims to investigate the morpho-syntactic flexibility and semantic properties inherent in Italian light verb constructions (LVCs), with a particular emphasis on the variation of determiners. It focuses on the use of definite and indefinite articles in conjunction with nouns representing events within LVCs. Illustrative examples include *fare una/ la doccia* 'to take a/ the shower' and *dare una/ l(a) opinione* 'to give an/ the opinion'. Special attention is given to LVCs featuring the verbs *dare* 'to give' and *fare* 'to make/ do'.

The analysis is conducted through the syntax-semantics framework developed by Role and Reference Grammar (RRG) (Van Valin and LaPolla, 1997, 317–440; Van Valin, 2005, 128–293), which provides a valuable theoretical perspective for describing the significant syntactic and semantic features crucial for understanding the LVCs in question and the event structure that the light verb (LV) and its noun phrase (NP) contribute to. The data for this study sourced from the itTenTen20 corpus, a web-based Italian dataset from 2020, accessible via the Sketch Engine corpus manager (Kilgarriff et al., 2004).

The paper is organized as follows. Section 2, along with its subsections 2.1 and 2.2, offers an overview of the semantic and syntactic properties of LVCs under investigation. Section 3 delves into the concepts of (in-)definiteness and the referentiality associated with Italian definite and indefinite articles. Section 4 presents the analysis of the data. Section 5 provides a succinct overview of Role and Reference Grammar (RRG). Section 6 addresses the syntax-to-semantics linking within both non-LVCs and LVCs according to the RRG framework, specifically focusing on the mapping between the syntactic roles in a sentence and the semantic roles of the participants in the event. Section 7 concludes the paper.

Anna Riccio, University of Foggia, Foggia, Italy

https://doi.org/10.1515/9783112229330-007

2 Light verb constructions

The investigated Italian verb-noun constructions are considered a semantically compositional subtype of idioms, structured syntactically through the application of regular syntactic rules in the language (cf. Fellbaum, 1993, 271–295; Nunberg et al., 1994, 491–538; McGinnis, 2002, 665–672; Svenonius, 2005, 227–263; Everaert, 2010, 76–98; Stone, 2013, 1–22; Wittenberg, 2016; Fleischhauer, 2021b, 31–62, and others). Thus, they can be positioned between the constructions represented in (1):

(1) a. *Il medico diede una penna a Tomàs.*
the doctor give.3SG.PST INDEF.F.SG pen to Tomàs
'The doctor gave a pen to Tomàs.'
b. *tirare le cuoia*
pull.INF DEF.F.PL leathers
'kick the bucket'

In (1a), we observe a "free", productive combination characterized by a lack of preconstruction, wherein the structure remains a semantically literal structure and the words retain their conventional meanings. In contrast, the construction in (1b) represents a preconstructed idiom that lacks semantic literalness. The connection between the NP *cuoia* 'leathers' and the semantics of the expression is not immediately apparent. The NP deviates from its typical reference within the discourse's object universe and cannot be decomposed into components aligning with the lexical components of the expression. Consequently, such constructions are frequently perceived as both semantically and syntactically monolithic, stored as unified entities within the lexicon and resistant to numerous syntactic operations, as any modification could potentially induce a shift in meaning (Gibbs and Nayak, 1989, 133; Nunberg et al., 1994, 509; Horn, 2003, 249).[1] Consider instead the following examples:

(2) *Il Presidente ha dato una risposta ai presenti in aula*
the president give.3SG.PRS.PERF INDEF.F.SG answer to presents in chamber
sulla situazione in corso.
on.the situation ongoing
'The President has given an answer to those present in the chamber on the ongoing situation.' (marche.it)

1 A good example of syntactic operations is passivization. Considering the idiomatic phrase *tirare le cuoia* 'kick the bucket', **le cuoia sono state tirate ieri da Gianni* 'the bucket was kicked yesterday by John' cannot mean that John died yesterday; the idiomatic phrase cannot be passivized.

(3) *Il bagno minuscolo si allagava dopo aver fatto la*
the bathroom tiny flood.3SG.PST after have.INF make.PST.PTCP DEF.F.SG
doccia (*non c'era nessun divisorio tra i sanitari e la*
shower no there=be.IMPF no partition between the fixtures and the
doccia).
shower
'The tiny bathroom would flood after taking a shower (there was no partition between the fixtures and the shower).' (expedia.it)

It is evident that both constructions *ha dato una risposta* in (2) and *aver fatto la doccia* in (3) do not fall into the category of free combinations and are clearly distinct from idioms. The verbs *dare* 'to give' and *fare* 'to make/ do' are not used in their usual meaning, and they lack any idiomatic connotation. Their formation is idiosyncratic and constructional, meaning that the structure for specific patterns and the specific way they resolve their partial composition are stored in the lexicon. Let's reconsider example (1a). The verb *dare* 'to give' in this context means 'to hand over', involving the specific transfer of a tangible item (such as the pen) from one person (like the doctor) to another (like Tomàs). This specific transfer is absent in example (2). The verb *dare* 'to give' only denotes a general sense of transfer that involves a metaphorical meaning, while the NP *risposta* 'answer' encapsulates the action itself.

The examples provided in (2) and (3) represent a common verb-noun construction in Italian, referred to as 'light verb constructions' (LVCs). These LVCs typically comprises a light verb (LV) with a reduced semantic weight, in conjunction with a NP that carries the principal substantive meaning of the construction, often referred to as a deverbal noun (cf. von Polenz, 1963; Harris, 1968; Gross, 1976, 97–119; Gross, 1981, 7–52; Gross, 1984, 123–156; Gross, 1993, 16–23; Gross, 2010, 16–35; Elia et al., 1985, 311–325; Giry-Schneider, 1987, 27–28; Salvi, 1988, 79; Fellbaum, 1993; Nunberg et al., 1994; Elia and D'Agostino, 1998, 317; Cicalese, 1999, 450; Alba-Salas, 2004, 283 McGinnis, 2002, 669; Mastrofini, 2004, 383; Svenonius, 2005, 227–263; Everaert, 2010, 76–98; Huddleston and Pullum, 2010, 290–292; Stone, 2013, 1–22; Plante, 2014, 82; Wittenberg, 2016, 9–11; Fleischhauer, 2021b, 31–32; Fleischhauer, 2021a, 63, Pompei and Piunno, 2023, 99–100 and others). LVCs typically involve lexical restrictions triggered by the noun (e.g., *prendere*/ **fare una decisione* 'to take/ *to make a decision') that vary across languages, as seen in constructions such as *fare una doccia* (lit. 'to make a shower') in Italian and its English translation *take a shower*.

The notion of 'light verb' originates from Jespersen's (1942) analysis of Modern English grammar, referring specifically to deverbal nouns paired with an "insignificant verb, to which the marks of person and tense are attached, before the really important idea" (Jespersen, 1942, 117). Various terms have been used in the literature to describe these constructions. In French and Italian, they are also known as 'con-

structions à verbe support' or 'costruzioni a verbo supporto' (La Fauci, 1979, 39–40; Gross, 1981, 37; Elia et al., 1985, 311–312; Giry-Schneider, 1987, 2–3; Salvi, 1988, 79; De Angelis, 1989, 223; Stichauer, 2000, 37–38; Mastrofini, 2004, 388-389; Ježek, 2004, 189; Pompei, 2017, 109–110), emphasizing the syntactic perspective where the verb serves as a function word or operator facilitating the predication of another lexical item. In English linguistics, this concept is occasionally discussed (Danlos, 1992, 2; Dras, 1995, 451; Igor, 1996, 38; Krenn, 2000, 74; Fillmore et al., 2003, 243; Langer, 2004, 171). While 'light verb constructions' and 'support verb constructions' are widely accepted in Italian, English, and French literature, German literature presents a broader range of labels for similar constructions. These include 'Verbaufspaltungen' ('verbal dispartments', Lüger, 1995, 118), 'nominale Umschreibungen' ('nominal paraphrases', Keller and Mulagk, 1995, 577), and notably, 'Funktionsverbgefüge' ('function verb constructions', von Polenz, 1963, 11).

2.1 Semantic and syntactic properties of LVs

In LVCs, the verb typically carries a neutral and generic meaning because the semantic weight shifts from the verb to the noun phrase. This makes the verb "semantically light", as its contribution to the overall meaning of the NP is limited. Specifically, the LV primarily provides (i) morphological features such as tense, number, aspect, and diathesis (which determines the syntactic roles of the nominal arguments), and (ii) an indication of the noun's semantic role as determined by the verb's subject. In examples (2) and (3), the verbs *dare* 'to give' and *fare* 'to make/ do' are used in a non-standard sense, contributing no additional meaning to their respective NPs *risposta* 'answer' and *doccia* 'shower'. Moreover, given that the subject of the verb in (2) and (3) functions as a semantic argument of the noun, it is possible to construct an acceptable NP where the verb's subject acts as a dependent, with the use of any preposition. Verb reduction for the LVC in (2) is represented in (4):

(4) *Il Presidente ha dato una risposta ai presenti* → *La risposta del Presidente ai partecipanti*
'The President gave an answer to those present' → 'The President's answer to the participants'

LVCs only express valency complements that relate to the semantic participants of the noun. For instance, the noun *risposta* 'answer', derived from an intransitive two-argument agentive verbs *rispondere* 'to answer', carries two thematic roles. When combined with the ditransitive verb *dare* 'to give', which has three argument slots, *risposta* 'answer' fits naturally. In contrast, the noun *doccia* 'shower', derived

from the intransitive pronominal one-argument verb *docciar*(*si*) 'to shower', has only one thematic role. It combines with *fare* 'to make/ do', offering two argument slots (Samek-Lodovici, 2003, 835–836). The underlying idea is that in LVCs, an argument transference (or a similar mechanism) allows the NP to assign its thematic-roles to the V (Grimshaw and Mester, 1988, 211–219). Therefore, broadening the scope of LVC analysis, any syntactic argument of the verb could potentially assume the semantic argument of the noun.

2.2 Semantic and syntactic properties of NPs

In LVCs, at least one semantic argument of the verb acts as a mandatory and specific participant in the event described by the noun, due to the noun's predicative nature. Predicative nouns (predNs) are nouns that express predicates whose meaning is fully specified by their semantic arguments. For example, the predN *risposta* 'answer' in (2) denotes an event with two semantic arguments: who gives the answer and who receives it, as in example (4). In contrast, the predN *doccia* 'shower' in (2) denotes an event with one semantic argument 'the one who takes a shower'.

Referring to the LVC in (3), the noun *doccia* 'shower' shares semantic features such as duration, dynamicity, and telicity that align with those of the verb *fare* 'to make/ do' (Grimshaw, 1990, 10). This is corroborated by the use of aspectual modifiers (IN/ FOR test) like the complement of duration *di dieci minuti* 'of ten minutes', that represents a natural beginning and end point, as illustrated in example (5):

(5) *Così per fare una doccia di dieci minuti anziché consumare*
thus to make.INF INDEF.F.SG shower of ten minutes instead consume
100 litri se ne usano circa la metà.
100 liters one of.it use.3PL.PRS about the half
'Therefore, to take a ten-minute shower instead of using 100 liters, about half is used.'

The verb *fare* 'to make/ do' provides a framework for the action described by the construction, but it does not specify the action in great detail. The noun determines the verb's meaning and retains its meaning within the construction, as in other contexts. The more specific information in (3) and (5) comes from the count predN *doccia* 'shower' that gives clarity to the nature of the action (shower-making). It refers to inherently bounded situations, signifying occurrences delimited by specific spatial and temporal boundaries. In accomplishment terms, example (5) denotes an event progressing toward a terminus (Vendler, 1967, 101). The bounded nature relies

not only on the verb but also on the majority of the predicative nominal constituent (Verkuyl, 1972, 59).

In terms of syntax, the question arises whether the NP truly functions as the direct object or if it has more restricted distribution suggesting a different syntactic role. It is necessary to explore whether the noun in Italian LVCs behaves like an argument or if it has a distinct syntactic role. We argue that the NP holds the same syntactic status as any other NP in an object position. The hypothesis of semi-compositional construction appears well-founded for examples (2) and (3), as the complete meaning of these constructions cannot be fully deduced from the meanings of their individual parts. LVs and NPs essentially consist of distinct word-level lexical entries, encompassing a nominal and a verbal component.

Notably, the NP functions as the semantic head of the construction, while the LV serves as the syntactic head. Each element contributes specific meaning and function to the overall construction. The semi-compositional LVCs demonstrate enhanced syntactic flexibility, influenced by the interaction between the syntactic verb head and the semantic noun head. This hypothesis finds support in various syntactic tests used as diagnostics for mobility (cf. Machonis, 1985, 299; Nunberg et al., 1994, 501). Returning to the example of *dare una risposta* 'to give an answer' in (2), this flexibility is evident in the results obtained from passivization in (6a), relativization in (6b), topicalization in (6c), and modification in (6d):

(6) a. *La risposta è stata data dalla giurisprudenza, cioè da*
DEF.F.SG answer give.3SG.PST.PASS by case law namely by
quei magistrati, in primo luogo quelli della Cassazione […].
those judges primarily those of_the Court of Cassation […]
'The answer was provided by case law, namely by those judges, primarily those of the Court of Cassation […].' (laleggepertutti.it)

b. *Le vorrei chiedere un altro chiarimento su una*
you like.1SG.COND ask.INF a further clarification on INDEF.F.SG
risposta che ha dato un lettore.
answer that give.3SG.PRS.PERF a reader
'I would like to ask you for further clarification on a response you gave to a reader.' (repubblica.it)

c. *Come dicevamo, la vera risposta la darà Google alla fine*
as say.1PL.PST DEF.F.SG real answer it give.3SG.FUT Google at end
del corrente mese, […].
of current month […]
'As we were saying, the real answer will be provided by Google at the end of this month, […].' (altervista.org)

d. *Dopo 4 ore di esami, i medici legali hanno*
after 4 hours of examinations the doctors forensic give.3PL.PRS.PERF
così dato la prima risposta agli interrogativi sospesi sulla
thus DEF.F.SG first answer to_the questions lingering on_the
fine del primo cittadino.
end of_the first citizen
'After 4 hours of examinations, the forensic doctors thus provided the initial answer to the lingering questions about the mayor's condition.' (ristretti.it)

In (6a), the NP *la risposta* 'the answer' shifts to become the subject of the passive sentence, with the verb immediately following it. For both (6b) and (6c), the NP precedes the LV, and in (6c), it acts as the antecedent for an anaphoric pronoun. In (6d), the LV and the NP are separated by an adjective.

Therefore, the LVCs are typically defined as "combinations of predicative nouns and semantically weak or reduced verbs, where the noun subcategorizes semantically and the verb subcategorizes syntactically" (Langer, 2004, 173).

3 (In)definiteness

Most linguists would concur that a definite NP presupposes that the speaker assumes the addressee is able to identify the specific entity being referenced, based on the sentence itself, previous discourse, the current context, the addressee's general knowledge, or other pertinent sources. Conversely, an indefinite NP indicates that the speaker assumes the hearer lacks this specific knowledge (cf. Lyons, 1999, 1–2; Birner and Ward, 1994, 93–94).

The complexity of the situation becomes apparent upon closer examination. Different forms of definiteness and indefiniteness exist, influencing levels of specificity and implying varied assumptions about existence. In this context, Prince (1981, 233, 245) observes that both definite and indefinite NPs reveal different ways of referring. Without engaging with the intricate discussions from various theoretical perspectives, including logical semantics, functionalism, psycholinguistics, and computational linguistics, the concept of (in)definiteness is frequently explained through factors such as uniqueness, identifiability and inclusiveness (Russell, 1905, 481; Christophersen, 1939, 72; Hawkins, 1978, 167; Hawkins, 1984, 649; Löbner, 1985, 299; Kadmon, 1990, 283; Neale, 1990; Lyons, 1999, 7–11; Lambrecht, 2000, 613–616), familiarity (Christophersen, 1939, 75; Karttunen, 1976, 364; Kamp, 1981, 279; Heim,

1982, 193–196; Lyons, 1999, 2–7), or salience (Lewis, 1979, 178; von Heusinger, 1997, 18–22).

The distinction between familiarity/ anaphoricity and uniqueness closely parallels the distribution of strong and weak definites in German (Cieschinger, 2006, 30; Puig-Waldmüller, 2008, 148; Schwarz, 2009, 22–32; Schwarz, 2019, 3–5). The weak/ strong distinction, first introduced by Milsark (1977, 8) and revisited among others by Diesing (1992, 8–11), has been examined in terms of cardinality, symmetry, or existentiality (cf. Keenan, 1987, 291; Blutner, 1993, 134; Poesio, 1994, 284–286; Zucchi, 1996; McNally, 1998, 3; Carlson and Sussman, 2005, 75; Carlson et al., 2006, 2013; Klein et al., 2013, and others). Based on this assumption, definite and indefinite articles may each have a weak and a strong form. Strong definites are mostly used in anaphoric contexts, reflecting familiarity. In contrast, weak definites mainly occur in immediate and larger-situation contexts, indicating uniqueness in Hawkins' terms. Unlike regular (strong) definites, weak definites do not refer to uniquely identifiable individuals.

Additionally, specificity and referentiality play significant roles across various theories (cf. Quine, 1960, 141–156; Kripke, 1977, 257; Givón, 1978, 293–294; Hawkins, 1978, 167–168, 203–209, 215; Fodor and Sag, 1982, 355; Lazard, 1984, 283; Croft, 1988, 161; Croft, 2003, 130–132; Enç, 1991, 21; Ludlow and Neale, 1991, 176; Farkas, 1995, 119; Farkas, 2002, 229; Larson and Segal, 1995, 334–342; Haspelmath, 1998, 95–98; Lyons, 1999, 58; Aissen, 2003, 438; von Heusinger, 2002; Leonetti, 2004, 76–79; de Swart, 2007, 135). Specificity relates to elements already established in the discourse. A specific noun phrase presupposes the existence of a uniquely identified entity, indicating that the speaker has a particular referent in mind. In contrast, a non-specific NP signals that the identity of the entity is not established and no specific referent is intended. Only unidentifiable expressions can have a non-specific interpretation. According to von Heusinger (2002, 252), specificity is a 'referential property' of NPs, which transcends the definite versus indefinite distinction and is akin to genericity (cf. Krifka and Gerstner-Link, 1993, 966–978). Both specificity and non-specificity can apply to indefinite NPs or to NPs marked by definite markers or definite articles, but with a generic reading.[2]

In Italian, the masculine and feminine definite articles *il/ la* 'the' and the indefinite articles *uno/ a* 'a' are used to highlight or distinguish a discourse referent that the speaker believes may or may not be uniquely identifiable to the hearer. This concept of definiteness is tied to the determiner system, which reflects scalar or hierarchical values depending on the type of determiner used (Croft, 1988, 166).

2 The theoretical discussion on definiteness has been revisited by see Reimer and Bezuidenhout (2004); Schwarz (2009); Coppock and Beaver (2015), and Aguilar-Guevara et al. (2019).

However, not all NPs in Italian include an article. For instance, expressions like *fare attenzione* (lit. 'to make attention') 'to pay attention' or *dare retta* (lit. 'to give reason') 'to listen to (someone)' lack articles. This study will not address LVCs that omit articles, as they typically function as idiomatic or fixed expressions. In these instances, the noun often refers to a general or abstract concept, which in Italian does not always require an article. Such expressions are usually understood as cohesive semantic units.

Consider the examples presented in (7) :

(7) a. *Gianni ha comprato una macchina questa mattina.*
John buy.3SG.PRS.PERF INDEF.F.SG car this morning
'John bought a car this morning.'

b. *Gianni ha comprato la macchina questa mattina.*
John buy.3SG.PRS.PERF DEF.F.SG car this morning
'John bought the car this morning.'

In (7b), the NP *la macchina* 'the car' is more definite and specific compared to *una macchina* 'a car'. However, as previously mentioned, *una macchina* 'a car' still refers to a particular car from the speaker's perspective. The key difference lies in the fact that the reference of *la macchina* 'the car' is assumed to be clear to both the speaker and the hearer. In contrast, the car mentioned in (7a) is not yet established in the hearer's mental context. This distinction supports the familiarity hypothesis. Example (7b) illustrates a situational use of *la* 'the', indicating that the context is immediate, identifiable, or relatively close, even if the specific referent is not physically present. The presupposition of identifiability suggests that the speaker believes the referent is already represented in the hearer's long-term memory during the conversation. Additionally, the definite article implies uniqueness, signaling that only one entity fits the given description. On the other hand, the indefinite article *una* 'a' allows for a similar interpretation while also suggesting that the car could be one of several options.

However, several factors beyond the noun and determiner can influence reference. These include linguistic and non-linguistic context, the salience of the referent, world knowledge, and the syntactic position and information status of the noun phrase. In example (7a), the indefinite NP *una macchina* 'a car' can be interpreted in two ways: with a stronger specific reading (indicating that the speaker has a particular care in mind) or with a weaker non-specific reading (referring to any care). A noun phrase achieves a referential interpretation within an utterance if it not only "refers to an entity in the world" (Bennett, 2002, 167) but also ensures that, as a consequence of hearing the noun phrase, both the speaker and the listener can identify and fully recognize the same entity in "the world" (Lyons, 1999, 254).

Definite articles can exhibit various types of referentiality. They typically refer to something that uniquely satisfies their predicate within a general or specific context. For example, a bridging definite refers to an object that is uniquely identified in relation to another specified object, as illustrated in example (7b). In addition to this, definite articles can refer to a specific NP that does not necessarily imply familiarity or uniqueness. For instance, in the sentence in (8):

(8) *Gianni legge il giornale tutti i giorni.*
John read.3SG.PRS DEF.M.SG newspaper every the days
'John read the newspaper every day.'

The definite article *il* here does not refer to a particular newspaper. Instead, it refers to the activity of reading newspapers in general. This usage illustrates a weak definite construction, where the focus is on the familiar activity rather than a specific newspaper. In the context of weak definites, familiarity helps to explain why certain noun phrases, like *andare all'ospedale* 'go to the hospital', do not refer to a unique entity but instead evoke a steroetypical activity, like receiving care.

To address this issue, various LVCs were collected and analyzed, alternating between the use of definite and indefinite articles, and vice versa.

4 Data analysis

This section examines the alternation between definite and indefinite NPs in Italian LVCs to analyze the nuanced meanings shaped by context, with a particular focus on the role of definite and indefinite articles. The attention is focused on constructions that can be manipulated syntactically and semantically only in accordance with the speaker's grammatical competence and the pragmatic conditions in a given discourse without involving a change of meaning.[3]

3 The definiteness/ indefiniteness alternation allows us to distinguish two main types of verbo-nominal constructions: on the one hand, those that do not involve a change of idiomatic meaning, and on the other hand, those that do so. Consider the following example in (i):

(i) *Sono andata a fare la spesa per papà.*
go.1SG.PRS.PERF to make DEF.F.SG grocery shopping for dad
'I went grocery shopping for Dad.'

The components of *fare la spesa* have a literal connection to the overall idiomatic meaning. They identify normally decomposable idioms (Gibbs and Nayak, 1989), in which *spesa* refers to an actual 'grocery'. The idiomatic use of the definite article corresponds to its standard use to establish

The analyses and reflections are based on examples taken from the itTenTen20 corpus, an internet-based Italian Web 2020 dataset accessible via the Sketch Engine corpus manager (Kilgarriff et al., 2004). The selected examples primarily focused on the most frequently collocates of the verbs *fare* 'to make/ do' and *dare* 'to give' used as definite or indefinite direct objects in constructions with denotative meaning. Examples with connotative implications are intentionally excluded from this study. Each collocate is examined to illustrate the alternation between definite and indefinite determiners, along with their respective occurrence counts in the corpus. This approach highlights how the choice of determiner influences the interpretation of LVCs in Italian, shedding light on the subtleties of meaning that arise in different contexts.

Consider examples (3), repeated for convenience in (9). Corpus data reveal that the LVC *fare la doccia* is the most common, with a total of 17,118 occurrences, compared to 9,057 for the LVC *fare una doccia*:

(9) a. *Il bagno minuscolo si allagava dopo aver*
the bathroom tiny flood.3SG.PST after have.INF make.PST.PTCP
fatto la doccia (*non c'era nessun divisorio tra i*
DEF.F.SG shower no there be.IMPF no partition between the
sanitari e la doccia).
fixtures and the shower
'The tiny bathroom would flood after taking a shower (there was no partition between the fixtures and the shower).' (expedia.it)

situational reference. Commonly, always in singular form, *spesa* means the purchases that are made every day, or almost every day, of foodstuffs or household products that are necessary for the sustenance of a family or a community (the combination of *spesa* with the definite article also means 'cost', 'price', 'amount'). This construction is a highly conventional and fixed expression. In example (ii), *fare una spesa* means, instead, 'to buy something (very expensive)':

(ii) *Non è pensabile che papà abbia fatto una spesa del genere da*
NEG be.3SG.PRS thinkable that dad make.3SG.PRS.PERF INDEF.F.SG expense like this
solo.
alone
'There is no way that Dad did all of this shopping by himself.'

The difference between *fare la spesa* (lit. 'make grocery') and *fare una spesa* (lit. 'to make expense') comes from a difference in the meanings of *spesa*, not from the definiteness/ indefiniteness alternation.

b. *Appena arrivato a casa ho fatto una*
as soon as arrive.PST.PTCP.1SG.M to house make.1SG.PRS.PERF INDEF.F.SG
doccia ghiacciata.
shower frozen
'As soon as I got home, I took an icy shower.' (lemieorobie.com)

The definite NP *la doccia* 'the shower' in (9a) suggests that the action of taking a shower is implied to be something one does regularly. However, it does not refer to an individuated entity, but to an entity that is representative of its kind, or for a general concept. Consider example (10):

(10) *Luca ti ha chiesto di comprargli il giornale.*
Luca 2SG.OBJ ask.3SG.PRS.PERF to buy.INF.3SG.OBJ DEF.M.SG newspaper
'Luca asked you to buy him the newspaper'.

The definite NP *il giornale* 'the newspaper' is understood to be specific, strong, and referential, meaning it is uniquely identifiable within the relevant discourse domain. This domain can be considered the common ground shared by the speaker and the hearer, which encompasses their mutual knowledge (Stalnaker, 1979, 321; Clark and Marshall, 1981, 16–21). Conversely, *la doccia* 'the shower' in example (9a) illustrates a non-specific (generic) definite NP. The definite article *la* 'the' does not refer to a specific, pre-established event, as seen in (10). Rather, *la doccia* allows for a weak interpretation typical of weak definite NPs. While it is not directly connected to the immediate linguistic or non-linguistic context, it retains a unique status within the broader situational framework. Lacking a concrete referential anchor, this weak definite NP serves as a generic term, indicating a category rather than an individual (Aguilar-Guevara and Zwarts, 2011, 180–182).

In contrast, the indefinite NP *una doccia* 'a shower' in example (9b) refers to a specific entity within a particular context, making it a strong indefinite NP. The modifier *ghiacciata* 'frozen' specifies a type of showering that lies outside the speaker' usual experience. This differs from non-specific indefinite NPs, which refer to any member of a category without particular reference. The indefinite NP *una doccia* 'a shower' does not introduce a new discourse referent like regular indefinites; rather, it restricts the range of a variable that has already been introduced in the discourse.

Consider other examples in (11) and (12). Similar to examples (8) and (9a), the corpus data show that the LVC *fare la spesa* is the most frequent, with a total of 83,985 occurrences, compared to 2,772 for *fare una spesa*:

(11) *Allora quando vai a fare la spesa per te,*
so when go.2SG.PRS to make DEF.F.SG grocery shopping for yourself
per la tua famiglia, ricordati di chi ha fame e non
for your family remember.2SG.IMP of who have.3SG.PRS hungry and no
ha nulla.
have.3SG.PRS nothing
'So when you go shopping for yourself, for your family, remember those who are hungry and have nothing.' (asdolr.it)

(12) *Come al solito nel weekend, hai fatto una*
as usual in_the weekend make.2SG.PRS.PERF INDEF.F.SG
spesa abbondante per l'intera settimana.
grocery shopping big for the-whole week
'As usual on the weekend, you did a big grocery shop for the whole week.' (altervista.org)

In these contexts, *spesa* in its singular form typically denotes a 'grocery'. In the LVC *fare la spesa* in (11), the definite NP indicates regular purchases, often daily, of food or household essentials necessary for sustaining a family or community.[4] In contrast, example (10) operates similarly to example (9b), as it refers to a particular grocery item recognized by the speaker or hearer in that context. Similarly to examples (9a) and (9b), the LVCs in (11) and (12) consist of a weak non-specific definite NP and an indefinite NP that denotes a specific entity.

Now, consider the following examples in (13a) and (13b). The indefinite NP *fare una passeggiata* is the most common in the corpus, with 39,488 occurrences, while *fare la passeggiata* appears 2,705 times:

(13) a. *Anche se sono numerosi coloro che regolarmente*
even though be.3PL.PRS numerous who that regularly
fanno la passeggiata domenicale attorno il Lago di
make.3PL.PRS DEF.F.SG walk Sunday around the Lake of
Poschiavo [...].
Poschiavo [...]
'Even though many people regularly take a Sunday walk around Lake Poschiavo [...]' (toltedalcassetto.it)

4 Additionally, using *spesa* with the definite article conveys meanings such as 'cost', 'price', 'amount'.

b. *Giocano a basket, fanno una lunga passeggiata*
play.3PL.PRS basketball make.3PL.PRS INDEF.F.SG long stroll in
nei boschi con il cane Ketti.
the woods with the dog Ketti
'They play basketball and take a long walk in the woods with their dog Ketti.' (menteinpace.it)

The LVC *fanno la passeggiata* in (13a) refers to a specific walk that occurs on Sunday, often as a weekly tradition. This activity is generally well-known and anticipated. The definite article and the modifier *domenicale* 'Sunday' emphasize the particular nature of this walk, providing a strong specificity to the LVC. In contrast, the indefinite NP in (13b) highlights a specific entity within a given context. Here, the modifiers *lunga* 'long' and *nei boschi* 'in the woods' denote a more general, indefinite reference that exists outside of one's usual experience.

Here are some examples of LVCs formed with the LV *dare* 'to give'. Let's start from example (2), repeated in (14), along with example (15). The indefinite NP *dare una risposta* is the most prevalent in the corpus, with 91,505 occurrences, whereas *dare la risposta* appears only 20,286 times:

(14) *Il Presidente ha dato una risposta ai presenti in aula*
the president give.3SG.PRS.PERF INDEF.F.SG answer to presents in chamber
sulla situazione in corso.
on_the situation ongoing
'The President has given an answer to those present in the chamber on the ongoing situation.' (marche.it)

(15) *I giocatori gridano a gran voce le loro risposte e il*
the players shout.3PL.PRS a loud voice their answers and the
Presentatore assegna i punti a ogni giocatore che ha dato
Host award.3SG.PRS the points to each player who give.3SG.PRS
la risposta esatta [...].
DEF.F.SG answer correct [...]
'The players shout their answers loudly, and the Host awards points to each player who gives the correct answer [...].' (wikipedia.org)

In example (14), *una risposta* 'an answer' serves as a non-specific indefinite NP that refers to any member of a category without specifying a unique referent. This is similar to example (16) below:

(16) *Voglio comprare una bicicletta.*
want.1SG.PRS buy.INF INDEF.F.SG bicycle
'I want to buy a bicycle.'

In this case, the indefinite NP *una bicicletta* 'a bicycle' is also non-specific, indicating any one entity from the category of 'bicycle', rather than a specific one. In contrast, the weak definite NP *la risposta* 'the answer' in (15) implies a specific, known entity that both the speaker and hearer identify.

Consider the examples provided in (17) and (18):

(17) *In ogni caso resta l'impegno dei volontari per dare*
in any case remain.3SG.PRS the=commitment of_the volunteers to give
un aiuto a chi soffre.
INDEF.M.SG help to who suffer.3SG.PRS
'In any case, the commitment of the volunteers to provide help to those who are suffering remains.' (agira.org)

(18) *La comunità internazionale forse non sta dando l'aiuto*
the community international maybe not give.3SG.PRS.PROG DEF.M.SG=help
che servirebbe.
that serve.3SG.PRS.COND
'The international community may not be providing the help that is needed.' (m.famigliacristiana.it)

In example (17) is similar to example (14). The indefinite NP refers to an action within a given context, even though it is not definitely identifiable. In (18), the definite NP serves a non-anaphoric function, pointing to something not explicitly mentioned but assumed to be known by the hearer. This assumed knowledge may arise from general awareness or inferences that the hearer can reasonably make within the context. This reflection can be extended to the following two examples:

(19) *Ogni tanto andate a dare* un'occhiata *per controllare la*
from time to time go.2PL.IMP give.INF INDEF.F.SG=look to check out the
pubblicazione di nuove proposte […].
publication of new proposals […]
'From time to time, visit to check out the publication of new proposals […].' (lannaronca.it)

(20) *Prima di iniziare, però, diamo la consueta occhiata*
before to begin however give.PL.IMP DEF.F.SG usual look
all'etimologia della parola regina di questo articolo.
at_the=etymology of_the word central of this article
'Before we begin, let's take a look at the etymology of the word central to this article.' (marinaatzori.it)

In example (19), the indefinite NP denotes a non-specific action within a given context, while the definite NP in (20) connects to the hearer's prior knowledge of the action. This last type of construction is rarely encountered in the corpus, with only 53 occurrences, compared to 111,014 for the definite NP.

Another key aspect of the collected examples is the role of NPs modifiers within the LVCs in (11), (12), (13a), (15), (18), and (20). Analyzing these constructions reveals that modifiers in indefinite constructions provide general qualities or characteristics, indicating a type of entity or event. In contrast, modifiers in definite constructions specify known items or previously mentioned entities, often serving a contrastive function that distinguishes them from others. For instance, the adnominal modification *ghiacciata* 'frozen' in *ho fatto una doccia ghiacciata* 'I took an icy shower' in (9b) defines a kind of shower, and makes the indefinite NP specific, whereas *la passeggiata* 'the stroll' in (13a) is a familiar LVC triggered by the adjective *domenicale* 'Sunday'.

5 Theoretical framework: Role and Reference Grammar (RRG)

Role and Reference Grammar (RRG), developed by Van Valin and LaPolla (1997) and further elaborated by Van Valin (2005), is a monostratal theory of syntax. It proposes a direct connection between syntactic and semantic representations without relying on abstract syntactic structures. Instead, RRG employs a linking algorithm that includes a set of rules governing the syntax-semantics interface. Additionally, discourse-pragmatics may influence the relationship between syntactic and semantic representations.

The RRG linking algorithm operates bidirectionally: it maps semantic representations to syntactic structures (semantics-to-syntax) and links syntactic representations back to their semantic counterparts (syntax-to-semantics). Characterized as a "projectionist" theory, RRG posits that the semantic representation of a clause emerges from the lexical representation of the verb, significantly influencing the clause's syntactic structure (Van Valin, 2013, 68).

The syntactic units within the layered structure of a clause, specifically the nucleus, core argument, periphery, core, and clause, are linked to a corresponding set of semantic elements: the predicate, semantic arguments of the predicate, non-arguments, the combination of predicate and argument, and the combination of predicate, argument, and non-argument.

Operators are closed-class grammatical categories that modify various layers of the clause, such as aspect, negation, and tense. They are represented in a distinct

projection of the clause. The syntactic relationships among units in a complex construction, known as nexus relations, can occur at the nuclear, core, or clause level (Van Valin, 2005, 188).

In RRG, the semantic representation is based on the lexical features of predicates in the nuclei, utilizing an Aktionsart-based decompositional approach (Van Valin, 2013). These representations, known as Logical Structures (LSs), describe the semantic and argument structure of predicates. They integrate features from Vendler's Aktionsart classes (1967, 149), States, Activities, Achievements, and Accomplishments, and Dowty's decompositional system (1979, 33). The general linking schema in RRG is represented in Figure 1.

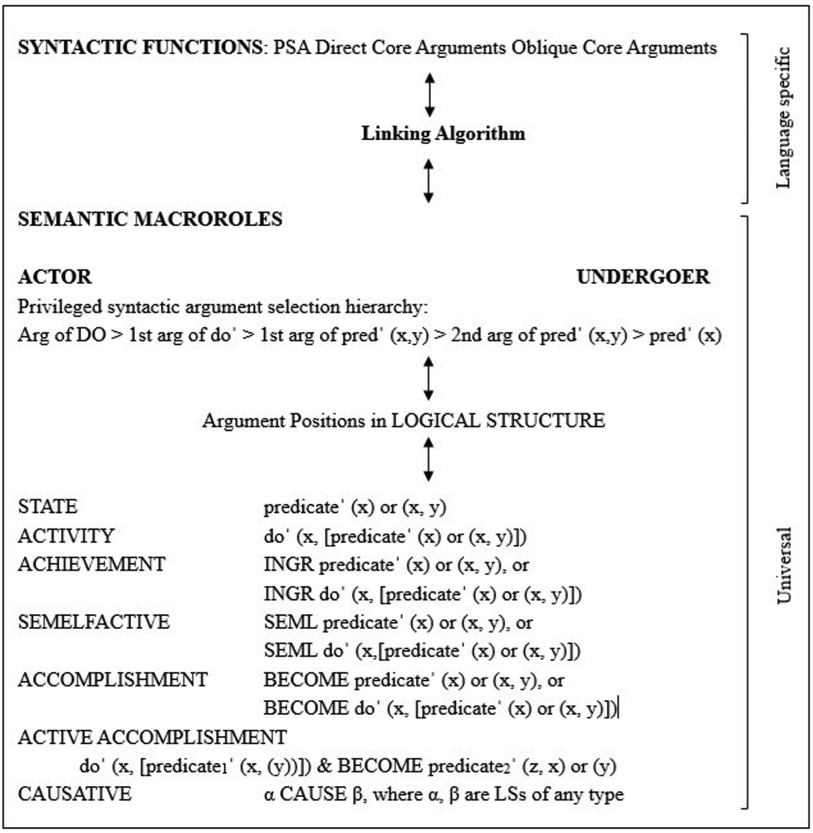

Fig. 1: Summary of RRG linking system (revisited from Van Valin, 2005, 129.)

In RRG, the theory of semantic roles identifies two primary macroroles: actor and undergoer (Van Valin, 2004, 62; Van Valin, 2005, 60). These macroroles are the main arguments in transitive predications, while intransitive verbs can have either an actor or an undergoer as their single argument, depending on their semantics. The selection of actor and undergoer in a Logical Structure (LS) is determined by the Actor-Undergoer Hierarchy (AUH). According to the AUH, for a given transitive verb, the leftmost argument in the LS is designated as the actor, and the rightmost as the undergoer (Van Valin, 2005, 61).

6 The RRG analysis of definiteness in Italian LVCs

Analyzing definiteness in Italian LVCs from an RRG perspective entails a thorough examination of how syntactic structures, such as the use of articles, map onto semantic roles and logical structures. By examining both the syntactic positions and the semantic implications of definiteness, we can better understand how Italian LVCs express specificity and generality in meaning. To begin, it is essential to examine the syntactic positions and semantic implications of the heavy verbs *fare* 'to make/ do' and *dare* 'to give.' This initial analysis will provide a foundation for comparing these findings with the structures of LVCs.

The connection between syntax and semantics is governed by the 'Completeness Constraint,' which posits that every specified argument in a sentence's semantic representation must be mirrored in its syntactic structure (Van Valin, 2005, 129). Conversely, each syntactic element must have a corresponding component in the semantic representation to ensure accurate interpretation. Consider the non-LVC examples (21) and (22), where the verb functions as a heavy verb and the noun has a concrete interpretation:

(21) *Han fatto il bagno nuovo nella casa vecchia.*
make.3PL.PRS.PERF DEF.M.SG bathroom new in_the house old
'They put the new bathroom in the old house.' (lettiseparati.it)

(22) *Mi date le risposte*: *io le posso consegnare*, *prima della fine del mese.*
me give.3PL.IMP DEF.F.PL answers I them can.1SG.PRS submit.INF before the end of_the month
'Give me the answers: I can submit them before the end of the month.' (atma-o-jibon.org)

The process of linking syntax to semantics during comprehension unfolds in three stages. The initial stage involves extracting all available information from the overt morphosyntactic form of the sentence, which includes the verb's voice (if applicable), case markings, word order, and adpositions. The second stage entails retrieving the logical structure of the clause from the lexicon and assigning macroroles where possible. The final stage involves mapping the arguments onto semantic relations.

The Layered Structures of the Clause (LSC) depicted in Figure 2 effectively represent the syntax-semantics linking for example (21). The essential components of the LSC in Figure 2 are the nucleus, which includes the verb predicate *fare* 'to make/ do', and the core, which contains both the nucleus and the arguments of the predicate within the nucleus: the null subject ∅ core argument and the direct core argument *il bagno nuovo* 'the new bathroom'. In terms of the AUH, the leftmost subject core argument in the LS (the null subject ∅) is selected as Actor, while the rightmost direct core argument is identified as the Undergoer (i.e., *il bagno nuovo* 'the new bathroom'). In the framework of RRG's lexical decomposition system (Van Valin and LaPolla, 1997; Van Valin, 2005), the general semantic representation for the predicate *fare* 'to make/ do' is illustrated in (23):

(23) **[do'** (x, ∅)] CAUSE [BECOME **predicate'** (y)]

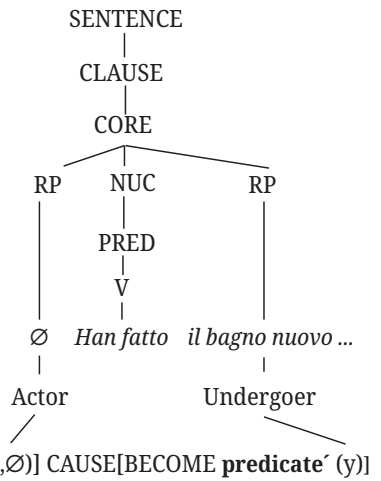

Fig. 2: From syntax to semantics for (21).

Han fatto il bagno nuovo is an Accomplishment predicate involving the unstructured activity of constructing and a culmination. These components are connected through

a BECOME event, which incorporates an incremental structure. They BECOME event associated with this scenario is BECOME-BUILT/ PUT, culminating when the bathroom (the theme) 'becomes built'.

Following the same general procedure outlined in the algorithm, the layered structures in Figure 3 represents the syntax and semantics linking of example (22).The components of the LSC illustrated in Figure 3 include the nucleus that contains the verb predicate *dare* 'to give', and the core, which incorporates the nucleus along with the predicate's arguments: the null subject ∅ core argument, the direct core argument *le risposte* 'the answers', and the receiver *mi* 'me'. According to the AUH, the leftmost subject core argument in the LS (the null subject ∅) is designed as the Actor, while the rightmost direct core argument is classified as the Undergoer (*il bagno nuovo* 'the new bathroom'). The theoretical label for the third argument in a ditransitive construction within the RRG framework is 'non-macrorole direct core argument' (NMR) (Van Valin, 2005, 64). In the RRG's system of lexical decomposition the general semantic representation for a three-place predicate is expressed in (24).

(24) **[do'** (x, ∅)] CAUSE [BECOME **predicate'** (y, z)]

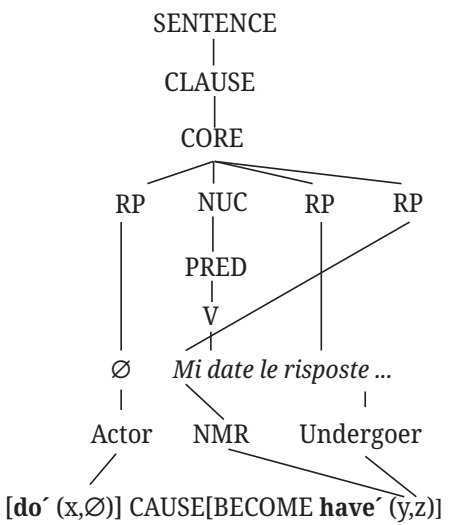

Fig. 3: From syntax to semantics for (22).

The verb *dare* 'to give' in this context means 'to hand over', specifically referring to the transfer of a tangible item (*le risposte*) from one individual (the null subject) to another (the recipient *mi*). In this scenario, the initiator of the action is represented

by the null subject ∅ core argument, the recipient is denoted as the first argument of BECOME (y, z), which is *mi* 'me', and the transferred object is the second argument, *le risposte* 'the answers'. Consider the LVCs in (25) and in (26):

(25) *Dopo cena poi Daniele e il suo corrispondente hanno fatto*
after dinner then Daniele and his correspondent make.3PL.PRS.PERF
il bagno in una vasca di acqua caldissima.
DEF.M.SG bath in a tub of water very hot
'After dinner, Daniele and his correspondent took a bath in a tub of very hot water.' (quinewsvaldera.it)

(26) *Il Presidente ha dato una risposta ai presenti in aula*
the president give.3SG.PRS-PERF INDEF.F.SG answer to presents in chamber
sulla situazione in corso.
on_the situation ongoing
'The President has given an answer to those present in the chamber on the ongoing situation.' (marche.it)

As discussed in Section 2.1, LVCs have a simple nucleus where the verb serves as the sole lexical head. This verb supplies morphological features such as tense, number, aspect, and diathesis, which in turn determine the syntactic roles of the nominal arguments. For instance, the NP *il bagno* 'the bath' naturally pairs with the transitive verb *fare* 'to make/ do', which requires two argument slots. Conversely, the NP *la risposta* 'the answer' is better suited to the ditransitive verb *dare* 'to give', which accommodates three argument slots. Syntactically, only the verb serves as the head of the core, while the noun typically acts as a core argument, most often the direct object. As a result, LVCs exhibit a semi-productive and semi-compositional nature, which is best analyzed through a system of lexical decomposition.

In forming LVCs, the verb takes an event-denoting noun as its argument. This predicative noun combines its meaning with part of the meaning provided by the verb. For LVCs that denote activities, such as those examined here, these constructions do not incorporate the entire [do'(x, ∅)] CAUSE [BECOME ...] segment of the LS of *fare* 'to make/ do' or *dare* 'to give'. Instead, they include only the portion pertaining to the activity do'(...). This representation for the NP fills the second position in the LS. This gives rise to the representations in (27a) for the LVC (25) and (27b) for the LVC in (26):

(27) a. **do'**(Daniele e il suo corrispondente, [**bath'**(Daniele e il suo corrispondente)])
b. [**do'**(Presidente, [**express.as.answer'**(Presidente, situazione)]] CAUSE [BECOME **aware.of'**(presenti, situazione)]

The NPs *bagno* 'bath' and *risposta* 'answer' with heavy verbs and light verbs derive parts of their meaning from different aspects of their qualia (Pustejovski, 1991, 418; Pustejovski, 1995, 76; Pustejovsky, 1998, 330; Pustejovsky, 2001, 91; Van Valin, 2005, 50–52; Van Valin and LaPolla, 1997, 183–184). The meaning of nouns consists of information concerning the type of entity (i.e., whether it is a physical object, information, etc.), its components, and the eventualities associated with those entities. The examples in (21) and (22) refer to concrete entities, while those in (25) and (26) refer to activities. The LS in (25) represents qualia indicating a bath-taking event. The event is conventionalized and often culturally embedded. The LS in (24) illustrates qualia indicating an answer that can be communicated and created (presumably by writing it). Thus, the sentence in (22) implies that an individual received a physical object containing the answer, while the example in (26) suggests that an individual was informed by receiving the answer through communication.

Both the PredN and the LV function as co-predicators, contributing to the LS (cf. Rappaport-Hovav and Levin, 2001, 782). Therefore, the LS is the result of the co-composition of the LS for the verb and the LS for the noun (Van Valin, 2013, 98–104). The verb provides the lexical template, either in whole or in part, and the qualia structure of the nominal replaces a portion of the lexical template provided by the verb; specifically, the nominal should supply the predicate in the logical structure (predicate'(x) or predicate'(x, y)). Consequently, the argument structure of the resulting LVC is not directly licensed by either the noun or the verb alone; rather, it emerges from the integrated LS.

This section concludes with final reflections on the domain of weak referentiality within LVCs, a phenomenon wherein definite or indefinite NPs lack their typical referential force. As elucidated in Section 4, the potential for weak interpretations is contingent upon information stored within the lexicon and shared by the speakers. The linkage of an entity with an activity in the verb's core meaning is explicitly incorporated into the NP's lexical entry, specifically within its functional qualia structure (Pustejovski, 1995, 76).

The defining characteristics of weak readings is that the verb-core denotes a prototypical activity. Consequently, the definite and indefinite articles in LVCs, such as those here examined in this study, contribute their conventional semantics not by combining with the RP but with the verb-argument relations in the clause. According to this analysis, Carlson et al. (2013, 17–19), building upon the work of Klein et al. (2013), suggests that weak definites and indefinites could be interpreted similarly, giving both incorporated interpretations following the compositional structure ARTICLE(V(N)). Therefore, the semantic composition in (25) and (26) can be integrated as in (28a) and (28b):

(28) a. DEF < **do'**(x, [**bath'**(x)]))
b. DEF < **do'**(x, [**express.as.answer'**(x, y)] CAUSE [BECOME **aware.of'** (z, y)]

The definiteness operator DEF is not linked to the NP itself, but rather to the resulting verb-argument combination. This assumes a close relationship between the LV and the NP such that both seem to form a semantic predicate. For instance, the definite article in (25) does not combine with the noun, rendering the noun semantically indefinite. This accounts for the apparent violation of the uniqueness requirement and the familiarity condition in cases with weak referentiality. It may function more like a Carlsonian weak definite which presupposes neither uniquiness nor prior mention (Carlson et al., 2006). The DEF operator denotes a familiar type of activity, one whose cultural significance is independently established and encoded within the grammar, thereby presumed as known.

7 Conclusion

Focusing on the semantic-pragmatic functions of definite and indefinite articles in NPs has given rise to some reflections on the concepts of (in-)definiteness and referentiality within Italian LVCs and non-LVCs. The findings reveal that the alternation between definites and indefinites influences the specificity of the event-denoting noun, thereby affecting the overall semantics of the LVC. For example, the distinction between *fare una doccia* 'to take a shower' and *fare la doccia* 'to take the shower' shows how definiteness modifies the focus and context of the event. Likewise, variations in the use of *una* and *la* in LVCs like *dare una risposta* 'to give an answer' versus *dare la risposta* 'to give the answer' underline the implication of article choice in determining the referentiality and scope of the NP.

The analysis of the LVCs carried out using the syntax-semantics interface developed by Role and Reference Grammar (RRG) has provided a useful theoretical perspective that is able to represent the verb meaning and definite and indefinite argument realization in LVCs and non-LVCs. The definiteness operator DEF, associated with the resulting verb-argument structure information, accounts for the hypothetical deviation from the uniqueness requirement and from the familiarity condition typically related to weak referentiality in LVCs. Its presence in the logical structure of the LVC suggests that the NP form a particularly tight unit with the LV of which it is argument. The NP appears as part of a "pseudo-incorporated" structure (cf. Massam, 2001, 157). In discussing how NPs are integrated into the verb's argument structure, it is noteworthy that while NPs in LVCs appear to be closely linked to the

LV, they do not completely align with the traditional notion of incorporation. Unlike full incorporation, where the NP become an integral part of the verb, in LVCs the NP retains its status as distinct element in the sentence. This means that although the NP and the verb are closely related, the syntactic and structural changes typically associated with complete incorporation are note fully present in LVCs.

Bibliography

Aguilar-Guevara, Ana, Julia Pozas Loyo and Violeta Vázquez-Rojas Maldonado. 2019. *Definiteness across languages*. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.3265959.

Aguilar-Guevara, Ana and Joost Zwarts. 2011. Weak definites and reference to kinds. In Nan Li and David Lutz (eds.), *Proceedings of Semantics and Linguistic Theory 20*, 179–196. Ithaca, NY: CLC Publications.

Aissen, Judith. 2003. Differential Object Marking: Iconicity vs. Economy. *Natural Language & Linguistic Theory* 21. 435–483. https://doi.org/10.1023/A:1024109008573.

Alba-Salas, Josep. 2004. *Fare* light verb constructions and Italian causatives: Understanding the differences. *Rivista di Linguistica* 16(2). 283–323.

Bennett, Paul. 2002. *Semantics: an introduction to the non-lexical aspects of meaning*. Munich: LINCOM.

Birner, Betty J. and Gregory Ward. 1994. Uniqueness, familiarity, and the definite article in English. *Berkeley Linguistics Society Annual Meeting* 20. 93–102.

Blutner, Reinhard. 1993. Dynamic Generalized Quantifiers and Existential Sentences in Natural Languages. *Journal of Semantics* 10. 133–159. DOI:10.1093/jos/10.1.33.

Carlson, Greg, Natalie Klein, Whitney Gegg-Harrison and Michael Tanenhaus. 2013. Weak definites as a form of definiteness: experimental investigations. *Recherches linguistiques de Vincennes* 42. 11–32. DOI:10.4000/rlv.2158.

Carlson, Greg, Sussman Rachel S., Klein Natalie and Michael Tanenhaus. 2006. Weak definite NPs. In Christopher Davis, Amy Rose Deal and Youri Zabbal (eds.), *Proceedings of the Thirty-Sixth Annual Meeting of the North East Linguistics Society*, 179–196. Amherst, MA: GLSA Publications.

Carlson, Greg and Rachel S. Sussman. 2005. Seemingly indefinite definites. In Stephan Kepsar and Marga Reis (eds.), *Linguistic Evidence: Empirical, theoretical, and computational perspectives*, 71–86. Berlin: De Gruyter.

Christophersen, Paul. 1939. *The articles. A study of their theory and use in English*. Cophenhagen: Munksgaard.

Cicalese, Anna. 1999. Le estensioni di verbo supporto: uno studio introduttivo. *Studi Italiani di Linguistica Teorica e Applicata* 3. 447–485.

Cieschinger, Maria. 2006. *Constraints on the contraction of preposition and definite article in German*. B.Sc. thesis: University of Osnabrück.

Clark, Herbert H. and Catherine R. Marshall. 1981. Definite Reference and Mutual Knowledge. In Aravind Joshi, Bonnie Webber and Ivan Sag (eds.), *Elements of Discourse Understanding*, 10–63. Cambridge: Cambridge University Press.

Coppock, Elizabeth and David Beaver. 2015. Definiteness and determinacy. *Linguistics and Philosophy* 28. 377–435. https://doi.org/10.1007/s10988-015-9178-8.

Croft, William. 1988. Agreement vs Case Marking and Direct Objects. In Michael Barlow and Charles A. Ferguson (eds.), *Agreement in Natural Language: Approaches, Theories, Descriptions*, 159–179. Chicago: University of Chicago Press.

Croft, William. 2003. *Typology and Universals*. Cambridge: Cambridge University Press 2nd edn.
Danlos, Laurence. 1992. Support verb constructions: linguistic properties, representation, translation. *Journal of French Language Studies* 2(1). 1–32.
De Angelis, Angela. 1989. Nominalizations with the Italian Support Verb *avere*. *Lingvisticae Investigationes* 13. 223–237.
de Swart, Peter. 2007. *Cross-linguistic Variation in Object Marking*. Utrecht: LOT.
Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: The MIT Press.
Dowty, David. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.
Dras, Mark. 1995. Automatic identification of support verbs: A step towards a definition of semantic weight. In *Proceedings of the Eighth Australian Joint Conference on Artificial Intelligence*, 451–458. Canberra: World Scientific Press.
Elia, Annibale and Emilio D'Agostino. 1998. Il significato delle frasi: un continuum dalle frasi semplici alle forme poliremathiche. In Federico Albano Leoni (ed.), *Ai limiti del linguaggio*, 287–310. Roma: Laterza.
Elia, Annibale, Emilio D'Agostino and Maurizio Martinelli. 1985. Tre componenti della sintassi italiana: frasi semplici, frasi a verbo supporto e frasi idiomatiche. In A. Franchi De Bellis and L. M. Savoia (eds.), *Sintassi e morfologia della lingua italiana d'uso. Teorie e applicazioni descrittive. Atti del XVII congresso internazionale della Società di Linguistica Italiana*, 311–325. Roma: Bulzoni.
Enç, Mürvet. 1991. The semantics of specificity. *Linguistics Inquiry* 22(1). 1–25. https://www.jstor.org/stable/4178706.
Everaert, Martin. 2010. The lexical encoding of idioms. In Malka Rappaport Hovav, Edit Doron and Ivy Sichel (eds.), *Lexical semantics, syntax, and event structure*, 76–98. Oxford: Oxford University Press.
Farkas, Donka. 1995. Specificity and Scope. In Léa Nash and Georges Tsoulas (eds.), *Actes du Premier Colloque*, 119–137. Paris: Langues & Grammaire.
Farkas, Donka. 2002. Specificity distinctions. *Journal of Semantics* 19. 213–243. https://doi.org/10.1093/jos/19.3.213.
Fellbaum, Christiane. 1993. The determiner in English idioms. In Christina Cacciari and Patrizia Tabossi (eds.), *Idioms: Processing, Structure, and Interpretation*, 271–295. Mahwah, NJ: LEA.
Fillmore, Charles J., Christopher R. Johnson and Miriam R.L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography* 16(3). 235–250. DOI:10.1093/ijl/16.3.235.
Fleischhauer, Jens. 2021a. Light Verb Constructions and Their Families – A Corpus Study on German '*stehen unter*'-LVCs. In *"Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021)*, 63–69. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.mwe-1.8.
Fleischhauer, Jens. 2021b. Simplex and Complex Predicates in Persian: An RRG Analysis. In Robert Van Valin (ed.), *Challenges at the Syntax-Semantics-Pragmatics Interface*, 31–62. Newcastle upon Tyne: Cambridge Scholars Publishing.
Fodor, Janet D. and Ivan Sag. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy* 5. 355–398.
Gibbs, Raymond and Nandini Nayak. 1989. Psycholinguistic studies on the syntactic behaviour of idioms. *Cognitive Psychology* 21. 100–138.
Giry-Schneider, Jaqueline. 1987. *Les prédicats nominaux en français: Les phrases simples à verbe support*. Geneva: Droz.
Givón, Talmy. 1978. Definiteness and referentiality. In Joseph Greenberg, Charles A. Ferguson and Edith Moravcsik (eds.), *Universals of human language, vol. 4*, 291–330. Stanford: Stanford University Press.
Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, MA: The MIT Press.
Grimshaw, Jane and Armin Mester. 1988. Light verbs and theta-marking. *Linguistic Inquiry* 19. 205–232.

Gross, Gaston. 1984. Compléments adverbiaux et verbes supports. *Revue québécoise de linguistique* 13(2). 123–156.
Gross, Gaston. 1993. Trois applications de la notion de verbe support. *L'Information Grammaticale* 59. 16–23.
Gross, Gaston. 2010. Les verbes supports et l'actualisation des prédicats nominaux. In Amr Helmy Ibrahim (ed.), *Supports et prédicats non-verbaux dans les langues du monde*, 16–35. Paris: CRL.
Gross, Maurice. 1976. Sur quelques groupes nominaux complexes. In Jean-Claude Chevalier and Maurice Gross (eds.), *Méthodes en grammaire française*, 97–119. Paris: Klincksieck.
Gross, Maurice. 1981. Les bases empiriques de la notion de prédicat sémantique. *Langages* 15(63). 7–52.
Harris, Zellig. 1968. *Mathematical Structures of Language*. New York: Wiley.
Haspelmath, Martin. 1998. *Indefinite Pronouns*. Oxford: Oxford University Press.
Hawkins, John A. 1978. *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction*. London: Croom Helm Humanities Press.
Hawkins, John A. 1984. A note on referent identifiability and co-presence. *Journal of Pragmatics* 8. 649–659.
Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases in English*. Doctoral dissertation: University of Massachusetts.
Horn, George M. 2003. Idioms, metaphors and syntactic mobility. *Journal of Linguistics* 39. 245–273.
Huddleston, Rodney and Geoffrey K. Pullum. 2010. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Igor, Mel'cuk. 1996. Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In Leo Wanner (ed.), *Lexical Functions in Lexicography and Natural Language Processing*, 37–102. Amsterdam: John Benjamins.
Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles, Part VI, Morphology*. Copenhagen: Ejnar Munksgaard.
Ježek, Elisabetta. 2004. Types et degrés de verbes supports en Italien. *Lingvisticae Investigationes* 27(2). 195–201.
Kadmon, Nirit. 1990. Uniqueness. *Linguistics and Philosophy* 13. 273–324.
Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen and Martin Stokhof (eds.), *Formal Methods in the Study of Language*, 277–322. Amsterdam: Mathematisch Centrum.
Karttunen, Lauri. 1976. Discourse referents. In James D. McCawley (ed.), *Syntax and semantics,vol. 7: Notes from the Linguistic Underground*, 363–385. New York: Academic Press.
Keenan, Edward L. 1987. A semantic definition of 'Indefinite NP'. In Eric Reuland and Alice ter Meulen (eds.), *The Representation of (In)definiteness*, 286–317. Cambridge: The MIT Press.
Keller, Rudolf E. and Karl-Heinz Mulagk. 1995. *Die deutsche Sprache und ihre historische Entwicklung*. Hamburg: Buske.
Kilgarriff, Adam, Pavel Rychlý, Pavel Smrž and David Tugwell. 2004. The Sketch Engine. *Proceedings of the 11th EURALEX International Congress* 105–116.
Klein, Natalie, Tim Gegg-Harrison, Greg Carlson and Michael Tanenhaus. 2013. Experimental investigations of weak definite and weak indefinite noun phrases. *Gognition* 128(2). 187–213. https://doi.org/10.1016/j.cognition.2013.03.007.
Krenn, Brigitte. 2000. *The Usual Suspects: Data Oriented Models for the Identification and Representation of Lexical Collocations*. Doctoral dissertation: DFKI & Universität des Saarlandes, Saarbrücken.
Krifka, Manfred and Claudia Gerstner-Link. 1993. Genericity. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann (eds.), *Syntax – An international handbook of contemporary research*, 966–978. Berlin: De Gruyter.

Kripke, Saul. 1977. Speaker Reference and Semantic Reference. *Midwest studies in philosophy* 2. 255–276.
La Fauci, Nunzio. 1979. *Costruzioni con verbo operatore in testi italiani antichi. Esplorazioni sintattiche*. Pisa: Giardini.
Lambrecht, Knud. 2000. When subjects behave like objects: a markedness analysis of sentence focus constructions across languages. *Studies in Language* 24. 611–682.
Langer, Stefan. 2004. A linguistic test battery for support verb constructions. *Lingvisticae Investigationes* 27(2). 171–184.
Larson, Richard and Gabriel Segal. 1995. *Knowledge of Meaning: An introduction to semantic theory*. Cambridge, MA: The MIT Press.
Lazard, Gilbert. 1984. Actance variations and categories of the object. In Frans Plank (ed.), *Objects. Towards a theory of grammatical relations*, 269–292. London: Academic Press.
Leonetti, Manuel. 2004. Specificity and object marking: the case of Spanish. *Catalan Journal of Linguistics* 3. 75–114.
Lewis, David. 1979. Scorekeeping in a language game. In Rainer Bäuerle, Urs Egli and Arnim von Stechow (eds.), *Semantics from different points of view*, 172–187. Berlin: Springer.
Ludlow, Peter and Stephen Neale. 1991. Indefinite descriptions: In defence of Russell. *Linguistics and Philosophy* 14(2). 171–202. https://www.jstor.org/stable/25001425.
Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.
Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4. 279–326. https://doi.org/10.1093/jos/4.4.279.
Lüger, Heinz-Helmut. 1995. *Pressesprache*. Tübingen: Niemeyer.
Machonis, Peter A. 1985. Transformations of verb phrase idioms: Passivization, particle movement, dative shift. *American Speech* 60. 291–308. DOI:10.2307/454907.
Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language & Linguistic Theory* 19. 153–197. https://doi.org/10.1023/A:1006465130442.
Mastrofini, Roberta. 2004. Classi di costruzioni a verbo supporto in italiano: implicazioni semantico-sintattiche nel paradigma V + N. *Studi italiani di linguistica teorica e applicata* 33. 371–398.
McGinnis, Martha. 2002. On the systematic aspect of idioms. *Linguistic Inquiry* 33(4). 665–672. https://www.jstor.org/stable/4179214.
McNally, Louise. 1998. Existential sentences without existential quantification. *Linguistics and Philosophy* 21. 353–392.
Milsark, Gary L. 1977. Toward an explanation of certain peculiarities of the existential construction of English. *Linguistic Analysis* 3. 1–29.
Neale, Stephen. 1990. *Descriptions*. Cambridge, MA: The MIT Press.
Nunberg, Geoffrey, Ivan A. Sag and Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538. https://doi.org/10.1353/lan.1994.0023.
Plante, Sebastien. 2014. Types of nominalized verbs in light verb constructions. *McGill Working Papers in Linguistics* 24(1). 81–91.
Poesio, Massimo. 1994. Weak definites. In Mandy Harvey and Lynn Santelmann (eds.), *Proceedings of the Fourth Conference on Semantics and Linguistic Theory*, 282–299. Ithaca, NY: CLC Publications.
Pompei, Anna. 2017. Verbi complessi con struttura [V+SP] e verbi supporto. Proprietà e test. *Studi e saggi linguistici* 55(2). 109–136. https://doi.org/10.4454/ssl.v55i2.208.
Pompei, Anna and Valentina Piunno. 2023. Light Verb Constructions in Romance. In Pompei Anna, Mereu Lunella and Piunno Valentina (eds.), *Light Verb Constructions as Complex Verbs*, 99–147. Berlin: De Gruyter. https://doi.org/10.1515/9783110747997-005.
Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–254. New York: Academic Press.

Puig-Waldmüller, Estela. 2008. *Contracted preposition-determiner forms in German: Semantics and pragmatics*. Doctoral dissertation: Universitat Pompeu Fabra, Barcelona.
Pustejovski, James. 1991. The Generative Lexicon. *Computational linguistics* 17. 409–441.
Pustejovski, James. 1995. *The Generative Lexicon*. Boston: The MIT Press.
Pustejovsky, James. 1998. The Semantics of Lexical Underspecification. *Folia Linguistica* 32(3–4). 323–348.
Pustejovsky, James. 2001. Type Construction and the Logic of Concepts. In Pierrette Bouillon and Federica Busa (eds.), *The Syntax of Word Meaning*, 91–123. Cambridge: Cambridge University Press.
Quine, Willard Van Orman. 1960. *Word and Object*. Cambridge, MA: The MIT Press.
Rappaport-Hovav, Malka and Beth Levin. 2001. An event structure account of English resultatives. *Language* 77. 766–797. https://www.jstor.org/stable/3086847.
Reimer, Marga and Anne Bezuidenhout. 2004. *Descriptions and beyond*. Oxford: Oxford University Press.
Russell, Bertrand. 1905. On Denoting. *Mind* 14. 479–493.
Salvi, Giampaolo. 1988. La frase semplice. In Lorenzo Renzi (ed.), *Grande grammatica italiana di consultazione, vol. 1*, 29–114. Bologna: Il Mulino.
Samek-Lodovici, Vieri. 2003. The internal structure of arguments and its role in complex predicate formation. *Natural Language & Linguistic Theory* 21. 835–881.
Schwarz, Florian. 2009. *Two types of definites in natural language*. Doctoral dissertation: University of Massachusetts, Amherst, MA.
Schwarz, Florian. 2019. Weak vs. strong definite articles: Meaning and form across languages. In Ana Aguilar-Guevara, Julia Pozas Loyo and Violeta Vázquez-Rojas Maldonado (eds.), *Definiteness across languages*, 1–37. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.3252012.
Stalnaker, Robert. 1979. Assertion. In Peter Cole (ed.), *Syntax and Semantics 9*, 315–332. New York: Academic Press.
Stichauer, Pavel. 2000. Su alcune construzioni con verbo supporto in italiano. *Linguistica Pragensia* 10. 37–50.
Stone, Megan Schildmier. 2013. Can Idioms be Passivized? *Coyote Papers* 21. 1–22.
Svenonius, Peter. 2005. Extending the extension condition to discontinuous idioms. *Linguistic Variation Yearbook* 5. 227–263. https://doi.org/10.1075/livy.5.08sve.
Van Valin, Robert. 2004. Semantic macroroles in Role and Reference Grammar. In Rolf Kailuweit and Martin Hummel (eds.), *Semantische Rollen*, 62–82. Tübingen: Narr.
Van Valin, Robert. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511610578.
Van Valin, Robert. 2013. Lexical representation, co-composition, and linking syntax and semantics. In James Pustejovsky, Pierrette Bouillon, Hitoshi Isahara, Kyoko Kanzaki and Chungmin Lee (eds.), *Advances in Generative Lexicon Theory*, 67–106. Dordrecht: Springer.
Van Valin, Robert and Randy LaPolla. 1997. *Syntax*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166799.
Vendler, Zeno. 1967. *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.
Verkuyl, Henk J. 1972. *On the Compositional Nature of the Aspects*. Dordrecht: Reidel.
von Heusinger, Klaus. 1997. *Salienz und Referenz. Der Epsilonoperator in der Semantik der Nominalphrase und anaphorischer Pronomen*. Berlin: Akademie Verlag. https://doi.org/10.1515/9783050073934.
von Heusinger, Klaus. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics* 19(3). 245–274. https://doi.org/10.1093/jos/19.3.245.
von Polenz, Peter. 1963. *Funktionsverben im heutigen Deutsch. Sprache in der rationalisierten Welt*. Düsseldorf: Schwann.

Wittenberg, Eva. 2016. *With Light Verb Constructions from Syntax to Concepts*. Potsdam: Universitätsverlag Potsdam.
Zucchi, Alessandro. 1996. *Subject clitics in Piacentino*. Manuscript.

Ramón Boldt

To come into bloom: On the emergence of light verb constructions in Old West Norse

1 Introduction

Although it seems that every publication on light verb constructions (LVCs)[1] must make reference to the fact that it is still not entirely clear, what it is that constitutes them as a class cross-linguistically as well as with a focus on a single language or closely related languages[2], some core characteristics of this class of more or less fixed expressions have become apparent by today (Butt, 2010). It is not my intention here to define in as much detail as possible what an LVC in Old West Norse is, and nor do I think a strict definition (in the sense that it does not allow for contradictions and issues without an evident solution) is possible or preferable.

What I am rather going to do here is the following: In section 2, I am going to present what is to be considered the core characteristics of prototypical light verb constructions in the Germanic languages. In section 3, I am going to present the morphosemantic core characteristics of the reconstructed Proto-Germanic verbal systems. The four classes of that system will be discussed in individual subsections. I am going to do the same for the Old West Norse verbal system in section 4. Section 5 is devoted to some early prototypical light verb and similar constructions as attested in Old West Norse, with special focus on semantics. In four subsections, causative, stative/durative, inchoative, and intensive LVC will be discussed. The article concludes with a summary of the findings and open questions.

I am doing all this in order to promote an idea concerning the emergence of Germanic LVCs that has been brought up before.[3] Namely, that it is connected to the

1 In the article, I am going to use both the abbreviation 'LVC' and the written form 'light verb construction' indiscriminately mainly because I vacillate between meeting the requirements for high reading efficiency and my reservations about using abbreviations too frequently outside a lexicon.

2 Cf. Harm's recent note, "[d]ie Klage über die Vagheit des Begriffs 'Funktionsverbgefüge' ist längst zu einem Topos geworden" 'the complaint about the vagueness of the term 'light verb construction' has long since become a topos' (2021, 5).

3 The history and development of LVCs have mostly been discussed for Old High German, see Filatkina (2018) with recent literature; Blum (1986); Relleke (1974). For case studies on German

Ramón Boldt, Friedrich-Alexander-Universität, Erlangen-Nürnberg

https://doi.org/10.1515/9783112229330-008

weakening (and subsequent collapse) of the Germanic system of weak verbs. Even though that connection has already been assumed, fleshing out the details is still in its early stages. However, I do not aim for a systematic analysis in the present article. My approach is of a rather exploratory nature.

2 Prototypical light verb constructions in Germanic

A prototypical light verb construction as understood in the present article consists of a verb, a preposition, and a noun. They jointly form a complex predicate and therefore coarticulate the situational content.

Not every verb can be part of a light verb construction, but the verbs involved are commonly verbs of movement or transfer. This criterion seems to be very consistent throughout history and languages.[4] I am therefore going to focus on them in this article. Typical light verbs are the correspondences of 'to go', 'to come' (for movement) and 'to give', 'to take', 'to put' (for transfer). It can be specified that these verbs of movement or transfer not only verbalise basic human concepts (Harm, 2021, 44), but it seems they are restricted to their basic form. That is why the German verb *gehen* 'to go' can function as a light verb, but not the more specific *rennen* 'to run'. Consider, for example, the following two sentences:

(1) a. *Mein Wunsch wird in Erfüllung gehen.*
my wish will in fulfillment go
'My wish will come true.'
b. **Mein Wunsch wird in Erfüllung rennen.*
my wish will in fulfillment run

kommen 'to come' as a light verb, see Fleischhauer and Hartmann (2021) and Fleischhauer and Hartmann (2023). For (Old) English, see Brinton and Akimoto (1999). For Germanic in general, see Boldt (2023, 107–136).

4 It can be assumed that those verbs are stable in this regard, i.e., there is no indication that they started out as main verbs and somehow developed a secondary use as a semantically reduced light verb. It has in fact been shown that already by the time of the earliest Indo-European attestations, those verbs did function as light verbs, cf. Proto-Indo-European **d*^{h}*eh*$_1$- 'to put' in Ancient Greek *ὄνομα τίθημι* [ónoma títhēmi], Vedic *nāma dhā-* etc. 'to give a name' from the Proto-Indo-European LVC **h*$_1$*neh*$_3$*m̥n d*^{h}*eh*$_{1/3}$- (on this type see Schutzeichel, 2013, Hackstein, 2002, 6–8, 15–16, and recently Ittzés, 2024). Throughout this process, noun and verb became so closely connected to one another that the former verb was phonologically reduced to *-*d*^{h}- and fused with the noun. The *-*d*^{h}- could even be abstracted and used to form new roots (Kölligan, 2018).

In sentence (1a), the basic verb of movement *gehen* 'to go' functions as light verb in the LVC *in Erfüllung gehen* 'to come true'. By comparison, sentence (1b) differs from (1a) only in the selection of the more specific verb of movement *rennen* 'to run'. However, this selection results in an ungrammatical expression *in Erfüllung rennen*.

The most striking property of the verb in a light verb construction is its lightness. That means that it differs from a non-light, i.e., a full lexical verb in that it carries a reduced, non-literal abstract meaning, even though it is identical in form and inflectional properties to such a verb. *Mein Wunsch* 'my wish' from the above sentence is not literally going anywhere, because a wish is not capable of doing so, the subject cannot move through space. Instead, the verb conveys semantic nuances that differ from the meaning of an etymologically related simple verb that often, but not always, exists. In the above example, *in Erfüllung gehen* differs from the simple passive verb *erfüllt werden* 'to be fulfilled' in that the verb in the LVC highlights that a change to the situation is taking place. It highlights the inchoative aspect of the construction as a whole.[5] When I am using the term 'aspect' here and in the following, I am always referring to 'lexical aspect' (German 'Aktionsart') as opposed to 'grammatical aspect' ('Aspekt').

While pioneering work on LVCs has regarded the marking of aspect as essential (von Polenz, 1963, 1987, but already Heringer, 1968), newer research promotes the view that this is only one of the characteristics of the LVCs (Harm, 2021, 6–49 with references). Other semantic characteristics are causativity and passivity, while higher syntactic flexibility, pragmatic reordering, and stylistic characteristics concern further parts of the language system.

Just like the verbs, not every noun can be part of an LVC. Instead, the slot is restricted to abstract nouns, i.e., nouns that denote actions, activities, processes, events, or states.[6] Since the noun is part of the predicate and not its argument, it can not be asked for or be replaced by a pronoun or another noun:

(2) *Mein Freund geht in ein Haus.*
my friend is.going into a house
'My friend is going into a house.'

5 Adopting a historical perspective, however, elucidates the fact that an etymologically related simple verb cannot always be expected to exist. Filatkina (2018, 260) shows that for the Old High German prototypical LVC *zi site habēn* 'to do something habitually' no such etymologically related simple verb is attested.

6 For my purposes, it is more convenient to stick to the wider term abstract noun instead of the narrower eventive noun. For the latter, see Fleischhauer (2022, 267). For a general discussion, see Harm (2021, 23–26).

While in the structurally identical LVCs (as (1a)), we could neither ask *where* the wish is going nor can the noun be replaced by *there* or *into a garage*. However, both is possible in (2): *Where is your friend going?* and *My friend is going there/into a garage* are perfectly fine sentences.

The noun in an LVC includes the semantic role mediated by the verb, expresses quantification, definiteness, and carries the meaning of the construction. As Hellan remarks, the noun "avails itself of the light verb to have its predicative content expressed in a sentential pattern" (Hellan, 2023, 47). In this sense, the noun has a richer content then it has by itself. While the verb in an LVC is considered 'light', the noun, in contrast, can be considered 'heavy'. A light verb construction, from this point of view, is also a heavy noun construction.

Very little is known about the role of the preposition in LVCs. It appears only spatial prepositions, i.e., the equivalents of 'in(to)', 'on(to)', 'to', are allowed. Like the verb, the preposition loses its literal meaning. Prototypical LVCs, that is constructions that contain a preposition, are attested from early on in the history of the Germanic languages (Boldt, 2023, 114–120) and are especially frequent in modern Dutch, e.g., *onder controle staan* 'to be controlled', literally 'to stand under control', and German, e.g., *in Beziehung setzen* 'to relate', literally 'to put in relation' (von Polenz, 1987, 171), but are also known in English, e.g., *to come into bloom*, or Norwegian, e.g., *sette i forbindelse* 'to connect', literally 'to put in connection'.

3 The morphosemantic system of the Germanic weak verbs

One of the key properties that sets apart the Germanic languages from the other Indo-European languages is that Germanic languages have developed a distinct threefold verbal system. One group within the threefold verbal system is populated by the strong verbs. This inflectional class mostly contains underived verbs that form their preterite with an ablauting root vowel, cf. PG[7] **beitō* 'I bite' with its preterite **bait* 'I bit'.[8] This type is inherited from Proto-Indo-European but has undergone several changes in Germanic depending on the phonological surrounding of the

7 PG = Proto-Germanic. Other recurring language abbreviations used throughout this paper are: Goth. = Gothic, Lat. = Latin; MHG = Middle High German; OE = Old English; OFris = Old Frisian; OHG = Old High German; OS = Old Saxon; OWN = Old West Norse; PIE = Proto-Indo-European.

8 For reasons of space and readability, I am only giving the reconstructed Proto-Germanic forms here, not the attested forms of the daughter languages, which make these reconstructions possible. The attested forms can easily be looked-up in, e.g., Ringe (2017), and Kroonen (2013).

root vowel (Harðarson, 2017, 931–935; Ringe, 2017, 263–279; Krahe and Meid, 1969, 97–118). Another group is populated by the preterite-presents. This rather limited group of about fifteen verbs formed its present tense like a strong preterite (that developed from the PIE perfect), while its finite past and past participle were weak, cf. PG **wait* 'I know', **wissōn* 'I knew', **wissaz* 'known' (Harðarson, 2017, 939; Ringe, 2017, 290–293; Krahe and Meid, 1969, 135–139). This leaves the last group, which is populated by the weak verbs. The majority of weak verbs are derived from nouns or strong verbs, some from adjectives or adverbs, some are primary, i.d., non-derived verbs. The weak verbs are divided into four classes and are here referred to as (I) *jan*-verbs, (II) *ōn*-verbs, (III) *ēn*-verbs, and (IV) *nan*-verbs depending on which of these suffixes was used to form the respective present stem. Crucially, what sets the weak verbs apart from the strong verbs is that they form their preterite by adding a dental element PG **d* to the originally non-ablauting root[9], cf. the class I verb PG **satjan-* 'to put, make sit' with its preterite form **satidō* 'I put, made sit'. The dental element developed into OWN *ð, t, d*, or *þ* depending on the preceding phoneme and whether the root syllable was light or heavy (Noreen, 1923, §507; Nedoma, 2010, 117–118 for details). The originally non-ablauting root vowel could become subject to secondary sound change and analogical leveling in Old West Norse. That is why we have OWN *setja* 'to put, make sit' (< **satjan-*) and *setta* 'I put, made sit' (< **satidō*) with *i*-umlaut.

Although the Germanic weak verbs are usually divided on formal grounds with regard to the present suffix, it has long been noticed that the formal differences are accompanied by semantic respectively functional differences of the suffix. As far as I can trace back this path of research, Jacobi (1843) was the first one to do so. The most recent overview regarding the semantics that the Germanic weak derived verbs show in relation to their nominal or verbal bases is given by Ringe (2017, 280–290), and thus will serve as the basis for my analysis. In the following sections, I am going to discuss the semantic properties of the individual weak classes, focusing on inconsistencies and ambiguities. These inconsistencies and ambiguities are understood here as symptoms of the beginning of the collapse of the system of weak verbs which in turn is the condition of the possibility of the emergence of light verb constructions.

9 The literature on the origin of the dental element is abundant. For recent overviews with references, see Fulk (2018, 292–294); Ringe (2017, 191–194).

3.1 Class I

The class I verbs are generally known to be causative, factitive or (rarely) intensive-iterative derivations from their individual bases. To stay with the above example: the weak verb PG **satjan-* 'to put, make sit' is a deverbal derivation from the strong verb **sitjan-* 'to sit'. The meaning of this and many other formally identical or similar formations can be paraphrased as 'to cause to happen what is expressed in the base', therefore the term is causative.[10] Very similar to the causatives are the factitives as they differ only in that they are derived from adjectives, cf. PG **daudijan-* 'to kill', a derivation from **dauda-* 'dead'. The paraphrase that can be used to describe how derivation and base relate to each other semantically is the same as for the causatives, that is with regards to our example 'to cause to be dead'. A few Proto-Germanic desubstantival derivations are reconstructable as well. It appears as if they are not so easily definable semantically as a group, but at least some of them are closely related to the causatives and factitives, cf. the PG noun **daili-* 'part' and its class I derivation **dailijan-* 'to divide', which is best understood as 'to make a part', or the PG noun **dōma-* 'judgment' and its derivation **dōmijan-* 'to judge' or 'to make a judgment'.[11] A little different is the case of **arbijan-* 'to inherit', a derivation from **arbija-* 'heritage; inheritance', where factitive semantics ('to make heritage') is noticeable, but not straightforward. Finally, some intensive-iteratives are reconstructable with certainty for Proto-Germanic already, cf. **draibijan-* 'to drive' from the strong verb **dreiban-* 'to drive' and **waljan-* 'to choose' from **wiljan-* 'to want'.[12] What is striking about the intensive-iteratives is that the reconstructions as given in Ringe (2017, 283) seem to differ only slightly in meaning from their derivational bases. That might reflect an artifact resulting form the methods of reconstruction, but it still is conceivable that the differences could be (or become) minimal and eventually contribute to the destabilization of the whole system. It is important here to point out and keep in mind that the earliest reconstructable stages of the system of Germanic weak verbs suggest discrepancies and ambiguities. In addition, there is evidence which suggests that already in Proto-Germanic times there were some fossilized weak verbs from bases no longer in use. These should

10 According to Ringe (2017, 281), more than two dozen causatives are securely reconstructable for Proto-Germanic.

11 Ringe's (2017, 284) statement, "[t]he semantics of those [class I verbs, R.B.] derived from nouns seem to have been governed by the semantics of the noun" is too superficial. Krahe and Meid (1969, 248) use the term *Objektverwirklichung* 'realization of objects' when talking about those verbs.

12 Ringe (2017, 283) takes four examples that he considers "reasonably clear" from Krahe and Meid (1969, 247). The last two are **wagjan-* 'to move' and **wrakjan-* 'to drive (out)'. See García García (2005, 41–42, 63–64) for Goth. *draibjan-* 'to drive; to plague, to harass', which she considers the only real intensive formation in Gothic.

have been at least deverbal PG **tawjan-* 'to make' and PG **wazjan-* 'to clothe, dress', both formally regular causative formations belonging to PIE **deh$_1$u̯-* 'put together, organize' (LIV add. s.v. **deh$_1$u̯-*) and PIE **u̯es-* 'to wear (clothes), to be clothed with' (Rix and Kümmel, 2001, 692–693), respectively.[13] Also, there is denominal **hauzijan-* 'to hear' < PIE **h$_2$ḱ-h$_2$ows-i̯é/ó-*, 'to be sharp-eared'.[14] While the derivation **tawjan-* is attested already in Proto-Norse runic inscriptions, the earliest being the 3SG.PRS form **tawide** on the Garbølle wooden box (Krause, 1971, 148) from around 160–400, the derivational basis is attested nowhere in Germanic. It might be helpful, though, to consider how little difference there is between the reconstructed semantics of the PIE root **deh$_1$u̯-* 'to put together, organize' and the actually attested 'to make'. If this is not an artifact related to the general methods of semantic reconstruction, it is conceivable that there is a genuine instability in the semantic relation between verbs for 'to do; to make' and other verbs that cover a wide semantic range on the one hand and their respective derivations on the other.[15]

In addition, it is possible that the derivational direction cannot be determined satisfactorily. This, among others, appears to be the case for PG **gaumijan-* 'to notice, pay attention to'. While for Ringe (2017, 284), the situation is "difficult to determine", Lühr et al. (1988–2024, IV: s.v. gouma) understand PG **gaumō-* as "wohl postverbal" 'probably postverbal' from **gaumijan-*, and Schuhmann (2024, 110) favors Goth. *gaumjan* to be a denominal formation. Grestenberger and Kastner (2022) have recently argued that questions of directionality are not only a matter of historical linguistics. Rather, they can show "that [even] overt categorizing and derivational morphology is no guarantee for an unambiguous synchronic morphological analysis of directionality (though it certainly helps)" (Grestenberger and Kastner, 2022, 47). That means that even for the speakers of a language it is sometimes not entirely clear what is the basis and what is the derivation. If that happens to a sufficient extent, it might contribute to the confusion of derivational means in general.

For questions that arise from the semantic opposition of the class I *jan-* and the class IV *nan*-verbs, see below.

---

13 Ringe (2017, 282) claims that also PG **sandijana-* 'to send' and **tandijana-* 'to kindle (trans.)' were fossilized formations with bases no longer in use. However, OWN *sinna* 'to travel, drive; pull, follow someone; adhere to, pay attention to, take care of, think about', OHG *sinnan* 'to strive, demand, go, come', OS *-sinnan* 'to recover consciousness, recall to one's mind', and OE *sinnan* 'to care for, mind, heed' point to a PG strong verb **sen(þ)na-*. For a discussion, see Lühr et al. (1988–2024, VII: s.v. sinnan). Also, OE *tinneð* 'ignites' and MHG *zinne* 'I ignite', both attested once, might point to a Germanic strong verb **tinnan-* 'to ignite (intr.)'. See Seebold (1970, 502); Kroonen (2013, 518).

14 A more literal German translation would be *spitze Ohren machen* 'to make pointed ears'.

15 But Kroonen (cf. 2013, 511), who understands the Germanic forms PN *tawjan** and Goth. *taujan* as factitive formations from PG **gatawa-* 'ready, finished' that would be only indirectly attested in Slavic loanwords going back to Proto-Slavic **gotovъ*.

3.2 Class II

The situation of the class II *ōn*-verbs is quite different from that of the class I *jan*-verbs. The oldest members of this group are denominals formed from *ō*-stems, with the case in point being PG **salbōn*- 'to anoint; to provide with ointment', derived from **salbō*- 'ointment, salve'.[16] However, in the PIE period it became possible to form *ōn*-verbs from other stems, as can for example be seen by Goth. *fiskon* 'to fish' and the corresponding Lat. *piscāri*, both derivations from PIE **pisk-o*- respectively **pisk-i*- 'fish' (Krahe and Meid, 1969, 239). The deverbal formations were mostly intensive-iteratives, cf. PG **wlaitōn*- 'to look around', a derivation from the strong verb **wleitan*- 'to look'. The denominal formations, however, formed a much more open group regarding semantics and this is particularly interesting for our purposes. According to Schaefer (1984, 363–383) there were eight types of desubstantival *ōn*-verbs alone, including agentives ('to act like the BS [= base substantive]'), fientives ('BS is taking place, happening'), efficientives ('to make BS [effected object]'), facientives ('to make BS [affected object]'), the *fiskōn*-type ('to catch, to collect BS') emotives ('to feel BS'), ornatives ('to equip someone with BS'), and instrumentatives ('to do something with BS'). Schaefer discusses words that are attested in more than one Old Germanic languages. This does not suggest that all the words mentioned here are of Proto-Germanic origin. It may, however, indicate that the processes of word-formation themselves are Proto-Germanic. As I will argue that the emergence of light verb constructions was connected to the collapse of the system of the Germanic weak verbs, let us keep this apparent diversity in mind, whether one agrees with it in detail or not.[17]

In this class, too, there were a few fossilised forms. As for the verb PG **frijōn*- 'to love', a synchronic derivational base **frija*- 'free' is reconstructable, but the two differ in meaning to such an extent that it seems likely that no speaker would connect them to one another.[18] In addition, there is a verb PG **wratōn*-, that looks like a derivation. However, no base is attested anywhere in the Germanic languages, and so it is likely that it never existed.

16 For an extensive discussion of the etymology, see Lühr et al. (1988–2024, VII: s.v. salbôn).

17 The same is true for Schwerdt (2008, 98), who identifies eighteen types of denominal derivations for the four weak classes. It might be questionable if this is necessary or particularly useful. Ringe (2017, 286) even states that the denominatives could be "formed freely".

18 On the etymology and for further references see Rix and Kümmel (2001, 490) and Schuhmann (2024, 91).

3.3 Class III

For the class III *ēn*-verbs, again, the situation is different from the other classes. The driving forces for the semantic inconsistencies of this class can be seen in the circumstance that the *ēn*-verbs contained mostly deverbal stative verbs and denominal factitive verbs with originally different suffixes that began to merge during the PIE period already (Yakubovich, 2013, 386–391). This group became secondarily productive in Old High German (forming inchoative verbs), but some of the verbs can be reconstructed for Proto-Germanic. By this time, however, they appear to have been a rather small group. I am citing here four Gothic examples that probably had cognates in Old High German, which would suggest that this type of word formation dates back to the Proto-Germanic period: Goth. *ana-*, *gaþiwan* 'to enslave', OHG *dewēn* 'to humiliate', continuing a derivation from the PG noun **þewa-* 'servant, slave'; Goth. *arman* 'to pity', OHG *armēn* 'to become poor; to suffer', continuing a deadjectival derivation from PG **arma-* 'poor, miserable'; Goth. *fastan* 'to hold fast, to maintain', OHG *fastēn* 'to fast, abstain from food', continuing a deadjectival derivation from PG **fasta-* 'fixed' (with a possible semantic narrowing 'to adhere (to the fixed Christian commandments)' > 'to keep (a commandment)' > 'to observe the fasting commandment (as a clearly recognizable sign of the Christian faith)' in OHG (Lühr et al., 1988–2024, III: s.v. fastên)) and Goth. *sweran* 'to honor', OHG *swārēn* 'to be heavy, to become heavy', both continuing a deadjectival derivation from PG **swēra-* 'weighty' (with metaphorical narrowing 'to make weighty' > 'to make (someone's) words weighty' > 'to consider important' > 'to honor' in Gothic, cf. Vedic *gurú-* 'heavy, weighty; honorable person; teacher, master' from a PIE root *$*g^{\underline{u}}reh_2$-* 'heavy', Mayrhofer, 1992, 490). Not only was this group comparatively small, it was also in competition with the deadjectival derivations that formed the class I-verbs of the **daudijan-* : *dauda-*type mentioned above.

Apart from the few factitives, there was the larger group of statives in Proto-Germanic. Some of them maintained a distinct inflection,[19] but most of them merged with the factitives. Again, some of those seem to have become fossilized forms that by the time of Proto-Germanic were no longer synchronically analyzable, e.g., PG **silēn-* 'to be still', the base of which should be an adjective **sila-* 'still, silent'. However, this is not attested anywhere.[20] PG **þulēn-* 'to endure' is a primary verb belonging to the PIE root *$*telh_2$-* 'to pick up, take upon oneself' (Schuhmann, 2024, 301 with references). PG **fijēn-* 'to hate' is another primary verb, belonging to the PIE root *$*peh_1i$-* 'to reprimand, taunt'. Additionally, the verb PG **hatēn-* 'to hate' is, according

19 Ringe (2017, 287), but **sagjan-* 'say' can hardly be understood as a stative.

20 OE *sāl* 'still, silent' (if emended correctly) continues the *o*-grade **saila-*, see Schuhmann (2024, 20).

to Ringe (2017, 288), not obviously derived from the *s*-stem noun **hataz* ∼ **hatiz-*. However, this example seems less clear as a synchronic motivation might have been given.

3.4 Class IV

The class IV *nan*-verbs remained a formally distinct class only in Gothic, while the Old Norse verbs are inflected exactly like the class II *ōn*-verbs.[21] This class has long been understood as containing deverbal inchoatives respectively deadjectival fientives. In recent years, however, some have begun to suggest that these verbs are best understood as being, above all else, anticausatives (Scheungraber, 2014; Ottósson, 2013). Indeed, if we consider the six deverbal examples that are, according to Ringe (2017, 289–290) securely reconstructable for Proto-Germanic, what is striking is that these verbs seem to require a non-agentive subject and cannot be passivized. In PG **libnan-*[22] 'to be left over', **(fra-)luznan-* 'to become lost', **þurznan-* 'to dry out (intr.), wither', **gasturknan-* 'to dry up (intr.), thicken', **waknan-* 'to wake up (intr.)', and **liznan-* 'to learn', the subject is the one to whom something happens, or, in Jacobi's words, "[d]ieses ist dabei nicht thätig [...], sondern leidend" 'it is not active, but permissive' (1843, 195). All of them were made to strong verbs that can also be reconstructed for Proto-Germanic. The same is true for the deadjectivals **$k^{w}ik^{w}nan$-* 'to come to life', which was formed to **$k^{w}ik^{w}a$-* 'alive', and **(ga-)batnan-* 'to get better', which was formed to **batizō-* 'better'.

Additionally, there was a group of strong (!) *nan*-verbs in Proto-Germanic, that could not show the same aspectual nuances as the weak derived *nan*-verbs and therefore might have challenged the otherwise apparently large conformity of this class. PG **spurnan-* 'to kick, spurn', that is reflected as strong verb in OE *spurnan*, *spornan*, OHG *-spurnan*, and (with less certainty) in OWN *sporna*, OFris. *spurna*, OS (*bi-*)*spurnan*, but also shows weak reflexes in OE *-sprynan*, OWN *spyrna*, OHG (-)*spurnen*, is very likely one of those words.[23]

A striking feature that might have contributed to the comparatively large semantic conformity of this class is the fact that they exhibited a strong formal semantic

21 On the pre-history of this class, see Scheungraber (2014, 169–177).

22 The verbs of this class are usually reconstructed as exhibiting the two stem variants **-nō-* ∼ **-na-*, reflecting an older ablauting suffix PIE **-neh$_2$-* ∼ **nh$_2$-*. Since the morphology of these verbs is not of primary importance here, I use the simplified notation **-nan-* to refer to them.

23 For the most recent reflections on this word, see now Lühr et al. (1988–2024, VIII: s.v. firspurnan) as well as Scheungraber (2014, 93–95), and Marti Heinzle (2014, 169–170).

opposition with the *jan*-verbs.[24] Katz recently explains that "a significant number of *-nan* predicates stand beside corresponding verbs of the first weak class in *-jan*, or beside strong verbs. For these, the *-nan* verb comprises half of a grammatical pairing" (Katz, 2021, 87). Indeed, I could find only two possible examples of verbs that might have been fossilized already by the Proto-Germanic period, both of which are only attested in Gothic and the first of which could very well be explained as secondary: PG **dōbnan-*, only attested in Goth. *afdobnan* 'to become dull, numb' (or secondarily nasalized, see Schuhmann, 2024, 68); PG **feinan-*, only attested in Goth. *infeinan* 'to have mercy'. The verb is probably a deadjectival derivation from **fija-* 'reconciled' (?), which is, however, attested nowhere in Germanic (Schuhmann, 2024, 160 with references).

3.5 Summary

The semantic nuances that the derived Proto-Germanic verbs show in relation to their bases are as follows:

- Class I: causative, factitive, intensive-iterative
- Class II: intensive-iterative, agentives, fientives, efficientives, facientives, *fiskōn*-type, emotives, ornatives, instrumentatives (probably because of the detailed study by Schaefer, 1984; a similar study by Schwerdt, 2008 suggests that the other classes were more diversified as well, but since she does not aim for a reconstruction of the Proto-Germanic system, her findings are of rather little value for the purposes discussed here.)
- Class III: stative, factitive
- Class IV: anticausative-inchoative

With that in mind, we can now summarize the semantic inconsistencies that were inherent to the Proto-Germanic system of weak verbs: a) For every class (except probably class IV) it is evident that the suffix carries out multiple functions. While it is probably true that there is a prototypical function, a single class forming feature should not be assumed. b) Some functions are in turn carried out by different suffixes (intensive-iterative: *-jan-* and *-ōn-*; factitive: *-jan-* and *-ēn-*), which means that the corresponding classes (I and II, respectively I and III) were in competition with one another semantically. c) Every class contains some unanalyzable fossils that

24 This opposition led Krämer (1971, 34) to suspect that *-j-* has a causative/factitive function only when it opposes a *nan*-verb or a strong verb. Therefore Goth. *hafjan*, lacking such an opposition, can keep its *-j-* which, according to Krämer, would have been otherwise likely to become subject of analogical removal.

synchronically do not relate to a base but look exactly like the rest of the group, i.e., verbs that were very well analyzable synchronically.

That state of linguistic development shown here indicates that the form of the word forming suffixes is partially detached from the content, i.e., the function of said suffixes. The former are thus partially reduced to mere inflectional markers and so become less capable of expressing semantic nuances.

4 The system of Old West Norse weak verbs

We have seen that already by the pre-historic Proto-Germanic period accessible to us by means of linguistic reconstruction the system of the weak verbs exhibited contradictions and discrepancies, which have led to the weakening of that said system in the daughter languages (except for Gothic). Since I want to argue that this weakening has contributed to the emergence of light verb constructions in Old West Norse, it is reasonable to now shed light on the system of weak verbs in this language and to pay special attention to the visible contradictions and discrepancies as well. In contrast to Gothic and the continental West Germanic languages (Schwerdt, 2008), the word formation of the Old West Norse weak verbs has not yet been subject to systematic investigations that are based on thorough corpus analyses and modern methods. It seems possible, however, to at least outline the main characteristics based on the works of Torp (1974, 38–46); Wessén (1970, 94–113) and a more recent article by Ottósson (2013).

Formally, the Old West Norse system distinguished between three weak classes of verbs, with the previously distinct class IV *nan*-verbs now inflecting exactly like those of class II. In section 4.1, I am going to examine the implications of this change.

4.1 Class I

As in Proto-Germanic, the Old West Norse class I contained causatives, factitives and intensive-iteratives. While the deverbal causatives “constitute a minority of that verb class, [...] the denominative [i.d, factitive; R.B.] *ja*-verbs seem to have been more productive in Old Nordic” (Ottósson, 2013, 367). This can be seen from young formations like OWN *hógværa* ‘to appease, calm down’, from *hógværr* ‘gentle, calm’, and OWN *prýða* ‘to adorn’, from *prúðr* ‘magnificent’, a loan from OE *prūd* ‘proud, arrogant’ (Fulk, 2018, 291).

Let us now return to the verb ‘to put, make sit’, reconstructable as PG **satjan-* (see Sections 3 and 3.1). It can be analyzed as having the *o*-grade of the root (with the

regular development PIE **o* > PG **a*) in contrast to its strong base **sitjan-* that shows the expected *e*-grade (with **e* > **i* as result of early *i*-umlaut). We know from similar cases that the *jan*-suffix in combination with the *o*-grade root forms causatives, factitives, and intensive-iteratives. However, in this particular case (and similar cases)[25] where the base verb continues a pre-Proto-Germanic *-i̯e/o-*present, the context is obscured because both the base **sitjan-* and the derivation **satjan-* show the element **-jan-*.

In Old West Norse, the *-j-* of the suffix causes umlaut, turning the **-a-* of the root into the palatal vowel *-e-*, in this case PG **satjan-* > OWN *setja*.[26] While those verbs might still have been analyzable to some degree, the *-j-* that triggered the umlaut in the first place could vanish in other cases making it unlikely that base and derivation could still be analyzed synchronically: while the *-j-* was retained after light root syllables before *a* and *u* (as in *setja*) and after heavy syllables between *g/k* and *a* or *u* (as in *drekkja* 'to drown, submerge (trans.)' < **drankijan-*), it was omitted in all other cases (Noreen, 1923, §§62–70; Nedoma, 2010, 47). That is why we have a pair like OWN *drjúpa* 'to drip' (< **dreupan-*) with a causative *dreypa* 'to let fall in drops' (< **draupjan-*). The "umlaut [comes] on top of ablaut […] making the phonological distance between base and derivation rather long" (Ottósson, 2013, 365). It is this great phonological distance that helps to obscure the connection between base and derivation. It finally yields a situation where once highly productive means of forming deverbal causatives becomes less applicable.

It is promising to understand other forms that were not analyzable synchronically in light of this context as well. Some of them were fossils like OWN *leifa* 'to leave behind', the causative of a strong verb that would have continued PG **leiban-* 'to stay'. However, these forms are not attested in OWN (and neither are they attested anywhere else in Germanic, but see the prefixed form PG **bileiban-* 'to stay' giving Goth. *bileiban*, OE *belīfan*, OHG *bilīban*). Another example might be OWN *beygja* 'to bend (trans.)', the causative of a strong verb that would have continued PG **beugan-* 'to bend (intr.)', which in its regular form is not attested on OWN but, for instance, in Goth. *biugan*, and OHG *biogan*. There is a strong form OWN *bugu* and a participle *boginn* 'curved, crooked, hooked, circular' (as well as Old Swedish *būgha* 'id.'), but they continue an unexpected root vowel **-ū-*, making it questionable as to whether speakers would associate the historically connected forms to each other.

Although most the *jan*-verbs were derived, some were primary and so had no synchronic base, cf. *søkja* 'to seek, fetch', a cognate with Goth. *sokjan*, OE *sēcan*, OHG

25 **ligjan-* 'to lie', **bidjan-* 'to ask for'. The *-i̯e/o-*presents were also common in the sixth strong verbal class with the present root vowel **a*, e.g., **swarjan-* 'to swear', **skapjan-* 'to make, fashion'. See Ringe (2017, 276).

26 But see the non-umlauting forms Goth. *satjan* : *satida* and Proto-Norse **satjan* : **satido** (Rö stone).

suohhen continuing PG **sōkijan-*, in turn a continuation from a *-i̯e/o-*present from the root PIE **seh$_2$g-* (Rix and Kümmel, 2001, 520; Kümmel, 2023, 71; Schuhmann, 2024, 260–261).

Some *jan*-verbs, in turn, were not analyzable as derived because the semantic difference between them and their respective base was too little, cf. OWN *þreifa* 'to touch, feel with the hand', historically a derivation from *þrífa* 'to grip, grasp, take hold of'. In addition, we have *hverfa* 'to (over-)turn' from PG **hwarbijan-* that at some point must have fallen together with its strong base **hwerban-*, yielding an isolated Old West Norse verb without base.

While we do not have detailed studies on the intensive-iterative group of the *jan*-verbs in Old West Norse, we know that at least some of them contributed to the weakening of the verbal system as the formal relation between base and derivation had become less apparent, cf. OWN *nema* 'to take' and the somewhat obscured *jan*-derivation *nœma* 'to deprive of'.

4.2 Class II

Class II is the biggest class in Old West Norse. The situation for the verbs of this class is quite different from that of class I. Perpetuating what we have seen for Proto-Germanic, the denominal verbs that belong to this class can show a great variety of derivational semantics in relation to their bases. Two examples may suffice here: 1) From the abstract noun OWN *afl* 'physical or mental strength, power' a verb *afla* 'to beget, arrange, achieve' was formed. The meaning of this verb can be generalized as 'to use what is mentioned in the base as an instrument'. It is therefore part of the group of instrumentals. 2) From the concrete noun OWN *dúkr* 'scarf, cloth, towel', a verb *dúka* 'to provide with a dúkr' was formed. The meaning of this verb can be generalized as 'to provide with what is mentioned in the base'. It is therefore part of the group of ornatives.

In contrast, the semantic variety of the deverbal derivations is much more limited in that they were mostly intensive-iteratives (with geminated voiceless root consonant, perhaps expressive in nature), cf. the OWN strong verb *ljúga* 'to lie' and the weak verb *lokka* 'to lure, entice' (< 'seduce' < 'to lie repeatedly'; Wissmann, 1932, 176; different Lühr et al., 1988–2024, V: s.v. liogan; Lühr, 1988, 348).

Old West Norse had some primary *ōn*-verbs, e.g., OWN *hrata* 'to crash, fall down', a descendant from PIE **(s)k/ḱer-* 'to jump, swing' (Rix and Kümmel, 2001, 556, 793) but "[e]lles er heile den uhorvelege mengdi av *ô*-verb avleidd" 'apart from that, most of the *ô*-verbs (= *ōn*-verbs, R.B.) are derived' (Torp, 1974, 38; similar Wessén, 1970, 95 "der weitaus größere Teil der *ō*-Verben (= *ōn*-Verben, R.B.) […] besteht indessen aus Ableitungen" 'the vast majority of ō-verbs (= ōn-verbs, R.B.), however, consist

of derivatives'). For a lot of those derived verbs, however, the base is not attested synchronically in the Old West Norse texts, e.g., *skoða* 'to view, look after'[27] or *mjólka* 'to milk'[28].

The former anticausative-inchoative *nan*-verbs of the fourth class gave up their existence as a morphologically distinct class in Old West Norse and inflect exactly like the old *ōn*-verbs. This can be seen in Table 1, which contains the present inflection of the former *nan*-verb *sofna* 'to fall asleep' and the old *ōn*-verb *kalla* 'to call, name':

Tab. 1: Present inflection of *sofna* and *kalla*.

SG	*sofna*	*kalla*
	sofnar	*kallar*
	sofnar	*kallar*
PL	*sofnum*	*kǫllum*
	sofnið	*kallið*
	sofna	*kalla*

Oddly enough, the former *nan*-verbs became secondarily productive in Old West Norse (Boldt, forthcoming; Torp, 1974, 41; Wessén, 1970, 100–110). While the early formal development of this class is not completely understood[29], synchronically it leads to class II-verbs being even more heterogeneous semantically as it adds weak verbs with anticausative-inchoative semantics to the already existing diversity. What can be seen in the case of *sofna* from the above table is that we have a non-agentive subject to which something is happening (and which therefore cannot be passivized). The same holds true for other former *nan*-verbs like *skriðna* 'to glide' with a base *skríða* 'to creep, crawl' or *rifna* 'to be rent, split' with a base *rífa* 'to tear'.

The majority of the *nan*-verbs were associated secondarily to the past participle of the strong verbs (*brotna* 'to break in two, break in pieces, crumble, collapse' to *brotinn* 'broken (off), fractured'; *bogna* 'to become bent, bend, crouch, buckle, become crooked' to *boginn* 'curved, crooked, hooked, circular'). However, there were some verbs with no such association and no attested synchronic base, e.g., *togna* 'to be stretched', the continuation of a derivation from PG **teuhan-* 'to pull', as attested in

27 With unclear etymology and derivational base; perhaps related to OE *sceāwian* 'to look, observe', OHG *skouwōn* 'to look, observe'. For those, see now Lühr et al. (1988–2024, VIII: s.v. skouwôn).

28 A deverbal derivation from strong PG **melkan-* 'to milk' which is attested in OE *melcan*, *meolcan*, OHG *melkan*, OS *melkan*; but probably with secondary association to *mjolk* 'milk'. See Lühr et al. (1988–2024, VI: s.v. melkan).

29 For the earliest tangible functions of the nasal verbal formation in Proto-Indo-European, see Scheungraber (2014, 182–183).

OE *teōn* and OHG *ziohan*. In the case of *staðna* ‘to stop, pause’, speakers did probably associate the verb with its strong base *standa* ‘to stand, stay’, but the formal as well as the semantic side must have been somewhat obscure.

What is remarkable about the *nan*-verbs is that they had a strong connection to the *jan*-verbs in that some of them formed an intransitive-transitive and anticausative-causative-opposition. The opposition does not seem to be as systematic as in Gothic (Ottósson, 2013, 335), but some *nan*-verbs are best understood as derived from weak *jan*-verbs rather than from their strong counterparts. For example, in the triad *sofa* ‘to sleep’ : *svejfa* ‘to lull to sleep, assuage, soothe’ : *sofna* ‘to fall asleep’, the derived verbs show the expected causative respectively anticausative semantics with regard to their common base. However, the situation is different for the triad *drekka* ‘to drink’ : *drekkja* ‘to submerge, sink, push below the surface’ : *drukkna* ‘to drown’. In Gothic the derived *jan*-verb *dragkjan* ‘to make drink’ is a straightforward causative formation to *drigkan* ‘to drink’, whereas the OWN derivations *drekkja* and *drukkna* are more closely connected to each other than to *drekka*. It seems as if *drekkja* has been reinterpreted from the *nan*-verb, while the old meaning ‘to make drink’ has been adopted by the verb *brynna* in the course of this process.

4.3 Class III

Finally, with regards to the *ēn*-verbs of the third class, the following general characteristics, contradictions, and inconsistencies can be observed:

In one of the very few recent works on Old West Norse verbal word formation, Ottósson (2013, 355) states that “[t]he stem suffix *-ē* - tended to form durative verbs, often stative verbs, or intransitive non-agentive verbs”, which seems especially true if there is an “Oppositionspartner” ‘opposition partner’ from another class (Krämer, 1971, 37). This, however, should be taken cum grano salis in light of the remarkable observation that only very little of the material that is contained in Torp (1974, 38) and Wessén (1970, 95) does actually suggest a synchronic relationship of a base and a weak *ēn*-verb. One of these few cases is OWN *trúa* ‘to believe, trust’, which historically continues a deadjectival stative formation derived from PG **truwa-* ‘loyal, faithful’ (Harðarson, 2018, 230, n. 11 with references), but synchronically was associated not only with the adjective OWN *trúr* ‘loyal, faithful’, but also with the noun *trú*(*a*) ‘loyalty, honesty’.

Most of the other verbs examined by Torp and Wessén only had strong counterparts as a derivational base in Gothic (OWN *lifa* ‘to be left; to live’ : Goth. *bileiban* ‘to remain, stay’) or were related to another OWN weak verb (*loka* ‘to lock, shut’ : *lokna* ‘to come to an end’).

Also, when compared to class II verbs, a lot of the verbs belonging to class III were primary and therefore could not be associated with a derivational base synchronically. Two instructive examples are the following: 1) OWN *duga* 'to make an effort, do one's best', either from PG **daug-*/*dug-*, a reformation of the stative present **d^hug^h-é(i)* with *o*-grade root after the preterite-presents, or straight from PG **dug-*, continuing the stative present PIE **d^hug^h-* from the root **d^heug^h-* 'to meet' (Rix and Kümmel, 2001, 148–149, fn. 4), and 2) OWN *gana* 'to yawn, gawp', continuing PIE **$g/ǵ^hn̥$-$h_1$ie-* from the root **$g/ǵ^han$-* 'to yawn' (Rix and Kümmel, 2001, 193) with an OWN back formation *gan* 'yawning, screaming'.

In addition, we have verbs like OWN *ljá* 'to lend' and *tjá* 'to show; to tell' (older *léa* and *téa*) that once were strong (< **leihwan-* and **teihan-*) but became weak and, accordingly, did not have a synchronic base. The paradigm of *ljá* was further obscured by the remnants of the strong form *lé* (1SG) and *léner* (PPNPLM) (Noreen, 1923, §483). Finally, a weak *jan*-verb OWN *leiga* 'to hire, rent' is attested, continuing PG **laigijan-* 'to make lend', a causative derivation from **leihwan-* which could not be analyzed synchronically because the phonological distance between *ljá* and *leiga* was too long.

Additionally, there were class III verbs with little or no difference in semantics compared to their strong base. The highly frequent verb OWN *hafa* 'to have, hold, use' was weak and thus was distinguished from its strong base *hefja* 'to lift, begin'. However, as Cleasby & Vigfússon have observed, "in sundry cases [...] it [*hafa*, R.B.] passes into the sense of this latter word; as also in some instances into that of another lost strong verb, *hafa*, *hóf* 'to behave', and *hœfa* 'to hit"' (Cleasby and Vigfússon, 1874, s.v. hafa).

4.4 Summary

All weak verbal classes in Old West Norse show general developments that point to a further weakening of the entire system. These developments are: 1) fossilisation through a) the loss of a synchronic base, b) phonological and morphological developments that lead to the complete coalescence of the formerly different classes II and IV as well as the coalescence of individual bases and derivations, and c) strong verbs becoming weak[30]; and 2) little or no semantic differences between base and derivation. For each individual class, the following peculiarities are to be mentioned:

30 There were some cases of labile verbs in OWN. These are verbs that vacillate between a transitive and an intransitive reading. They usually denote movement, for example OWN *renna* 'to run; to make run' and *koma* 'to come; to bring'. Nonetheless, because they are quite rare, I don't take them into account here.

The *jan*-class contained causatives, factitives and intensive-iteratives. The relations between base and derivation became obscured due to *i*-umlaut and the partial loss of the factor that triggered it. These developments resulted in a phonological distance between base and derivation in a way that the letter appeared as fossilised form, since it could not be associated with the former. Nonetheless, the class was still productive to a certain extent.

The *ōn*-class continued the semantic variety of the Proto-Germanic denominal formations. Deverbal formations, on the other hand, were intensive-iteratives, which resulted in partial competition with the *jan*-verbs. This class contained some primary formations but most of the *ōn*-verbs were derived. A lot of the bases, however, are not actually attested in the Old West Norse texts, rendering the derivations unanalyzable fossils. The semantic variety of the class continued to increase as the distinct inflection of the former anticausative-inchoative *nan*-verbs was given up and they assimilated to the old *ōn*-verbs. These *nan*-verbs, many of which were Nordic innovations, were often in secondary opposition to the *jan*-verbs. Moreover, the *jan*- and the *nan*-verbs were semantically closer to each other than to their base.

It might be speculated here that there could have been a contrasting development to the (secondary) productivity of the former *nan*-verbs. In the class II verb *bákna* 'to wave, give a sign', the *-n-*, which originally belonged to the stem of the derivational base *bákn* 'sign' (in turn a loan from pre-Old Frisian **bāken*, see Maini, 2017), might have been interpreted as part of the suffix, disintegrating its close connection to the anticausative-inchoative semantics.

Most of the *ēn*-verbs were intransitive stative verbs. However, what set them apart from the other classes was that a good part of them could not be associated to a base synchronically. It therefore appears that, in contrast to the other classes, the class II denominal verbs showed no prototypical derivational semantics. The class is hence best understood as default weak class in which new denominal verbs could be integrated regardless of the semantic relation they may have had to their bases. This includes the former class IV *nan*-verbs, as well as verbs from other classes, cf. *fá* 'to adorn, decorate' continuing Proto-Norse *faihijan**, a class I primary verb meaning 'to color, paint (runes)'. Class II was highly productive and so was its aspectual variety.

As can be seen by the secondary productivity of the *nan*-verbs, the fact that class II became the default class for new verbs does not mean that the semantic system of the verbs had altogether collapsed in Old West Norse. The fact, however, that new verbs could be integrated in class II with little regard to their semantic features points to a situation where the semantic system was weakened or where semantic nuances, that were once expressed by different verbal suffixes could possibly be expressed by other means.

I am focussing here on the semantic side of the weak verbs, but it is helpful to look at the formal side as well in order to better understand their development. This

can be seen not only on the basis of the *nan*-verbs becoming *ōn*-verbs, but also on the basis of the rest of the verbal paradigm. If we compare the endings of the present paradigm of the three weak verbal classes in OWN in Table 2[31], we can see that their coalescence is much more developed than, for instance, in OHG, where *-ō-* and *-ē* - as the stem forming suffixes are carried out consequently throughout the paradigm of the second and third class.

Tab. 2: OWN endings of the three weak classes.

	I	II	III
SG	*-0*	*-a*	*-i*
	-r	*-ar*	*-ir*
	-r	*-ar*	*-ir*
PL	*-um*	*-um*	*-jum*
	-iþ, *-it*	*-iþ*	*-iþ*
	-a	*-a*	*-a*

Formal coalescence and overlapping entails the loss of the possibility of associating form and function, and subsequently the loss of the possibility of expressing semantic nuances in the traditional way.

5 Old West Norse LVCs

When semantic categories that were once sufficiently formally separated give up their corresponding formal separation to a certain degree for whatever reason, they have to find new, sufficiently unambiguous means to be expressed. Having considered discrepancies and inconsistencies in the semantic system of the Old West Norse weak verbs, and partially coalesced verbal endings, it is plausible to assume that the semantic categories of causality and aspect once expressed by the weak verbs had to find new, sufficiently unambiguous means to be expressed. The observation that the LVCs express some of those semantic categories has led research to assume that there is a connection between both phenomena in that the first have taken over expressing those categories from the latter. In the following, I am going to assess the viability of this assumption with regards to the development in Old West Norse.

31 The endings are taken from Nedoma (2010, 110). It can be seen there that the coalescence is even more developed in the subjunctive, imperative, as well as the indicative preterite paradigms.

In order to do that, I am going to consider four types of LVCs from the Old West Norse texts that express semantic nuances that are known to have been expressed by derived weak verbs. Those types are: causative, stative/durative, inchoative, and intensive light verb constructions. What is important to keep in mind is that those types are not just any types amongst many others, but the only types of light verb constructions that are attested in Old West Norse, at least to my knowledge.

5.1 Causative light verb constructions

The first type of LVCs in OWN that I want to present here with regards to the semantics of the whole construction is the causative type. It is represented by the expression *setja í bann* 'excommunicate', which according to ONP is attested at least six times in the Old West Norse prose literature (ONP, 2005, s.v. bann 2) setja í bann)[32] and at least twice in Old East Norse (Boldt, 2023, 150–151). The sentence in (3) is taken from the Old West Norse version of the Statuta Vilhjalms kardinála (ONP, 2005, s.v. bann 2):

(3) *Item setium ver alla i bann er nunnur vilia taka naudgar.*
also put we all in ban who nuns want take forced
'We also excommunicate all those who want to force nuns to have intercourse.'

The text is a translation from a Latin original and the words *setjum ver …i bann* translate Lat. *excommunicavimus* 'we have excommunicated' (!) which means that the OWN light verb construction is not taken over from the source but seems to draw from a pattern that is genuinely Old West Norse. According to ONP (2005, s.v. bann), the abstract noun *bann* can convey three meanings in free use: (1) 'prohibition, ban'; (2) 'state of excommunication, sentence of excommunication'; (3) 'swearing, cursing'. As is typical for LVCs, the noun keeps its literal meaning (in this case (2)) within the construction.

In contrast to the noun, both the verb and the preposition lose their literal meaning within the construction. Focusing on the semantics of the LVC, it should be noted that the verb is not entirely stripped of meaning but rather determines the causative semantics of the entire formation: the bringing about of the excommunication.[33]

32 Formations like OWN *falla í bann* or *renna í bann* (cf. ONP, 2005, s.v. bann), seem to contradict what has been said under 1., namely that only basic verbs can be part of modern LVCs.

33 In Boldt (2023, 155), I described this light verb construction as ingressive-inchoative, which I now believe to be incorrect as *setja* primarily stresses the bringing about of the situation, not the

In modern LVCs, verb and preposition can usually not be replaced without rendering the sentence ungrammatical (see (1a) and (1b)). It seems this was the case with historical LVCs as well. This differentiates them from free word combinations where a replacement is possible, as can be seen in the following sequence (4) which is taken from Fix et al. (2006, 528):

(4) *setja sverðs eggina í ana*
put swords tip into water
'to put the tip of the sword into the water'

In this word group, *í* and *setja* (although the verb is used metaphorically) could be replaced by the OWN equivalents of 'above' and 'put' or any other preposition and verb from the same semantic field (i.e., a spatial preposition and a verb of movement/transfer) without rendering the sentence ungrammatical.[34] This is only possible because in (4), the noun retains its literal, concrete meaning.

In example (5), taken from Gulathing Law IV 102 (ONP, 2005, s.v. fjǫturr, fjaturr 2), there is another option regarding constructions that superficially look like LVCs:

(5) [...], *þa scal hann setia hann i fiotur*.
then should he put him in shackles
'[...] then he should put him in shackles.'

In this example and just like in prototypical LVCs, the preposition and the verb seem to be replaceable to a very low extent.[35] In contrast to prototypical LVCs, however, the noun is not an abstract, but a concrete noun. This entails that the verb, although used metaphorically, cannot function as a light verb denoting mainly aspectual nuances, but instead functions as a full lexical verb.

What is interesting and crucial for determining the state of development of the LVCs in Old West Norse is the observation that the noun as seen in (5) seems to allow for alternatives in similar contexts, see (6) from the Vǫlsunga saga 5:

(6) *Þess vil ek bidia þik, at* [...] *latit þa helldr settia i stock*
this want I ask of you that let them rather put in log of wood
'What I want to ask of you is for you to rather put them in a log of wood.'

beginning of the situation. One could also think about understanding the construction as ornative in the sense of 'to assign/place the ban onto someone' or 'to equip someone with a ban', but that is not something I can discuss in detail here. On the semantics of the ornative verbs, see Kuroda (2017).

34 See Fleischhauer (2022, 264–265) for a similar approach with regards to German *stehen unter*-LVCs.

35 From 11 instances that are listed in (ONP, 2005, s.v. *fjǫturr*, *fjaturr*; *setja*/*...í fjǫtur*), 9 show the verb *setja*.

Moreover, in this example, the verb cannot function as a light verb. Instead, the verb functions as a metaphorical full verb, since the literal bringing about of a concrete object such as a log of wood (and a shackle in (5)) is impossible.

Additionally, in the LVC *setja í bann*, the concrete, spatial meaning of *bann* as a place where someone could be sent is still tangible. Therefore, the literal meaning of the verb *setja* is still tangible as well, making it less light than, for example, in the Modern Norwegian LVC *å sette i forbindelse* 'to connect', literally 'to put in connection'.

5.2 Stative/durative light verb constructions

The second type of light verb constructions presented here is the stative/durative type. The expression *sitja í arfi* 'to be in possession of an inheritance', attested at least three times in the Old West Norse prose according to ONP (2005, s.v. arfr, sitja í arfi), may serve as an instructive example. The sentence in (7) is taken from Frostathing Law IX 27 (ONP, 2005, s.v. arfr):

(7) *Nú callar sá þann eigi arfa réttan er í arf sitr*
now asserts he him not heir right is in inheritance sits
'He [a minor] now asserts that the one who is in possession of the inheritance is not the right heir.'

According to ONP (2005, s.v. arfr), the noun *arfr* bears the meaning 'inheritance, what is (to be) inherited, inheriting, right to inherit'. One can observed in example (3) *setja í bann* that the literal, spatial meaning of the noun *bann* as a concrete location where someone could be sent during the process of banning, is still tangible. The same holds true for example (7), but to an even greater extent. While it is conceivable that someone is physically located in his or her inheritance (e.g., is sitting in an inherited house), this is most likely not what is meant here. Instead, the usage of *sitja í arfi* expresses that someone is in permanent possession of an inheritance. *Arfr* does not refer to a specific inherited item and is therefore an abstract noun. *Sitja* is not a full verb denoting the absence of movement, but a semantically reduced light verb that is denoting semantic nuances that highlight the stative/durative aspect of the LVC compared to the simplex *erfa* which is unspecific in that respect.[36]

36 OWN *erfa* historically is a *jan*-verb continuing PG **arbijan*-. The semantics of this type of *jan*-verbs aims for the realization of the object with regards to the base. See above fn. 11. Cf. OWN *dómr* 'judgement, ruling, decision' : *dǿma* 'to judge, rule, past a sentence' (i.d. 'to realize a judgment etc.').

This entails that *sitja í arfi* has a) a literal reading consisting of a predicate *sitja* plus a dative object *í arfi*, and b) a reading as one complex predicate structure. Only that latter reading qualifies it a as light verb construction.

A very similar, yet slightly different situation can be observed in the examples (8) and (9), taken from Óláfs saga helga and Bøjarlǫg (ONP, 2005, s.v. vǫrðr). They both contain the expression *sitja á verði*, which I translate literally first:

(8) *Enn er þeir sǫto a verþinom þa heyrdu þeir grat oc veinon*
and as they sat on watch-out they heard their weeping and wail
micla, […]
intensive
'And as they sat on their watch-out, they heard their weeping and intensive wail.'

(9) *þa skulu bœar men a værði sitia hina fimtu not siðan vorðr er*
then shall citizens on watch-out sit the fifth night since watch is
boðen
ordered
'Citizens shall sit on the watch-out for five nights after the watch has been ordered.'

In (8) and (9), the expression *sitja á verði* shows the meaning 'to keep watch'. However, in contrast to (7), *sitja í arfi* 'to be in possession of an inheritance', keeping watch does require an immediate physical presence. That in turn makes the verb less light than in sentence (7). The verb covers the stative/durative aspect of being present at the watch, but it still has the metaphorical qualities of a full verb. It is contexts like these where the noun vacillates between its literal and its metonymic meaning that may be understood as hinge points in the emergence of light verb constructions[37] under the condition that simple weak verbs can no longer sufficiently be used to express aspectual nuances.

5.3 Inchoative light verb constructions

The third type of light verb constructions is the inchoative type. The expression *koma til enda* 'to come to an end'[38] does not have a separate entry in ONP (2005) but is

37 Similar now also Fleischhauer and Hartmann (2023, 131), who state that "einzelne konkrete Muster […] als Ausgangspunkte […] fungieren" 'individual concrete patterns serve as starting points'.
38 In Boldt (2023, 308), I cited *koma fyrir enda*, which is perhaps best paraphrased as 'to come to the beginning of the end'.

interesting for our purposes. The following sentence (10) is taken from Karlamagnúss saga, 7 (ONP, 2005, s.v. endir):

(10) *mal mitt er sua til enda komit at ek skal nu vera j fylgd med*
speech my is thus to end come that I must now be in service with
godum monnum.
good men
'My speech has thus come to an end, as I must now serve with good men.'

As in the previous examples, the noun *endi* 'end' is an abstract noun. However, the concrete, spatial meaning is still clearly present. This is further reinforced by two observations, 1) the fact that, according to ONP (2005), *endi* can also be a concrete noun as it can refer to 'tip, arm, branch', and 2) the possibility that *mál mitt* 'my speech' in (10) refers to a written document of some kind, the reading of which has ended.

The usage of *koma*, however, does not entail the physical act of approaching a specific place (as *mál mitt* can not do such a thing). Rather, it denotes the inchoative aspect of *koma til enda*, which is therefore best understood as light verb construction.

How precarious the state of that word combination can be becomes apparent when the noun is part of a compound, as it is in (11). The passage is taken from the Prose Edda, 30 (ONP, 2005, s.v. koma A. 7):

(11) *Þa taka þeir en skeið, en er Hvgi er kominn til skeiðsenda ok*
then take they a race and as Hugi is come to end.of.run/track and
snyzt aptr [...]
turned back
'They then begin another race. And as Hugin has come to the end of the run / the track and turned around, [...]'

In this text passage, the reading vacillates. The raven Hugin may have reached the end of the running track, or the end of the running. In the first case, *koma* would be a full lexical verb denoting the physical motion of approaching the end of a concrete place. In the latter case, *koma* would denote the beginning of the end of the running and would therefore rather qualify as light verb. It is, again, ambiguous contexts like the given one that may have played a crucial role in the rise of light verb constructions.

In order to illustrate the differences of a light verb construction to an idiom[39], an expression like *koma í ljós* 'to be born' (literally 'to come into light') can be cited here. It is taken from Grágás 118 (ONP, 2005, s.v. ljós I. 6), and given in (12):

39 A semantically opaque fixed expression.

(12) *Ef sa maðr andaz er barn á i vánum þa er barn eigi*
if the man dies who child has in expectation then is child not
arfgengt nema lifanda komi i lios [...]
entitled.to.inherit unless living comes into light
'If the man dies who is awaiting a child, then the child is not entitled to inherit unless it is born alive [...]'

ONP (2005) lists seven instances of *koma í ljós* within the corpus of Old West Norse prose. Superficially, its structure looks similar to one of the LVCs as it consists of a verb of movement, a preposition of direction, and a noun. The noun *ljós*, however, is not an abstract noun but a metaphorically used concrete noun. In free use, its usual meaning is '(sun-)light; radiance, shine; public', while 'life (on earth)' seems rather marginally attested. It mainly bears the latter meaning in two idioms: *í/ór þvísa ljósi* 'in/from this life' and *koma í ljósa* 'to be born'.

Since *ljós* is a metaphorically used concrete noun and the verb *koma* cannot take the step from being a full verb to being a semantically reduced light verb, the whole expression does not qualify as an LVC. However, it is plausible that LVCs, idioms, and free word combinations that share the same or a very similar structure may have influenced each other and that problems in keeping them apart are not due to inadequate scientific reasoning but mirror linguistic reality.

5.4 Intensive light verb constructions and the role of OWN *gera* 'to do; to make'

It is striking that the LVCs discussed so far cover the semantic range of the classes I, III, and IV of the Germanic weak verbs–but not that of class II. The latter, as we have seen, inflect exactly like those of the former class IV, and we would not expect this finding to have purely formal reasons. It should be remembered that, while the deverbal derivations are mostly intensive-iteratives, the denominal derivations of class II verbs show a high degree of semantic variety. It is this variety and accompanying lack of semantic clarity that prevents corresponding LVCs from emerging. Another reason is the lack of suitable verbs. For instance, it is conceivable that the verb *come* can become desemanticised and so lose its literal meaning of movement towards a goal in order to merely denote an abstract meaning of inchoation. However, no verb of movement or transfer can become desemanticised and lose its literal meaning of movement or transfer in a way that would qualify it for taking over emotive or ornative semantics. These, among other features, are typical characteristics of class II-verbs.

There might, however, be some constructions with intensive semantics that are comparable to LVCs. Relleke (1974, 35) quotes OHG *anaruafti tuon* 'to call for help, beg' and *in feste tuon* 'to reinforce, confirm', suggesting that these constructions can give an event a higher degree of intensity than the corresponding simplicia. OHG *habēn* 'to have', *lāzan* 'to let, allow', and *sīn* 'to be' are rarely used in similar environments. However, in most cases (as can be seen in the two examples above), the verb in those constructions is *tuon* 'to do'. No cognate of *tuon* is attested in North Germanic. Instead it was replaced very early by a verb that surfaces in Old West Norse as *gera*.[40] A perusal of the entries given in ONP (2005, s.v. gera), however, suggests that there are no prototypical LVCs with *gera* (+ abstract noun + preposition) attested in Old West Norse. There are indeed similar constructions like *gera skaða* 'to do harm' or *gera verk* 'commit an act, a crime'[41], but it is questionable if those qualify as light verb constructions for three reasons: a) they can hardly be understood as complex predicates simply because, as in contrast to LVCs with a preposition of direction, the noun can be asked for; b) the verb *gera* is not desemanticised to the degree that it conveys mere abstract aspectual semantics, but rather functions as a full verb conveying a general 'bringing about, producing' of the noun. It likely does not primarily cover intensive semantics in relation to its simplicia; c) words for 'to do; to make' do not usually surface in prototypical light verb constructions in other Old Germanic languages.[42]

6 Summary, conclusive observations, and open questions

This article has demonstrated a number of features that accord with the emergence of light verb constructions. First, the reconstructable system of weak verbs in Proto-Germanic already exhibits discrepancies and inconsistencies in both form and content that point to a dynamic linguistic situation, i.e., to a system that is undergoing restructuring. Second, it has shown that these discrepancies and inconsistencies not only persist in Old West Norse but have intensified. The semantic coherence of the system of verbal classes has further decreased, as has the possibility of keep-

40 For the etymology, see Lühr (2000, 226).

41 For an overview of those constructions in Old West Norse legal texts, see Boldt (2023, 330). For Old Swedish, see Sundquist (2018, 2020). Also, compare already Proto-Norse *tawo laþodu* 'I make/prepare the invitation' on the Trollhättan bracteate, see Schulte (2023, 7), and Schuhmann (2016, 449–450).

42 The only case that is known to me is Old Frisian *tō bonne dwā(n)* 'to ban'. For a short discussion of verbs for 'to do; to make' as light verbs in the Old Germanic languages, see Boldt (2023, 132–136).

ing the classes apart formally. Although the causative-factitive *jan*-class was to a certain extent still productive, and the anticausative-inchoative *nan*-class became highly productive, the durative *ēn*-verbs in particular appear to have already been fossilized to a large extent, while numerous new formations were entering the class of *ōn*-verbs as the default class.

The instability of the system of weak verbs entails that the semantic nuances expressed in it must look for new means of expression. Against the background of what has been shown here, it seems likely that they find those new means in the light verb constructions. In Old West Norse, these have at least causative, durative and inchoative semantic properties and thus cover part of the semantic range that the derived weak verbs have shown in comparison to their bases. However, the observation that the semantic diversity of the denominal verbs of the second class has no equivalent in the light verb constructions remains striking and in need of explanation. This also means that no straightforward correlation can be postulated between the weakening of the system of weak verbs and the emergence of LVCs.

What characterises the early Old West Norse LVCs is that the concrete spatial dimension of the abstract noun is often still tangible. It is likely that the development originated here and was gradually extended to more abstract contexts. It is further possible that the emergence of LVCs goes back to Proto-Germanic. This is supported by the fact that an LVC **tō andijai kwemanan/kumanan* 'to come to an end' can be reconstructed, since the cognates OE *tō ende cuman*, OHG *zi ente queman*, OFris. *tō ende koma* and (with substitution of the preposition) OWN *koma til enda* are attested in the daughter languages (Boldt, 2023, 308). Prototypical LVCs wherein the spatial component of the noun is still tangible are also found outside of Germanic, cf. Lat. *in suffragium mittere* 'to let vote' (literally 'to send for voting'; Relleke, 1974, 13). The extent to which old or more recent, possibly areal linguistic patterns play a role here would also have to be examined.

Further questions might be raised. Why are there no LVCs that correspond to the semantically diverse denominal verbs of class II? We have seen that prototypical LVCs contain a verb of movement or transfer. Are those not suitable for that purpose? Another idea could be that denominal class II verbs were already formally weakened and semantically diversified to a degree that by the time LVCs emerged, they were not sufficiently unambiguously associated with certain semantic properties, which then could have been taken over by LVCs. Furthermore, how does the fact that at least the causative *jan*- and the former anticausative-inchoative *nan*-verbs were still partially productive (*jan*-verbs) or became highly productive (*nan*-verbs) relate to the fact that there were causative and inchoative LVCs? Is that view an artifact because we have too little data, or did both formations exist side by side?

Related to those is also the question how analytical LVCs like *koma til enda* 'to come to an end' relate synchronically to synthetic forms like the anticausative-

inchoative *setna* ‘sink (in); be digested’ but also ‘to come to an end, cease’, a derivation from *sitja* ‘to sit’.

Having presented some ideas of a rather exploratory nature, this study contributed to the understanding of the emergence of light verb construction Old West Norse. It further shows that a comprehensive study of the Old West Norse system of weak verbs and a systematic analysis of the character of light verb constructions in Old West Norse and the other Old Germanic languages is necessary to deepen this understanding.

Acknowledgments

For polishing my English and for valuable comments, I owe thanks to Charles Comerford. I would also like to thank the reviewers and editors for their comments.

Bibliography

Blum, Siegfried. 1986. Ahd. *habên* in Funktionsverbgefügen. *Beiträge zur Erforschung der deutschen Sprache und Literatur* 6. 80–95.

Boldt, Ramón. 2023. *Aspekte der Vergleichenden Phraseologie. Untersucht und dargestellt am Beispiel der ältesten germanischen Rechtstexte*. Dettelsbach: J.H. Röll Verlag.

Boldt, Ramón. forthcoming. *To be or to become, that is the question. Word formation of the weak verbs formed from color adjectives in Old West Norse.* In Laura Grestenberger, Viktoria Reiter and Melanie Malzahn (eds.), *Deadjectival verb formation in Indo-European and beyond*, Berlin: Language Science Press.

Brinton, Laurel J. and Minoji Akimoto. 1999. *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*. Amsterdam: John Benjamins.

Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Mengistu Amberber, Bret Baker and Mark Harvey (eds.), *Complex Predicates in Cross-Linguistic Perspective*, 48–78. Cambridge: Cambridge University Press.

Cleasby, Richard and Guðbrand Vigfússon. 1874. *An Icelandic-English Dictionary*. Oxford: Clarendon.

Filatkina, Natalia. 2018. *Historische formelhafte Sprache: Theoretische Grundlagen und methodische Herausforderungen*. Berlin: De Gruyter. https://doi.org/10.1515/9783110494884.

Fix, Hans, Norbert Endres, Andrej Schabalin, Andreas Braml, Sebastian Holtzhauer, Jana Ilgner, Ann-Kathrin Müller and Fabian Schwabe. 2006. *Walter Baetke: Wörterbuch zur altnordischen Prosaliteratur digital*. Greifswald: Ernst-Moritz-Arndt-Universität.

Fleischhauer, Jens. 2022. *stehen unter*-Funktionsverbgefüge und ihre Familien. *Zeitschrift für germanistische Linguistik* 50(2). 247–288. https://doi.org/10.1515/zgl-2022-2055.

Fleischhauer, Jens and Stefan Hartmann. 2021. The emergence of light verb constructions: A case study on German *kommen* ‘come’. *Yearbook of the German Cognitive Linguistics Association* 9. 135–156. https://doi.org/10.1515/gcla-2021-0007.

Fleischhauer, Jens and Stefan Hartmann. 2023. Der Weg von kommen zum Funktionsverb. In Christian Braun and Elisabeth Scherr (eds.), *Methoden zur Erforschung grammatischer Strukturen in historischen Quellen*, 117–138. Berlin: De Gruyter. https://doi.org/10.1515/9783110784282-008.

Fulk, Robert D. 2018. *A comparative grammar of the early Germanic languages*. Amsterdam: John Benjamins.

García García, Luisa. 2005. *Germanische Kausativbildung: die deverbalen jan-Verben im Gotischen*. Göttingen: Vandenhoeck & Ruprecht.

Grestenberger, Laura and Itamar Kastner. 2022. Directionality in cross-categorial derivations. *Glossa* 7(1). 1–64. https://doi.org/10.16995/glossa.8710.

Hackstein, Olav. 2002. Uridg. *CH.CC > *C.CC. *Historische Sprachforschung* 115. 1–22. https://www.jstor.org/stable/41289070.

Harm, Volker. 2021. *Funktionsverbgefüge des deutschen: Untersuchungen zu einer Kategorie zwischen Lexikon und Grammatik*. Berlin: De Gruyter. https://doi.org/10.1515/9783110661255.

Harðarson, Jón Axel. 2017. The morphology of Germanic. In Klein Jared, Joseph Brian and Fritz Matthias (eds.), *Handbook of Comparative and Historical Indo-European Linguistics, Vol. 2*, 913–954. Berlin: De Gruyter. https://doi.org/10.1515/9783110523874-010.

Harðarson, Jón Axel. 2018. Germanisch *truwa-, *trewwa- und *-trewwija- und ihre urindogermanischen Grundlagen. In Olav Hackstein and Andreas Opfermann (eds.), *Priscis Libentius et Liberius Novis*, 229–264. Hamburg: Baar-Verlag.

Hellan, Lars. 2023. Unification and selection in Light Verb Constructions. A study of Norwegian. In Pompei Anna, Mereu Lunella and Piunno Valentina (eds.), *Light Verb Constructions as Complex Verbs*, 45–70. Berlin: De Gruyter. https://doi.org/10.1515/9783110747997-003.

Heringer, Hans Jürgen. 1968. *Die Opposition von 'kommen' und 'bringen' als Funktionsverben*. Düsseldorf: Pädagogischer Verlag Schwann.

Ittzés, Máté. 2024. Proto-Indo-European support verbs and support-verb constructions. In Victoria Beatrix Fendel (ed.), *Support-verb constructions in the corpora of Greek: Between lexicon and grammar?*, 3–32. Berlin: Language Science Press. 10.5281/zenodo.14017921.

Jacobi, Theodor. 1843. *Beiträge zur deutschen Grammatik*. Berlin: T. Trautwein.

Katz, R. Moses. 2021. *The Gothic resultative: non-agentive verbs and perfect expression in early Germanic*. Leiden: Brill.

Krahe, Hans and Wolfgang Meid. 1969. *Germanische Sprachwissenschaft*. Berlin: De Gruyter.

Krause, Wolfgang. 1971. *Die Sprache der urnordischen Inschriften*. Heidelberg: Winter.

Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Leiden: Brill.

Krämer, Peter. 1971. *Die Präsensklassen des germanischen schwachen Verbums*. Innsbruck: Institut für Vergleichende Sprachwissenschaft der Universität Innsbruck.

Kuroda, Susumu. 2017. Formen der ornativen Verben. *Jahrbuch für Germanistische Sprachgeschichte* 8(1). 121–134. https://doi.org/10.1515/jbgsg-2017-0009.

Kölligan, Daniel. 2018. Funktionsverbgefüge und Sekundärwurzeln. In Elisabeth Rieken, Ulrich Geupel and Theresa Maria Roth (eds.), *100 Jahre Entzifferung des Hethitischen. Morphosyntaktische Kategorien in Sprachgeschichte und Forschung. Akten der Arbeitstagung der Indogermanischen Gesellschaft vom 21. bis 23. September 2015 in Marburg*, 219–238. Wiesbaden: Reichert.

Kümmel, Martin. 2023. *Addenda und Corrigenda zu LIV*^{2}. https://www.gw.uni-jena.de/phifakmedia/42185/kuemmel-liv2-add.pdf, last access 26/9/2024.

Lühr, Rosemarie. 1988. *Expressivität und Lautgesetz im Germanischen*. Heidelberg: Winter.

Lühr, Rosemarie. 2000. *Die Gedichte des Skalden Egill*. Dettelbach: J.H. Röll.

Lühr, Rosemarie, Albert Larry Lloyd and Otto Springer. 1988–2024. *Etymologisches Wörterbuch des Althochdeutschen*. Göttingen: Vandenhoeck & Ruprecht.

Maini, Andrea. 2017. The relationship between early old nordic **bākn* 'fire beacon', pre-old frisian **bākən* 'id.' and germanic **baugō-*, **baukkaz* 'ring'. *Filologia Germanica* 9. 109–134.
Marti Heinzle, Mirjam. 2014. Die schwachen Verben der dritten Klasse im Altfriesischen – eine Spurensuche. *Amsterdamer Beiträge zur älteren Germanistik* 73(1). 145–184. https://doi.org/10.1163/9789401211918_007.
Mayrhofer, Manfred. 1992. *Etymologisches Wörterbuch des Altindoarischen. I Band*. Heidelberg: Winter.
Nedoma, Robert. 2010. *Kleine Grammatik des Altisländischen*. Heidelberg: Winter.
Noreen, Adolf. 1923. *Altisländische und altnorwegische Grammatik*. Halle a. d. Saale: Niemeyer.
ONP. 2005. Old Norse Prose Dictionary. https://onp.ku.dk/onp/onp.php, last access 27/9/2024.
Ottósson, Kjartan. 2013. The anticausative and related categories in the Old Germanic languages. In Folke Josephson and Ingmar Söhrman (eds.), *Diachronic and Typological Perspectives on Verbs*, 329–382. Amsterdam: John Benjamins.
Relleke, Walburga. 1974. Funktionsverbgefüge in der althochdeutschen Literatur. *Amsterdamer Beiträge zur älteren Germanistik* 7. 1–46.
Ringe, Donald A. 2017. *From Proto-Indo-European to Proto-Germanic*. Oxford: Oxford University Press.
Rix, Helmut and Martin Joachim Kümmel. 2001. *Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen*. Wiesbaden: Reichert.
Schaefer, Christiane. 1984. Zur semantischen Klassifizierung germanischer denominaler ōn-Verben. *Sprachwissenschaft* 9(4). 356–383.
Scheungraber, Corinna. 2014. *Die Nasalpräsentien im Germanischen. Erbe und Innovation*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.
Schuhmann, Roland. 2016. IK 639. Eine Neuinterpretation der abschließenden Sequenz auf dem Brakteaten von Trollhättan II. *Amsterdamer Beiträge zur älteren Germanistik* 76(4). 447–454. https://doi.org/10.1163/18756719-12340054.
Schuhmann, Roland. 2024. *Grundriss der Gotischen Etymologie*. =https://www.academia.edu/56815812/GRUNDRISS_DER_GOTISCHEN_ ETYMOLOGIE, last access 26/9/2024.
Schulte, Michael. 2023. Bauprinzipen runischer Formeln. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 145(1). 1–34. https://doi.org/10.1515/bgsl-2023-0001.
Schutzeichel, Marc. 2013. *Indogermanische Funktionsverbgefüge*. Münster: Universitäts- und Landesbibliothek Münster.
Schwerdt, Judith. 2008. *Morphosemantik der schwachen Verben im Ostgermanischen und Kontinentalwestgermanischen*. Frankfurt am Main: Peter Lang.
Seebold, Elmar. 1970. *Vergleichendes und etymologisches Wörterbuch der germanischen starken Verben*. The Hague: De Gruyter Mouton.
Sundquist, John D. 2018. A Diachronic Analysis of Light Verb Constructions in Old Swedish. *Journal of Germanic Linguistics* 30(3). 260–306. https://doi.org/10.1017/S1470542717000137.
Sundquist, John D. 2020. Productivity, Richness, and Diversity of Light Verb Constructions in the History of American English. *Journal of Historical Linguistics* 103. 349–388. https://doi.org/10.1075/jhl.19009.sun.
Torp, Alf. 1974. *Gamalnorsk ordavleiding*. Lund: CWK Gleerup.
von Polenz, Peter. 1963. *Funktionsverben im heutigen Deutsch. Sprache in der rationalisierten Welt*. Düsseldorf: Schwann.
von Polenz, Peter. 1987. Funktionsverben, Funktionsgefüge und Verwandtes. Vorschläge zur satzsemantischen Lexikographie. *Zeitschrift für germanistische Linguistik* 15. 169–189. https://doi.org/10.1515/zfgl.1987.15.2.169.
Wessén, Elias. 1970. *Schwedische Sprachgeschichte II*. Berlin: De Gruyter.

Wissmann, Wilhelm. 1932. *Nomina postverbalia in den altgermanischen sprachen, nebst einer voruntersuchung über deverbative ō-verba*. Göttingen: Vandenhoeck & Ruprecht.

Yakubovich, Ilya. 2013. Reflexes of Indo-European 'ē-Statives'" in Old Indic. *Transactions of the Philological Society* 112. 386–408. https://doi.org/10.1111/1467-968X.12039.

Index

https://doi.org/10.1515/9783112229330-009