



Universiteit
Leiden
The Netherlands

Verb-based restrictions on noun incorporation across languages

Olthof, M.; Lier, E. van; Claessen, T.; Danielsen, S.; Haude, K.;
Lehmann, N.; ... ; Wolvengrey, A.

Citation

Olthof, M., Lier, E. van, Claessen, T., Danielsen, S., Haude, K.,
Lehmann, N., ... Wolvengrey, A. (2020). Verb-based restrictions on
noun incorporation across languages. *Linguistic Typology*, 25(2),
211-256. doi:10.1515/lingty-2020-2054

Version: Publisher's Version
License: [Creative Commons CC BY 4.0 license](#)
Downloaded from: <https://hdl.handle.net/1887/3201040>

Note: To cite this publication please use the final published version
(if applicable).



Article

Marieke Olthof, Eva van Lier*, Tjeu Claessen,
Swintha Danielsen, Katharina Haude, Nico Lehmann,
Maarten Mous, Elisabeth Verhoeven, Eline Visser,
Marine Vuillermet and Arok Wolvengrey

Verb-based restrictions on noun incorporation across languages

<https://doi.org/10.1515/lingty-2020-2054>

Abstract: Although some characteristics of incorporating verbs and non-incorporating verbs have been proposed in previous studies, little systematic cross-linguistic research has been done on restrictions on the types of verbs that incorporate nouns. Knowledge about possible verb-based restrictions on noun

***Corresponding author: Eva van Lier** [*e:va van 'li:ɹ*], Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 134, 1012 VB Amsterdam, The Netherlands, E-mail: E.H.vanLier@uva.nl

Marieke Olthof [*ma'ri:kə'ɔltɦɔf*]: Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 134, 1012 VB Amsterdam, The Netherlands, E-mail: M.D.Olthof@uva.nl

Tjeu Claessen [*tʃø 'kla:sə*]: Leiden University Centre for Linguistics, Leiden University, Van Wijkplaats 4, 2311 BX Leiden, The Netherlands, E-mail: tjeuclaessen@gmail.com

Swintha Danielsen [*'svinta 'danjelzən*]: Karlstr. 27, 31785 Hameln, Germany, E-mail: swintha@hotmail.com

Katharina Haude [*katar'ina 'hawdə*]: Structure et Dynamiques des Langues, Centre national de la recherche scientifique, 7 Rue Guy Môquet, 94800 Villejuif, France, E-mail: Katharina.HAUDE@cncrs.fr

Nico Lehmann [*ni:.ko le:.man*] and **Elisabeth Verhoeven** [*e'li:zabet fεɣ'hu:vŋ*]: Institut für deutsche Sprache und Linguistik, Humboldt University Berlin, Dorotheenstraße 24, 10117 Berlin, Germany, E-mail: nico.lehmann@hu-berlin.de (n. le:man), verhoeve@cms.hu-berlin.de (e. fεɣ'hu:vŋ)

Maarten Mous [*ma:rtə ma's*]: Leiden University Centre for Linguistics, Leiden University, P.N. van Eykhof 3, 2311 BV Leiden, The Netherlands, E-mail: M.Mous@hum.leidenuniv.nl

Eline Visser [*ɛi'li:nə 'fisəɹ*]: Centre for Languages and Literature, Lund University, Helgonabacken 12, Lund, Sweden, E-mail: eline.visser@ling.lu.se

Marine Vuillermet [*maʁin vuʝjεʁmε*]: Department of Comparative Language Science, University of Zurich, Plattenstrasse 54, 8032 Zürich, Switzerland, E-mail: marinelouise.vuillermet@uzh.ch

Arok Wolvengrey [*'ε.ɹɪk 'wɔlvən gɹɛi*]: Department of Indigenous Languages, Arts and Cultures, First Nations University, 1 First Nations Way, Regina, Saskatchewan, S4S 7K2 Canada, E-mail: awolvengrey@firstnationsuniversity.ca

incorporation may, however, provide important insights for theoretical approaches to noun incorporation, in particular regarding the question to what extent incorporation is a lexical or a syntactic process, and whether and how languages may vary in this respect. This paper therefore investigates to what extent languages restrict noun incorporation to particular verbs and what types of restrictions appear to be relevant cross-linguistically. The study consists of two parts: an explorative typological survey based on descriptive sources of 50 incorporating languages, and a more detailed investigation of incorporating verbs in corpus data from a sample of eight languages, guided by a questionnaire. The results demonstrate that noun incorporation is indeed restricted in terms of which verbs allow this construction within and across languages. The likelihood that a verb can incorporate is partly determined by its degree of morphosyntactic transitivity, but the attested variation across verbs and across languages shows that purely lexical restrictions play an important role as well.

Keywords: corpus, lexical restrictions, incorporating verbs, morphosyntax, noun incorporation, transitivity

1 Introduction¹

This paper reports on a cross-linguistic investigation of verb-based restrictions on noun incorporation. Although some characteristics of incorporating and non-incorporating verbs have been proposed in previous studies, little systematic cross-linguistic research has been done on restrictions on the types of verbs that incorporate nouns. Restrictions on properties of incorporated nouns relating to their semantic role, syntactic function, modifiability, and referential status are relatively well-known. By contrast, the properties of verbs that are likely or unlikely to show incorporation across languages have received less attention.

Verb-based restrictions on noun incorporation may, however, be highly relevant for theoretical approaches to noun incorporation. One of the main questions addressed in the literature on noun incorporation concerns the status of incorporation as a lexical or syntactic process (Mithun 2000: 923–925; Massam

1 Author contribution statement:

Olthof and van Lier conducted the typological survey, designed the questionnaire, analysed the data, and wrote the paper. The data for the questionnaire-based studies, including the examples, were provided by the respective expert author(s), who also gave feedback on earlier versions of the paper.

2009: 1083–1086, 2017; Haugen 2015: 414–421): Is incorporation a lexically restricted type of word formation? Or is it rather a productive process that can be described by purely syntactic principles? Based on the characteristics and possibilities of incorporated nouns, arguments supporting each of these alternatives have been put forward. For instance, some studies state that incorporation, in contrast to most syntactic processes, is sensitive to the semantic roles of potentially incorporated nouns (Mithun 1984: 875; Anderson 2000: 16), while other works emphasize that only nouns in particular syntactic positions can be incorporated (Baker 1988: 81, 88, 90). In addition, some researchers have addressed the formal properties of incorporated nouns, demonstrating that in some languages not only noun stems but also inflected nouns and noun phrases can be incorporated (Barrie and Mathieu 2016; Olthof 2020), which may be regarded as evidence for the syntactic status of noun incorporation in these languages. By contrast, others show that in some languages incorporation is limited to specific semantic type of nouns, such as body-part nouns (Aikhenvald 2007: 20; Caballero et al. 2008: 391).

Knowledge about possible verb-based restrictions on noun incorporation may also provide important insights concerning the question to what extent incorporation is a lexical or a syntactic process, and whether and how languages may vary in this respect. Several studies have suggested that verb-based restrictions are relevant for noun incorporation. Caballero et al. (2008: 392), for instance, state that “it is not unexpected that some verbs might be able to select for a nominal object that will incorporate into them”, i.e. for any transitive verb it may be specified lexically whether or not it can incorporate its object noun. Moreover, for certain languages it has been noted that noun incorporation is only likely or possible with particular verbs. For instance, Mithun (2010: 52) notes that in Mohawk, noun incorporation is restricted in terms of both the nouns and verbs that can be involved in incorporation constructions, in that “some stems occur exclusively in such constructions, some often, some occasionally, some rarely, and some never”. For Ket, it is known that “[o]nly two transitive bases allow incorporation of their patient-role noun object with any productivity” (Vajda 2017: 911). In addition, verb-based restrictions have been observed for several voice- and valency-affecting alternations (Tsunoda 1985: 391–392; Kemmer 1993: 42–74; Næss 2007: 124–141; Polinsky 2013; Say in prep.; Vigus 2018: 370–371). For example, in some languages antipassives may only be formed on the basis of “a certain subset of transitive predicates” (Polinsky 2013). Such restrictions may also be relevant for noun incorporation, which in many languages has a valency-changing effect in that incorporated nouns may lose their morphosyntactic status of core argument (Gerds 1998: 88; Mithun 1984: 856, 859; Rosen 1989: 310–311).

This paper therefore aims to investigate to what extent languages restrict noun incorporation to particular verbs and what types of restrictions appear to be relevant cross-linguistically. Section 2 introduces the definition of noun incorporation used in the study, discusses earlier studies relevant for verb-based restrictions on noun incorporation, and formulates the research questions. Section 3 presents the results of the first part of the study, which consists of an explorative typological survey based on descriptive sources of 50 incorporating languages (cf. Olthof and van Lier 2018). Section 4 discusses the second part of the study, which investigates verb-based restrictions on noun incorporation more systematically in a sample of eight languages, guided by a questionnaire and based on data from spoken language corpora. Finally, in Section 5 we conclude that noun incorporation is indeed restricted in terms of which verbs allow this construction within and across languages. The likelihood that a verb can incorporate appears to be partly determined by its degree of morphosyntactic transitivity, but the attested variation across verbs and across languages shows that purely lexical restrictions play an important role as well.

2 Definition, theoretical background and research questions

2.1 Defining noun incorporation

While many different definitions of noun incorporation are used throughout the literature (see e.g. Johns 2017; Massam 2009), this study follows Caballero et al. (2008: 385) in defining noun incorporation as a construction in which a noun occurs “between parts of the inflected verbal complex”. The relevant “parts” are, in addition to the verbal stem, typically affixes, but they may also be clitics, particles, or other separate words that appear strictly adjacent to verbs (Caballero et al. 2008: 385).² Thus, in example (1b) from Chukchi, the noun *utt* ‘stick’ is considered to be incorporated because it is preceded by the first part of the verbal person marking circumfix *t...-γʔek* and followed by the verbal stem and the second part of the circumfix.

² Note that in the questionnaire-based case studies of noun incorporation, discussed in Section 4, a few additional language-specific criteria are used (see Section 4.1.2 and Appendix 4 for more details).

- (1) Noun incorporation in Chukchi (Kurebito 2001: 79)³
- a. *yəm-nan t-ə-mle-γʔen-Ø ott-ə-lyən*
 I-ERG 1SG.A-E-break-3SG.P-PST stick-E-ABS.SG
 ‘I broke the stick.’
- b. *yəm-Ø t-utt-ə-mle-γʔek-Ø*
 I-ABS 1SG.S-stick-E-break-1SG.S-PST
 ‘I broke a stick.’

However, the construction from Niuean in (2b) is also regarded as a noun incorporation construction, because the noun *ika* ‘fish’ is preceded by the verbal stem *takafaga* ‘hunt’ and followed by the verbal clitics *tūmau* and *nī* (cf. Seiter 1980: 22–24). Example (2a) provides the non-incorporated counterpart of the construction.

- (2) Noun incorporation in Niuean (Seiter 1980: 69)
- a. *Takafaga=tūmau=nī e ia e tau ika.*
 hunt=always=EMPH ERG he ABS PL fish
 ‘He’s always fishing.’
- b. *Takafaga ika=tūmau=nī a ia.*
 hunt fish=always=EMPH ABS he
 ‘He’s always fishing.’

Note that the requirement that incorporated nouns must occur between parts of the inflected verbal complex is only used in this study to identify a language as noun-incorporating. Individual constructions may not satisfy this requirement, for instance because some values of the relevant inflectional feature have forms that do not appear in the relevant positions or are zero-marked. Thus, example (1b) from Chukchi demonstrates that this language makes use of noun incorporation, as it shows a construction in which a noun is included between the first part of the first-person circumfix on the one hand, and the verb stem and the second part of the circumfix on the other. However, the affix marking third person is a suffix rather than a circumfix, such that the noun *ηekk* ‘daughter’ in example (3) does not appear between parts of the inflected verbal complex.

- (3) Noun incorporation in Chukchi (Kurebito 2001: 76)
- əlləγ-ə-n ηekk-imti-γʔi-Ø*
 father-E-ABS.SG daughter-carry.on.the.back-3SG.S-PST
 ‘The father carried his daughter on his back.’

³ The vowel difference between *ott* in (1a) and *utt* in (1b) is due to a vowel harmony rule (Kurebito 2001: 66).

Nevertheless, example (3) is included in the study as a case of noun incorporation, because the existence of examples such as (1b) shows that Chukchi can be regarded as a noun-incorporating language according to our definition.

Similarly, in some languages the features whose marking may show that the incorporated noun appears between parts of the inflected verbal complex may only be required in particular contexts or constructions. For example, in Western Frisian, constructions with the verbal infinitive marker *te*, exemplified in (4a), show that this language makes use of noun incorporation. In the finite construction shown in example (4b), by contrast, there is no verbal marking preceding the noun *hier* ‘hair’. Nevertheless, this noun is considered to be incorporated because there is evidence for noun incorporation in Western Frisian in the form of constructions like the one in example (4a).

(4) Noun incorporation in Western Frisian (Dijk 1997: 41, 44)⁴

- a. *De kapper begjint te hier-knipp-en*
 The barber Begins to hair-cut-INF
 ‘The barber begins to cut the hair.’
- b. *Ik sil him hier-knipp-e*
 I will him hair-cut-INF
 ‘I will cut his hair.’

It is also important to mention that the definition we use does not delimit incorporation in terms of the semantic or syntactic role of the incorporated noun. Nor does it pose any restrictions on the formal characteristics of incorporated nouns and incorporating verbs. Although incorporated nouns in most languages are identical or at least very similar in form to corresponding independently occurring nouns, in a few languages such as Halkomelem (Gerds 2003: 345–346) some incorporated nouns have non-incorporated counterparts that are formally completely unrelated. Such nouns may, however, be seen as suppletive versions of non-incorporated nouns and the relevant constructions are not excluded from the study (cf. Caballero et al. 2008: 387–388). In addition, in some languages, including Movima (Haude 2006: 72–73) and Washo (Bochnak and Rhomieux 2013), some or all incorporated nouns are bound in the sense that they never occur without an additional morpheme or are even obligatorily incorporated into a verb. Cases with such incorporated nouns are

⁴ Note that Western Frisian distinguishes two infinitives formed with two different suffixes: *-en* and *-e*. Both forms are used in several different contexts, but one of the contexts in which the infinitive with *-en* is used is a construction with the infinitive marker *te*, as in (4a), while one of the contexts in which the infinitive with *-e* is used is after a modal auxiliary like *sil* ‘will’ in (4b) (Dijk 1997: 178–182).

included in the study as well. Similarly, we include constructions with obligatorily incorporating elements that are sometimes called affixes in languages like Eastern Ojibwa (see e.g. Mathieu 2013) and Kalaallisut (see e.g. Fortescue 1980), as long as these elements have action semantics and the resulting constructions conform to the definition of inclusion of the noun inside the inflected verbal complex.⁵ For instance, the Kalaallisut construction in example in (5) is considered to be a noun incorporation construction, even though the element *isur* ‘fetch’ cannot occur independently without a noun (Fortescue 1980: 274, 1984: 322).

- (5) Noun incorporation in Kalaallisut (Fortescue 1984: 322)
tiiturvi-isur-put
 cup-fetch-3PL.IND
 ‘They fetched (the) cups.’

Note that a language like Kalaallisut is included as a noun-incorporating language by Caballero et al. (2008: 412) as well.

2.2 Theoretical background

2.2.1 Noun incorporation and semantic transitivity

In the literature on noun incorporation various factors have been suggested to co-determine a verb’s likelihood to appear in noun incorporation constructions cross-linguistically. These factors can generally be linked to the notion of transitivity, in that a verb’s (degree of) transitivity appears to affect its ability to incorporate nouns. Both semantic and morphosyntactic aspects of transitivity have been mentioned in this respect.

From a semantic argument-structure perspective, Mithun (1984: 875) argues that transitive verbs are more likely to incorporate their patient arguments than intransitive verbs. More specifically, she proposes that all incorporating languages at least allow the incorporation of patient arguments of transitive verbs, i.e. P-arguments, such that intransitive verbs can only incorporate their patient

⁵ The inclusion of obligatorily incorporating or bound verbs is only relevant for the typological survey presented in Section 3; in the languages studied in the corpus-based case studies discussed in Section 4, such verbs only play a very marginal role. Overall, the meanings of the bound verbs included in the typological survey do not seem to be very different from the meanings of the other incorporating verbs. Therefore, we do not distinguish between bound verbs and other incorporating verbs in the discussion of the results of the typological survey in Section 3.

arguments, i.e. Sp-arguments, in languages that also show the incorporation of P-arguments (see also Haspelmath 2018: 318, fn. 9).⁶

In addition, the ability of different verbs to incorporate has been linked to certain semantic characteristics of their prototypical patient argument, which are also related to these verbs' degree of semantic transitivity. Firstly, Mithun (1984: 863) states that verbs with highly affected patient arguments are more likely to incorporate these arguments than verbs with less affected patient arguments. As highly affected P-arguments are seen as prototypical of highly transitive verbs (Hopper and Thompson 1980: 252; Malchukov 2005: 80; Tsunoda 1981: 393), the preference for incorporation into verbs with such P-arguments suggests a relation between noun incorporation and high semantic transitivity.

Secondly, Mithun (1984: 863) maintains that verbs that tend to take inanimate, non-agentive, and non-individuated patient arguments are more suitable for incorporation than those with animate, agentive, and individuated patient arguments. Importantly, non-individuated P-arguments are characteristic of verbs that are semantically low in transitivity (Hopper and Thompson 1980: 252–253; Tsunoda 1981: 393). Thus, the non-individuated status of the P-arguments of frequently incorporating verbs suggests that semantically low-transitive verbs are most likely to incorporate.

With respect to semantic transitivity, Mithun's claims that incorporating verbs tend to have patient arguments that are on the one hand highly affected but on the other hand inanimate, non-agentive, and non-individuated thus appear to be contradictory: high affectedness of the P-argument is a characteristic of semantically high-transitive verbs, whereas non-individuation of the P-argument is a characteristic of low-transitive verbs. However, the relation between affectedness and individuation of P-arguments is not clear-cut. Vigus (2018: 373), focussing on antipassive constructions, shows that low individuation is not correlated with low affectedness: P-arguments that are low in individuation are not necessarily low in affectedness at the same time.

Note also that the noun incorporation process itself has often been regarded as a way to mark the referent of a noun as low in individuation. For instance, Hopper and Thompson (1980: 257) consider P-argument incorporation to correlate with low individuation of P-arguments. Similarly, for the seven P-incorporating languages included in her study, Vigus (2018: 360) finds that the function of incorporation is to indicate the lower individuation of P-arguments. Interestingly, if the

⁶ According to Mithun (1984: 875), languages that both allow the incorporation of P-arguments into transitive verbs and the incorporation of Sp-arguments into intransitive verbs may additionally show the incorporation of instruments and/or locations. Instrument incorporation and location incorporation thus appear to be more marginal types of noun incorporation.

function of noun incorporation is to mark a P-argument as low in individuation, we may expect that highly transitive verbs, which tend to take highly individuated P-arguments, are most likely to show incorporation. These verbs would then be used in an incorporation construction when their P-argument is, unexpectedly, less individuated.

2.2.2 Noun incorporation and morphosyntactic transitivity

There are also indications that incorporation potential is related to morpho-syntactic transitivity. First of all, Baker (1988) makes claims about the types of syntactic arguments that can be incorporated. He proposes that noun incorporation is a head-movement process in which internal arguments are moved to, i.e. incorporated into, a verb (Baker 1988: 82–83). Correspondingly, transitive verbs can incorporate their objects and unaccusative verbs their subjects, while unergative verbs do not allow incorporation of their subjects because these are external arguments (Baker 1988: 81–82, 87–90).⁷ This pattern has been described in other studies as well (e.g. Gerds 1998: 87).

Secondly, based on studies using data from the Valency Patterns Leipzig (ValPaL) Project (Hartmann et al. 2013), a possible relation between the incorporation of P-arguments and degree of morphosyntactic transitivity can be identified. The ValPaL project investigates the argument-coding properties of the translational equivalents of 80 verb meanings in 36 languages. As part of this project, Haspelmath (2015: 143) assigns a so-called “transitivity prominence” score to 70 of the 80 verb meanings. This means that for these verb meanings he calculates the percentage of transitively encoded verbs among all translational equivalents across the sample languages. Transitive encoding is defined as the coding used for the A- and P-arguments of the verb meaning “break” in a particular language. A score of 1 for a particular verb meaning indicates that its translational equivalents use basic transitive coding in all 36 languages, whereas a score of 0 means that the verb meaning does not have a translational equivalent with basic transitive coding in any of the 36 languages.

In the context of the same project, Malchukov (2015) and Wichmann (2015) study the ability of the 80 verb meanings to participate in object-demoting and object-deleting alternations, including P-argument incorporation, across the

⁷ According to Baker (1988: 86–87), adjuncts can never be incorporated. Note also that there is some similarity between Baker’s claim that only internal arguments can be incorporated and Mithun’s (1984: 875) proposal that all incorporating languages show the incorporation of patient arguments. However, whereas Mithun predicts that transitive verbs are more likely to incorporate nouns than intransitive ones, Baker does not make a distinction between transitive and unaccusative verbs but argues that incorporation into unergative verbs is impossible. In addition, Mithun (1984: 875) does not exclude the incorporation of nouns that are not arguments.

sample languages.⁸ Their findings lead to a hierarchy for object-demoting and object-deleting alternations in which verb meanings at the top are cross-linguistically most likely to undergo object-demoting and object-deleting alternations and verb meanings at the bottom are cross-linguistically least likely to undergo such alternations.⁹ This hierarchy, henceforth called the *object dem/del hierarchy*, is represented in (6) below. Note that in (6), only the 70 verb meanings also studied by Haspelmath (2015) are included and the transitivity prominence scores of these verbs are given between parentheses.¹⁰

According to the methods developed by Wichmann (2015, 2016; Aldai and Wichmann 2018), the object dem/del hierarchy can be interpreted as statistically implicational. Thus, verbs lower on this hierarchy can usually only be involved in object-demoting and object-deleting alternations in a particular language, if verbs higher on the hierarchy allow these alternations as well. Given that P-incorporation is also an object-demoting construction, we expect to find overlap between high-ranking verbs on this hierarchy and verbs that incorporate frequently across languages in our study.

8 The number of incorporation constructions included in their work is in fact fairly limited (see Wichmann 2015: 178). However, P-argument incorporation constructions share functional and sometimes also formal characteristics with other object-demoting constructions, such as anti-passives (Heaton 2017: 17; Vigus 2018), such that the hierarchy may nevertheless be quite relevant for the study of noun incorporation.

9 Note that this hierarchy also includes intransitive verb meanings. Although these verb meanings all appear in the lower part of the hierarchy, only one verb meaning, FEEL PAIN, appears at the lowest end of the hierarchy. This means that, surprisingly, translational equivalents of most of these intransitive verb meanings are able to undergo object-demoting and/or object-deleting alternations in at least some languages. On closer inspection, however, it appears that translational equivalents of many of the verb meanings in the lower part of the hierarchy can only be used in object-demoting and/or object-deleting alternations in very few of the languages studied and/or if they are first transitivized. For instance, translational equivalents of the verb meaning DIE can only undergo alternations that Wichmann (2015) classifies as object-demoting or object-deleting in Russian and Sliammon. In addition, in Bezhta the translational equivalents of BE DRY and BOIL can undergo an antipassive alternation, but only if they are combined with the causative suffix *-l*.

10 In (6), as in the remainder of this paper, we follow the ValPaL practice to write comparative verb meanings in small caps. Yet, for the sake of terminological simplicity, we will use the term “verb” as a shortcut for “(comparative) verb meaning”, unless we think it is important to explicitly differentiate between the comparative verb meaning and its translational equivalents, i.e. the actual language-specific lexical items, for which we also use the word “verb”. The lexical items will be written in italics and their translations given between single quotation marks, in accordance with general typological practice.

(6) Object dem/del hierarchy (adapted from Malchukov 2015: 105–106; Wichmann 2015: 166–167).

EAT (0.93), WASH (0.94), GIVE (0.98) >
 SHAVE (0.93) >
 CUT (1.00), SEARCH FOR (0.88), HIT (1.00) >
 KILL (1.00), ASK FOR (0.95), TAKE (1.00), BEAT (1.00) >
 SEE (0.93), THROW (0.98), TOUCH (0.84), LOOK AT (0.73) >
 BREAK (1.00), FILL (0.98), HUG (0.90), COVER (0.95), POUR (0.95), THINK (0.52), LOAD (0.96) >
 TELL (0.78), KNOW (0.88), TEAR (1.00), HELP (0.78), TIE (0.98), SHOW (1.00), CARRY (0.95) >
 SING (0.38), DRESS (0.92) >
 CLIMB (0.49), BUILD (0.93), FEAR (0.53) >
 SMELL (0.78), PUT (0.98), SEND (0.93), LEAVE (0.42) >
 PEEL (0.96), BLINK (0.11), SAY (0.41), TALK (0.40), SHOUT AT (0.45), NAME (0.80), RUN (0.05) >
 JUMP (0.00), HIDE (0.97), FRIGHTEN (0.98), LIKE (0.78), PLAY (0.10), FOLLOW (0.74), LIVE (0.05), BE DRY (0.00) >
 ROLL (0.00), LAUGH (0.03), BURN (INTR.) (0.00), SCREAM (0.03), GO (0.00), SINK (INTR.) (0.03) >
 MEET (0.70), DIE (0.00), COUGH (0.00), BE A HUNTER (0.00) >
 FEEL PAIN (0.12), SIT (0.05) >
 BE SAD (0.00) >
 SIT DOWN (0.03), BE HUNGRY (0.00) >
 RAIN (0.00) >
 FEEL COLD (0.00)

Crucially, it appears that the verbs higher on the hierarchy generally show higher transitivity prominence scores than the verbs lower on the hierarchy, i.e. they show transitive coding in more languages. The ranking of the verbs based on their transitivity prominence scores and their position on the object dem/del hierarchy show a strong and statistically significant correlation, as demonstrated by their Spearman's rank order correlation coefficient ($\rho = 0.78$, $p < 1e^{-16}$).¹¹ We gather from this that morphosyntactic transitivity as measured by Haspelmath's (2015) transitivity prominence scores is at least an important factor underlying the object

¹¹ Note that the object dem/del hierarchy results from a procedure based on Guttman scaling (see Aldai and Wichmann 2018; Wichmann 2015, 2016), while the ranks in Haspelmath's scale are based on simple counting. Yet, as shown in Aldai and Wichmann (2018: 270), although based on a smaller data set, the results are quite similar.

dem/del hierarchy, which suggests that it is also relevant for P-argument incorporation, in that verbs with higher transitivity prominence scores may be expected to be more likely to incorporate their P-arguments.

Malchukov (2015: 103–104) and Wichmann (2015: 167) indeed acknowledge that morphosyntactic transitivity plays a role in their hierarchy. On the other hand, they observe that telicity or the distinction between “manner” and “result” verbs proposed by Levin (2015) is important, in that atelic or “manner” verbs are generally ranked higher than telic or “result” verbs (Malchukov 2015: 105–106; Wichmann 2015: 167). This pattern is interesting, because Hopper and Thompson (1980: 252) and Tsunoda (1981: 393, 1985: 388) state that telicity is characteristic for high semantic transitivity. Thus, whereas the transitivity prominence scores of the verbs on the hierarchy suggest that morphosyntactically highly transitive verbs are more likely to undergo object-demoting and object-deleting alternations, the telicity of the verbs suggests the reverse pattern for semantic transitivity. Finally, Wichmann (2015: 167) notes that verbs expressing “actions that habitually involve a certain kind of object” tend to appear high on the object dem/del hierarchy. Although he explains this observation by proposing that such verbs often show object omission, it may also be relevant for noun incorporation, because noun incorporation has often been argued to express conventionalized or institutionalized activities (Massam 2017; Mithun 1984: 848).

2.3 Research questions

This study investigates verb-based restrictions on noun incorporation on the basis of the following research questions.

- (7) Research questions
 - a. Which verbs are most likely to incorporate nouns across languages?
 - b. To what extent are verb-based restrictions on noun incorporation determined by morphosyntactic transitivity?
 - c. What other factors affect the likelihood that a verb is able to incorporate nouns?
 - d. To what extent do languages differ in terms of how many and which verbs allow noun incorporation and how frequently these verbs show noun incorporation?

Note that the research question in (7b) specifically focuses on morphosyntactic transitivity as discussed in Section 2.2.2 rather than on the semantic characteristics that can be related to transitivity presented in Section 2.2.1. Whereas the effect of and relation

between these semantic characteristics remains somewhat unclear, morphosyntactic transitivity can be measured systematically on the basis of Haspelmath's (2015) notion of transitivity prominence and, as shown in Section 2.2.2, is also involved in the object dem/del hierarchy proposed by Malchukov (2015) and Wichmann (2015). In addition, morphosyntactic transitivity presumably reflects some aspects of semantic transitivity.

We try to answer the research questions in (7) on the basis of a study consisting of two parts. The first part is an explorative typological survey of 50 incorporating languages, while the second part focuses on eight incorporating languages, on the basis of more systematic and detailed questionnaire-based case studies using corpus data. The methodologies and results of each part of the study are now discussed in turn.

3 Typological survey

3.1 Method and data

The typological survey of incorporating verbs makes use of a sample of 50 languages drawn from a list of ca. 250 languages that are described as incorporating languages in the literature on incorporation. This list includes languages from 82 language families, and the 50 languages in our sample are all from different families.¹² We selected those languages for which most data could be obtained. The data are extracted from reference grammars and from articles on noun incorporation in the relevant languages. The sample languages and the data sources are included in Appendix 1 below.

For each language we listed the meanings of all verbs for which it is mentioned or shown in the sources that they can be used in noun incorporation. Note that, in accordance with the definition of noun incorporation introduced in Section 2.1, we did not restrict our search to P- or Sp-incorporation constructions, i.e. verbs allowing the incorporation of locations and instruments were included as well. However, as the most frequently incorporating verbs in the languages of our sample, to be presented in the next subsection, generally show P- or Sp-incorporation in our data, we focus on P- and Sp-incorporation in the remainder of this section and only mention other types of noun incorporation where they are particularly relevant.

Considering the exact glosses used for the incorporating verbs in the sources, the data collection resulted in a list of 808 different verb meanings. However, this list included many near-synonyms. For instance, a verb glossed as “roast” was found in one language and a verb glossed as “toast” in some others. Further

¹² There is one exception: both Panamint and Ute-Southern Paiute belong to the Uto-Aztecan family. Note also that, since noun-incorporating language families are not evenly distributed around the world (Velupillai 2012), geographical distribution was not taken into account in the sampling procedure.

examples include “look after” and “take care of” or “happen” and “occur”. For this reason, the next step of the data collection involved combining the near-synonyms into single entries in our list. To further reduce the number of meanings, we also merged certain stative verbs that are expressed in English by means of the verb “to be” combined with an adjective, grouping them in accordance with the property concept classes distinguished by van Lier (2017). For example, the meaning BE + PHYSICAL PROPERTY represents glosses such as “be dry” and “be dirty”, while the meaning BE + EXPERIENTIAL STATE includes e.g. “be tired” and “be hungry”.¹³ All in all, this merging procedure reduced the original list of 808 verb meanings to 526.

The method of data collection for the typological survey has some obvious limitations. Firstly, as can already be inferred from the list of sources included in Appendix 1, there are large differences in the amount of available data for the different languages. This means that the data gathered for some languages are much more likely to represent a substantial proportion of all verbs that can be used in noun incorporation than the data for other languages. Especially in the case of languages in which noun incorporation is a very productive process, our data necessarily cover only a small subset of the possibilities. Secondly, because the data were collected primarily on the basis of reference grammars, the data include information about verbs that can incorporate nouns, but very little information about which verbs cannot be used in noun incorporation. The latter form of evidence is found only in an indirect way in those few cases where incorporation is restricted to a very limited set of verbs (cf. the example of Ket in Section 1). Thirdly, the method does not take into account the frequency with which the relevant verbs are used in noun incorporation. The data thus only state that a particular verb can incorporate nouns, and do not include any information about how often the verb occurs in noun incorporation, compared to other constructions.

3.2 Results and discussion

3.2.1 Frequently incorporating verbs

Table 1 shows the meanings of the verbs that are found as incorporating verbs most frequently across the sample. Each of these verbs is found in at least 10 different languages. The number of languages in which a verb is found to incorporate is given in the second column. For those verbs which are also studied by Haspelmath (2015), we give the transitivity prominence scores between parentheses. Note that

¹³ The classes used in van Lier (2017) are similar to the ones proposed by Dixon (2004 and earlier work), but experiential states are called “corporeal properties” by Dixon and treated as a subclass of “physical properties”. See van Lier (2017) for a more detailed description and motivation of the various classes.

most of the verb meanings in Table 1 in fact represent mergings of multiple glosses, as explained and illustrated in the previous subsection. The exact set of glosses included under each verb meaning can be found in Appendix 2.¹⁴

Table 1: Verbs found as incorporating verbs in at least 10 out of 50 languages in the survey.¹⁵

Verb meaning	Number of languages (of total 50)
CUT (1.00)	22
MAKE/DO	21
EAT (0.93)	19
SEARCH FOR (0.88)	17
GIVE (0.98)	16
WASH (0.94)	16
PUT (0.98)	14
BREAK (TR.) (1.00)	13
KILL (1.00)	13
BE + PHYSICAL PROPERTY (0.00)	12
BUY	12
FEEL PAIN (0.12)	12
GO (.05)	12
HAVE	12
TAKE (1.00)	12
FALL	11
HIT (1.00)	11
HUNT	11
PUT DOWN	11
REMOVE	11
SEE (0.93)	11
BE + EXPERIENTIAL STATE (0.00)	10
CATCH	10
DIE (0.00)	10

3.2.2 Morphosyntactic transitivity

Table 1 shows that almost all of the most frequently incorporating verbs either have a (di)transitive meaning or are patientive intransitive verbs. This is in accordance with Baker's (1988) claim that only transitive verbs and unaccusative verbs can incorporate

¹⁴ Appendices 2–5 can be found in the online Supplementary materials to this paper.

¹⁵ The transitivity prominence score given for the meaning BE + PHYSICAL PROPERTY corresponds to the transitivity prominence score for the meaning BE DRY in Haspelmath's (2015) study, and the transitivity prominence score given for the meaning BE + EXPERIENTIAL STATE corresponds to the transitivity prominences score for the meanings BE SAD AND BE HUNGRY in Haspelmath's (2015) study.

nouns. Nevertheless, Table 1 also includes an exception to this pattern: the verb GO is generally an agentive intransitive. In addition, among the other, less frequently incorporating verbs (not included in Table 1) we also find, for instance, COME (in four languages), JUMP (in four languages), and RUN (in three languages). However, with these agentive intransitive verbs it is typically a goal or location that is incorporated rather than the agent (Sa) argument, as for example in (8) from Southern Tiwa:¹⁶

- (8) Noun incorporation in Southern Tiwa (Allen et al. 1984: 309)
Te-fiesta-mĩ-ban
 1SG.S-party-go-PST
 ‘I went to the party.’

Note also that it is known that, in some languages, agentive verbs of manner of motion such as RUN and JUMP show unaccusative behaviour when they combine with directional phrases (Levin and Rappaport Hovav 1995: 182–186). The ability to incorporate nouns may be one example of such unaccusative behaviour.

As mentioned, Table 1 also shows the transitivity prominence scores of those verbs that are included in Haspelmath (2015) between parentheses. As can be seen in the table, most of the frequently incorporating verbs have a transitivity prominence score of 0.88 or higher, except the intransitive verbs BE + PHYSICAL PROPERTY, FEEL PAIN, GO, BE + EXPERIENTIAL STATE, and DIE. It thus appears that of the transitive verbs, those with high transitivity prominence as defined by Haspelmath (2015) are good candidates for noun incorporation across languages.

The hierarchy in (9) shows the object dem/del hierarchy introduced in Section 2.2 and indicates the number of languages in which each of the verbs included in this hierarchy is found as an incorporating verb in our typological survey.

- (9) Object dem/del hierarchy (adapted from Malchukov 2015: 105–106; Wichmann 2015: 166–167). For each verb, the number between parentheses indicates the number of languages included in the sample of 50 languages in which it is found as an incorporating verb.

EAT (19), WASH (16), GIVE (16) >
 STEAL (4), TEACH (1), SHAVE (2), COOK (6) >
 CUT (21), WIPE (5), SEARCH FOR (17), HIT (11) >
 KILL (13), ASK FOR (2), TAKE (12), BEAT (4) >

¹⁶ In a few cases, the semantic role of the incorporated noun was hard to determine. Note, though, that both the incorporation of adjuncts and the incorporation of Sa-arguments are not predicted by Baker's (1988) theory, whereas Mithun (1984) does recognize the incorporation of instruments and locations. ¹⁷ In our data, we could not distinguish between the meanings SEE VERSUS LOOK AT and we included all examples under SEE. Hence, the “0” reported here is due to this choice, rather than to the fact that LOOK AT does not incorporate in any of the sample languages.

SEE (11), THROW (8), HEAR (3), TOUCH (6), LOOK AT (0)¹⁷ >
 GRIND (2), BREAK (TR.) (13), FILL (2), HUG (2), COVER (6), POUR (4), THINK (1), LOAD (4) >
 TELL (4), KNOW (4), TEAR (3), HELP (1), TIE (7), SHOW (2), CARRY (7) >
 SING (2), DIG (6), DRESS (0) >
 CLIMB (2), BUILD (6), FEAR (1) >
 SMELL (TR.) (0), PUSH (3), PUT (14), SEND (3), LEAVE (TR.) (6) >
 PEEL (2), BLINK (0), SAY (2), TALK (2), SHOUT AT (0), NAME (1), RUN (3) >
 JUMP (4), HIDE (2), FRIGHTEN (1), LIKE (7), PLAY (TR.) (3), FOLLOW (7), LIVE (0), BE DRY (1) >
 BRING (6), ROLL (0), LAUGH (0), BURN (INTR.) (1), SCREAM (0), GO (12), SINK (INTR.) (0) >
 MEET (0), DIE (10), COUGH (0), BOIL (INTR.) (0), BE A HUNTER (0) >
 FEEL PAIN (12), SIT (5) >
 BE SAD (0) >
 SIT DOWN (1), BE HUNGRY (1) >
 RAIN (1) >
 FEEL COLD (2)

It can be observed in (9) that some of the verbs that were expected to be prone to noun incorporation because they are high on the object dem/del hierarchy are indeed found as incorporating verbs in many of the sample languages. These include, for instance, EAT, WASH, GIVE, CUT, SEARCH FOR, KILL, and BREAK (TR.). Correspondingly, some of the verbs that are low on this hierarchy, such as LAUGH, SIT, SIT DOWN, BE HUNGRY, and FEEL COLD, are found as noun-incorporating verbs in few or none of the sample languages. Thus, there seems to be some overlap between verbs that are unlikely to undergo different types of object-demoting and object-deleting alternations and verbs that are unlikely to undergo noun incorporation. The data of the present study are, however, not completely in line with the expectations based on the object dem/del hierarchy. This will be discussed further in the next subsection.

3.2.3 Other factors

Some verbs that appear high on the object dem/del hierarchy are not found as noun-incorporating verbs in many of our sample languages. For a few of these cases we can offer a tentative explanation. Firstly, verbs such as THINK, TELL, and KNOW may be unlikely candidates for incorporation because they potentially or even typically take clausal complements rather than nominal objects.¹⁸ It may be

¹⁸ Also, Aldai and Wichmann (2018: 271, 273) show that THINK and KNOW are relatively likely candidates for, respectively, oblique-object and inverted coding frames, which may reduce the incorporability of their P-arguments even if they are noun phrases (rather than complement clauses).

the case that these verb meanings more easily allow object deleting alternations, which may explain their different behaviour in our data compared to the object dem/del hierarchy. Secondly, a few semantically quite specific verbs such as WIPE and GRIND we may have found in relatively few languages as noun-incorporating verbs simply because they are not used in many data sources. More generally, of course, the fact that a particular verb does not appear often in our data could well be a side effect of the method of data collection.

On the other hand, there are also some verbs that are low on the object dem/del hierarchy, yet are found as incorporating verbs in relatively large numbers of languages. Three of these, GO, DIE, and FEEL PAIN, correspond to intransitive verbs, which explains why they are unlikely candidates for object-demoting and object-deleting alternations but do occur as incorporating verbs. Another verb that is found more frequently as an incorporating verb than expected based on this hierarchy is PUT. The relatively high number of languages that show noun incorporation with this verb may be related to the observation made by Mithun (1984: 863) that verbs with very general semantics are likely to incorporate nouns. Note here that we also find MAKE/DO, HAVE, and TAKE, all of which may be considered to be general in semantic scope, among the most frequently incorporating verbs, as can be seen in Table 1.¹⁹

3.2.4 Variation across languages

It may be noted that there are large differences in our data between languages in the number of verbs found to be able to incorporate nouns; figures vary between 1, for instance for Atsugewi, and 101, for Western Frisian. We cannot, however, draw any firm conclusions from the attested variation, because it is strongly influenced by the sources we used: while Palancar (1999) mentions only a single concrete example for Atsugewi in a comparative study, Dijk (1997) devotes an entire dissertation to noun incorporation in Western Frisian. Although in many cases the sources used may indeed give a reasonably representative impression of the verb-based productivity of noun incorporation in a given language, cases like Atsugewi show that the study does not amount to an overall reliable picture. This issue is addressed by the second part of this study to which we now turn.

¹⁹ Another question that presents itself is the influence of the combination of specific verbs with specific incorporated nouns. We checked which noun was incorporated in 950 of our assembled examples involving the most frequently incorporating verbs. This non-systematic exploration confirmed the oft-noted cross-linguistic preference for body-part noun incorporation: 350 out of the 950 examples involved a body-part noun. However, we did not find clear evidence for cross-linguistic collocations, i.e. for very highly frequent combinations of specific verbs and nouns.

4 Questionnaire-based case studies

4.1 Method and data

4.1.1 Questionnaire design

The second part of the study consists of eight systematic case studies of verb-based restrictions on noun incorporation. For these case studies we used a questionnaire, which was filled out on the basis of data from spoken language corpora. The questionnaire consists of 47 verb meanings for which we checked in each sample language whether or not its translational equivalent can occur in a noun incorporation construction and, if so, how often this happens relative to the verb's total token frequency. The questionnaire contains both verb meanings expected to favour incorporation cross-linguistically and verb meanings expected to disfavour or disallow incorporation.

The inventory of verb meanings in the questionnaire takes into account previous research on the role of morphosyntactic transitivity in incorporating verbs (see the discussion in Section 2.2.2) as well as the results of the typological survey discussed in Section 3.²⁰ First, the questionnaire verbs cover the full range of transitivity prominence scores calculated by Haspelmath (2015), in order to verify the finding from the typological survey that verbs with a relatively high transitivity score *and* verbs with a very low transitivity score, i.e. intransitive verbs, are most likely to show noun incorporation. Second, verb meanings representing each of the levels of the object dem/del hierarchy (Malchukov 2015; Wichmann 2015) are selected because P-incorporation is expected to pattern with other object-demoting and object-deleting alternations, such that verbs ranking high on this hierarchy would also be frequently used in P-incorporation. Thirdly, the questionnaire contains both typically patientive and agentive intransitive verbs, in order to test the idea from earlier literature (and to a certain extent supported by the typological survey) that many patientive intransitive verbs show noun incorporation, while agentive intransitive verbs do not or only rarely allow it. Fourth, most verbs found most frequently as incorporating verbs in the typological survey (listed in Table 1) are also part of the questionnaire.²¹ Finally, we take into account expectations based on other factors, namely that verbs with a very general

²⁰ We tried to balance the representativeness of the range of verb meanings in the questionnaire with general concerns of feasibility of the data collection. Therefore, we did not, for instance, include all the verb meanings from the ValPaL project.

²¹ We excluded REMOVE and PUT DOWN, since these are semantically similar to TAKE and PUT, which we did include.

meaning are often used in incorporation, as well as verbs expressing habitual activities in combination with certain incorporated nouns. In contrast, verbs that can take complement clause objects are not expected to be prone to incorporation. Verb meanings representing each of these verb types are also part of the questionnaire.

In addition to the list of verb meanings, the questionnaire includes some meta-questions on the corpus on which its answers are based, as well as some general questions about restrictions on incorporation in the relevant language, and about other verbs than the selected 47 that allow incorporation in that language. In addition, the questionnaire asks for at least one example of each incorporating verb used in a noun incorporation construction. The complete questionnaire is included as Appendix 3.

4.1.2 Language sample and data

The questionnaire was filled out by the expert authors, for one (or two, in Danielsen's case) of the eight noun-incorporating languages, as represented in Table 2.²²

The data gathered in the questionnaire-based case studies are mostly from electronic corpora of spoken language data assembled during fieldwork, typically as part of documentation projects and often in the context of language endangerment. Details about the respective language corpora can be found in Appendix 4. In some cases, the corpus data were supplemented by information from dictionaries, reference grammars, and other published sources, as well as elicitation or volunteering by native speakers, and/or personal language knowledge of the respective expert authors. Roughly, the corpora range in size between ca. 29,000 and 160,000 words and consist mostly of (spoken) narrative and conversation.²³ It goes without saying that these corpora are relatively small compared to corpora of many Indo-European and other well-studied languages often used in corpus linguistics, and this may impact the reliability of the frequency data extracted from

²² Note that the size and composition of the sample is merely a matter of convenience. The fact that four languages are from Bolivia is purely coincidental. We are not aware of any direct contact between (some of) these languages, but we cannot exclude the possibility that there are similarities between them due to areal effects. Of the languages in the questionnaire-based study, the following also figure in the 50-language sample used for the typological survey: Ese Ejja, Iraqw, Movima, and Yucatec Maya.

²³ For Baure, Ese Ejja, and Guarayu the number of words in the corpus is unknown. Information about the number of hours of recorded speech for these languages is included in Appendix 4. For the other languages included in the questionnaire-based study, the number of words in the corpus can be found there.

Table 2: Languages included in the questionnaire-based study.

Language	Glottocode	Family	Location	Expert
Baure	baur1253	Arawakan	Bolivia, Plurinational State of	Swintha Danielsen
Ese Ejja	esee1248	Pano-Tacanan	Bolivia, Plurinational State of	Marine Vuillermet
Guarayu	guar1292	Tupian	Bolivia, Plurinational State of	Swintha Danielsen
Iraqw	iraq1241	Afro-Asiatic	Tanzania, United Republic of	Tjeu Claessen and Maarten Mous
Kalamang	kara1499	West Bomberai	Indonesia	Eline Visser
Movima	movi1243	Isolate	Bolivia, Plurinational State of	Katharina Haude
Plains Cree	plai1258	Algic	Canada, United States	Arok Wolfengrey
Yucatec Maya	yuca1254	Mayan	Belize, Guatemala, Mexico	Nico Lehmann and Elisabeth Verhoeven

them. Despite this limitation, we consider the corpus-based methodology advantageous, especially because it allows for a much more systematic search for particular verbs compared to the typological survey.²⁴

Appendix 4 explains how noun incorporation constructions are identified in the sample languages. In all but one case, namely Kalamang, these identification criteria match the general definition employed in the typological survey (see Section 2.1), in the sense that there are at least some conditions under which incorporation involves the inclusion of the noun inside the inflected verbal complex. In Kalamang, noun incorporation is defined by the absence of an object-marker on the incorporated noun in combination with a phonological criterion: the noun and verb have a single prosodic contour and thus form a single phonological word. Also in other sample languages the main definition is supplemented by various additional diagnostics, which may be morphosyntactic and/or phonological in nature. In Ese Ejja, for instance, they include the lack of the *e*-marker on incorporated nouns from the *e*-class, which always take this marker when they occur independently, as well as the phonological word-status of the incorporation construction in terms of stress assignment (Vuillermet 2012: 514, 515). For concrete examples of noun incorporation constructions in all sample languages we refer to Appendix 4.

In some languages, noun incorporation constructions may be nominalized. Nominalized incorporation constructions are included in the study, except when the relevant nominalization strategy makes it impossible to verify if the relevant constructions really involve noun incorporation or not, as is the case in Iraqw. In this language patients of nominalized verbs immediately precede the nominalized

²⁴ In addition, we believe it is important to capitalize on the available resources of lesser studied languages and to stimulate collaboration between native speakers, fieldworkers, and typologists.

verb with no additional marking just as an incorporated noun would precede an inflected verb. However, this patient requires an object pronoun in the verbal complex of the inflected main verb, suggesting that the patient is a separate constituent and not an incorporated noun (Mous and Qorro 2010: 73–75). Such a criterion is absent in nominal clauses containing nominalized verbs with a patient noun. For this reason, such constructions were not counted as noun incorporation here.

On a final note, it should be stressed that not only the form, but also the function of noun incorporation may differ between the sample languages. For example, in Baure, so-called “Ground incorporation” and “classifying incorporation” portray the referent of the incorporated noun as generic, non-individuated, and backgrounded (Danielsen 2007: 99). In Ese Eja, by contrast, incorporation does not have this function, which is rather served by the antipassive construction. Incorporation in this language always involves possessed nouns and serves to promote the possessor to argument status (Vuillermet 2012: 514, 518–519). As some languages show more than one type of noun incorporation, there are also differences within languages. In fact, the different functions of noun incorporation are expected to play an important role in determining the verbs’ incorporating potential.²⁵ However, differentiating between these different functional types of incorporation must be left for a future study.

4.1.3 Method

For each of the 47 verb meanings in the questionnaire we checked whether or not the translational equivalent in each sample language appears with noun incorporation in the relevant corpus. Just as in the typological survey, not only cases of P- and Sp-incorporation but also cases of incorporation of nouns with other semantic roles are included. Appendix 4 contains the results of this query. For verb meanings with more than one translational equivalent, the translational equivalents are numbered as (i), (ii), etcetera, and data are included for each of them (see further below on the selection of translational equivalents). Those translational equivalents that are found in the relevant corpus at least once with an incorporated noun are counted as incorporating verbs in the corresponding language. The frequency of occurrence in the corpus of these incorporating verbs, both with and without noun incorporation, is also given in

²⁵ An anonymous reviewer points out that different functions of incorporation, especially syntactic versus semantic ones, also influence the type of incorporated nouns. For instance, in Ese Eja, where incorporation serves the syntactic function of possessor raising, body part nouns are preferentially incorporated.

Appendix 4. Translational equivalents that are not attested with incorporation in the corpus of a language are interpreted as non-incorporating verbs in the relevant language. Of course, however, the absence of noun incorporation with a particular verb in a (relatively small) corpus does not prove that noun incorporation is impossible. Therefore, we also included, as much as possible, information from published sources and native speakers (see Section 4.1.2), in order to verify whether or not noun incorporation is (im)possible for a verb. For verbs that are not found in the corpus at all (neither with nor without an incorporated noun) and for which additional sources are not conclusive either, the question whether or not they allow noun incorporation cannot be answered and these cases are treated as missing data points. Finally, for some verb meanings no verbal translational equivalent was found in one or more of the sample languages. In these cases the question whether the verb allows noun incorporation is irrelevant, and these cases are also considered missing data points.

One further methodological issue is important to interpret the results of the case studies and their comparison. As pointed out by Haspelmath and Hartmann (2015: 51–53), finding translational equivalents of particular verb meanings is not always a trivial matter. Mosel (in prep.) illustrates this problem in the context of a corpus investigation similar to ours, focussing on the Austronesian language Teop and using the verb *CUT* as an example. Looking for translational equivalents of *CUT* in this language, she finds the following: (i) Teop words translated by “cut”; (ii) Teop words that in addition to “cut” are translated by other English words (e.g. “carve, cut, shave”); (iii) Teop words that are *not* translated by “cut”, but by a word that shares an English translation with another Teop word that is also translated by “cut” (e.g. “shave”). Our study is mostly limited to cases like (i). In addition, when there is a choice within such cases, we choose the verb with the semantically least specific translation. Thus, if we find, for instance, two verbs glossed as “cut” and “cut with knife”, respectively, we choose the former. Only if we cannot make a motivated choice between two alternative verbs, for instance between “cut with knife” and “cut with machete”, which are equally semantically specific, we take both (or in rare cases all three or four) verbs into account. Cases with multiple translations, as in Mosel’s scenario (ii), are considered only when there is no candidate of type (i). Cases of type (iii) are not considered at all: a verb glossed as “shave” would not be counted under *CUT*, but rather under *SHAVE*. Despite this procedure, however, decisions were not always straightforward to make. In Appendix 4 we therefore provide the verbs we chose as translational equivalents for the verb meanings in the questionnaire in each of the sample languages.

4.2 Results and discussion

4.2.1 Frequently incorporating verbs

Table 3 shows the 47 verb meanings included in the questionnaire, ordered according to their incorporation scores across the eight sample languages: the higher the incorporation score, the more frequently the verb meaning's translational equivalents are used in noun incorporation in the data from the eight sample languages. The data on which Table 3 is based are included in Appendix 5. Note that those verb meanings that were found to be noun-incorporating in 10 or more languages in the typological survey are presented in bold (cf. Table 1).

For each verb meaning, the cross-linguistic incorporation score is the average of the language-specific incorporation scores for this verb meaning. These language-specific incorporation scores can have one of three values: 1 if the only or all translational equivalents of the relevant verb meaning allow noun incorporation; 0 if the only or all translational equivalents of the verb meaning do not allow noun incorporation; and 0.5 if there is at least one translational equivalent that allows noun incorporation and at least one that does not allow noun incorporation. No language-specific score was assigned to a verb meaning for which the language does not have any translational equivalents or for which it is unclear if the translational equivalent(s) allow noun incorporation. The cross-linguistic incorporation scores were calculated by dividing the sum of the scores for a particular verb across the eight languages by the number of languages for which a score could be calculated, resulting in a cross-linguistic incorporation score between 0 and 1.

As mentioned, to allow for a comparison between the findings from our questionnaire-based case studies and the findings from the typological survey, in Table 3 the verbs that were found to allow noun incorporation in at least 10 out of the 50 languages in the typological sample are indicated in bold-face. Most of these verbs appear at the top of the table, which shows that many of the verb meanings that are found frequently as incorporating verbs in the typological survey are also among the most frequently incorporating verbs in the eight languages studied on the basis of the questionnaire and corpus data. The only exception to this general pattern involves the verb GO, which has a quite low cross-linguistic incorporation score and correspondingly appears near the bottom of Table 3. In general, the results of the typological survey are thus quite comparable to those of the questionnaire-based case studies.

In order to evaluate to what extent the ranking of verbs in terms of their cross-linguistic ability to incorporate can be considered a statistically implicational hierarchy, just as the verbs in the object dem/del hierarchy, we applied the method developed by Wichmann (2015, 2016; Aldai and Wichmann 2018) to determine its

Table 3: Verbs included in the questionnaire ordered based on their cross-linguistic incorporation scores calculated on the basis of the data from the eight languages.

Verb meaning	Cross-linguistic incorporation score
WASH	1.00
CUT	0.94
EAT	0.88
CATCH	0.79
KILL	0.79
FEEL PAIN	0.75
HAVE	0.75
THROW	0.75
BE DRY	0.71
PUT	0.71
BREAK (TR.)	0.69
HIT	0.69
SHAVE	0.67
GIVE	0.63
SEARCH FOR	0.63
COVER	0.56
ASK FOR	0.50
COOK	0.50
FALL	0.50
HUNT	0.50
TAKE	0.50
BUY	0.43
TELL	0.43
FEEL COLD	0.40
MAKE/DO	0.38
SEE	0.38
FRIGHTEN	0.33
SIT DOWN	0.33
SIT	0.31
DIE	0.25
JUMP	0.25
NAME	0.25
SING	0.25
RUN	0.21
SAY	0.21
BE SAD	0.20
SINK (INTR.)	0.20
BURN (INTR.)	0.13
KNOW	0.13
LEAVE	0.13
PLAY	0.13
GO	0.06

Table 3: (continued)

Verb meaning	Cross-linguistic incorporation score
BE HUNGRY	0
FEAR	0
HELP	0
MEET	0
THINK	0

Guttman's coefficient. The coefficient is 81.98, which is below the conventional 85%. Also the p-value is not statistically significant: $p = 0.15$.²⁶ This means that the eight languages in our sample are not sufficient to determine a reliable ordering of the verbs with some form of implicational power. This result can be visualized with NeighborNet (Huson and Bryant 2006), as in Figure 1: While, as expected, the verbs that have comparable cross-linguistic incorporation scores appear in the same areas of the tree, they are connected by boxes rather than lines, indicating non-treelike or non-implicational behaviour.²⁷ Yet, as will be discussed in the next subsection, the ordering of the verbs in our study does correlate with the verb rankings found in the ValPaL project.

4.2.2 Morphosyntactic transitivity

The cross-linguistic incorporation scores of the verb meanings shown in Table 3 are largely in line with the expectations about which verbs are likely to incorporate

²⁶ We thank Søren Wichmann for helping us with this calculation, which was carried out using his software at <https://github.com/Sokiwi/Guttman>. It is important to realize that the Guttman's coefficient can only be calculated on the basis of a binary (1 for "yes" or 0 for "no") value, in this case of incorporation potential. Therefore, we transformed the data on which the verb ordering in Table 3 is based, such that all language-specific incorporation scores of 0.5 were changed to 1 scores (to indicate that at least one out of multiple translational equivalents analysed for a specific verb meaning is able to incorporate). While this does not influence the overall ordering of the verb very strongly, there are some differences, as can be seen when comparing the levels of verbs in Table 3 with those based exclusively on binary values: WASH, CUT – EAT, HIT – CATCH, KILL – FEEL PAIN – BREAK (TR.), GIVE, HAVE, THROW – BE DRY, PUT – SHAVE – COVER, SEARCH FOR – ASK FOR, COOK, FALL, HUNT, TAKE – BUY, TELL – FEEL COLD – MAKE/DO, SEE, SIT – FRIGHTEN, SIT DOWN – RUN, SAY – DIE, JUMP, NAME, SING – BE SAD, SINK (INTR.) – BURN (INTR.), GO, KNOW, LEAVE, PLAY – BE HUNGRY, FEAR, HELP, MEET, THINK. Especially for the verb meaning HIT three languages have multiple translational equivalents with language-internally distinct values for incorporation potential. These transformed data were also used to create the NeighborNet visualization in Figure 1.

²⁷ We thank Alena Witzlack-Makarevich for creating the NeighborNet visualization.

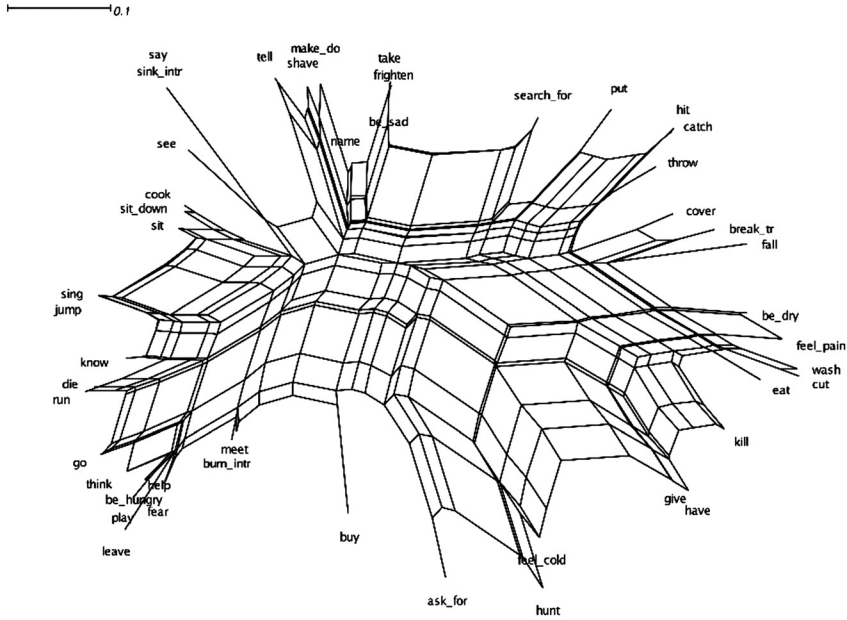


Figure 1: NeighborNet visualization of the incorporation ability of the verbs in the eight sample languages.

based on their transitivity prominence scores as calculated by Haspelmath (2015) and based on their position on the object dem/del hierarchy (Malchukov 2015; Wichmann 2015). The figures in Table 4 show that, as expected, most verbs with high cross-linguistic incorporation scores also have high transitivity prominence scores. Note that this table is an adapted version of Table 3, including only the verb meanings of our questionnaire that overlap with Haspelmath's (2015) study, with their transitivity prominence scores given in parentheses.

The Spearman's rank order coefficient for the ranking of the verb meanings based on their cross-linguistic incorporation scores and their ranking based on their transitivity prominence scores shows that there is a moderate, but statistically significant positive correlation between these rankings: $\rho = 0.56$ ($p < 0.001$). It can be seen in Table 4 that verbs with cross-linguistic incorporation scores of 0.50 and higher have transitivity prominence scores that range between 0.88 and 1.00, with the exceptions of FEEL PAIN and BE DRY. These latter two verb meanings have very low transitivity scores; hence, they can be regarded as intransitive verbs cross-linguistically and seem to favour incorporation for that reason. FEEL COLD and DIE are two other verbs that belong to the patientive intransitive class, with transitivity prominence scores of 0.00, showing noun incorporation with a considerable cross-

Table 4: Verb meanings included in the questionnaire and studied by Haspelmath (2015) ordered based on their cross-linguistic incorporation score. Transitivity prominence scores are given between parentheses.

Verb meaning	Cross-linguistic incorporation score
WASH (0.94)	1.00
CUT (1.00)	0.94
EAT (0.93)	0.88
KILL (1.00)	0.79
FEEL PAIN (0.12)	0.75
THROW (0.98)	0.75
BE DRY (0.00)	0.71
PUT (0.98)	0.71
BREAK (TR.) (1.00)	0.69
HIT (1.00)	0.69
SHAVE (0.93)	0.67
GIVE (0.98)	0.63
SEARCH FOR (0.88)	0.63
COVER (0.95)	0.56
ASK FOR (0.95)	0.50
TAKE (1.00)	0.50
TELL (0.78)	0.43
FEEL COLD (0.00)	0.40
SEE (0.93)	0.38
FRIGHTEN (0.98)	0.33
SIT DOWN (0.03)	0.33
SIT (0.05)	0.31
DIE (0.00)	0.25
JUMP (0.00)	0.25
NAME (0.80)	0.25
SING (0.38)	0.25
RUN (0.05)	0.21
SAY (0.41)	0.21
BE SAD (0.00)	0.20
SINK (INTR.) (0.03)	0.20
BURN (INTR.) (0.00)	0.13
KNOW (0.88)	0.13
LEAVE (0.42)	0.13
PLAY (0.10)	0.13
GO (0.05)	0.06
BE HUNGRY (0.00)	0
FEAR (0.53)	0
HELP (0.78)	0
MEET (0.70)	0
THINK (0.52)	0

linguistic frequency across the sample languages. By contrast, some other members of the patientive intransitive class, such as BE SAD, SINK (INTR.), BURN (INTR.), and BE HUNGRY, appear unexpectedly near the bottom of the table.

The agentive intransitive verbs with a very low transitivity prominence score are generally found to be unlikely to incorporate. This finding is expected based on the preference for incorporation into unaccusative rather than unergative verbs, identified in earlier work (Baker 1988: 81–82, 87–90).²⁸ Thus, JUMP, RUN, PLAY, and GO have transitivity prominence scores of 0.10 and lower and are only very rarely used as incorporating verbs in our data. Note also that, when they do incorporate, they incorporate goals or instruments rather than agents. The only agentive intransitive verb that has a relatively high cross-linguistic incorporation score is SIT DOWN. However, the examples of incorporation into verbs meaning SIT DOWN involve an incorporated goal or an incorporated body-part noun, with its possessor expressed as the subject of the construction, rather than an incorporated agent argument.

Finally, there are a few verbs with fairly high transitivity prominence scores of 0.70 and higher that nonetheless have quite low cross-linguistic incorporation scores of 0.25 or lower. These include NAME, KNOW, HELP, and MEET. We will come back to these verb meanings in the next subsection.

Turning to the relation between our data and the object dem/del hierarchy, consider (10) below, where we represent the verbs from this hierarchy that are also part of the present study, with their cross-linguistic incorporation scores included in parentheses. These figures indicate that many of the verbs that are likely to incorporate based on our questionnaire data also rank high on the object dem/del hierarchy, while many of the verbs that do not incorporate frequently according to our data have a correspondingly low position on this hierarchy. The primary exceptions to these patterns are, as expected, some patientive intransitive verbs like BE DRY and FEEL PAIN, which are quite likely to incorporate but rank low on the general object dem/del hierarchy. Also as expected, the agentive intransitive verbs (e.g. RUN, PLAY) are typically low-ranking on both accounts.

- (10) Object dem/del hierarchy (adapted from Malchukov 2015: 105–106; Wichmann 2015: 166–167). Only the verbs that overlap with our questionnaire are included. The cross-linguistic incorporation scores of the verbs are included between parentheses.

EAT (0.88), WASH (1.00), GIVE (0.63) >
 SHAVE (0.67), COOK (0.50) >
 CUT (0.94), SEARCH FOR (0.63), HIT (0.69) >

²⁸ Notably, Baker (1988) claims that unergative verbs cannot incorporate at all. Here, we rather look at semantically defined agentive and patientive intransitive verbs.

KILL (0.79), ASK FOR (0.50), TAKE (0.50) >
 SEE (0.38), THROW (0.75) >
 BREAK (TR.) (0.69), COVER (0.56), THINK (0.00) >
 TELL (0.43), KNOW (0.13), HELP (0.00) >
 SING (0.25) >
 FEAR (0.00) >
 PUT (0.71), LEAVE (0.13) >
 SAY (0.21), NAME (0.25), RUN (0.21) >
 JUMP (0.25), FRIGHTEN (0.33), PLAY (0.13), BE DRY (0.71) >
 BURN (INTR.) (0.13), GO (0.06), SINK (INTR.) (0.20) >
 MEET (0.00), DIE (0.25) >
 FEEL PAIN (0.75), SIT (0.31) >
 BE SAD (0.20) >
 SIT DOWN (0.33), BE HUNGRY (0.00) >
 FEEL COLD (0.40)

The Spearman's rank order coefficient for the ranking of the verbs based on their cross-linguistic incorporation score and their ranking on the object dem/del hierarchy (Malchukov 2015; Wichmann 2015) shows a moderate but statistically significant positive correlation for these rankings ($\rho = 0.54$, $p < 0.001$), just as for the ranking of the verbs based on their cross-linguistic incorporation score and their ranking in terms of transitivity prominence. This is expected given the high correlation between the object dem/del hierarchy and the ranking of the verbs based on their transitivity prominence score reported in Section 2.2.

4.2.3 Other factors

The data summarized in Table 4 and in (10) above suggest that a few factors other than morphosyntactic transitivity also affect the likelihood that a verb allows noun incorporation. Firstly, verbs like KNOW, THINK, and SAY have relatively low cross-linguistic incorporation scores compared to other verbs with similar transitivity prominences scores and on similar positions on the object dem/del hierarchy. This finding matches the results for these verbs in the typological survey (see Section 3.2.3) and it suggests again that these verbs do not incorporate frequently because their P-arguments may be complement clauses rather than nouns. Secondly, for the verbs NAME, HELP, MEET, and, to a lesser extent, LEAVE and FEAR, which have relatively low cross-linguistic incorporation scores even though they have quite high transitivity prominence scores and appear quite high on the object dem/del hierarchy, it may be noted that their meanings imply a P-argument which is typically human or at least animate. As discussed in Section 2.2.1, it has been claimed that verbs that tend

to take animate patient arguments are not very likely to incorporate, and animacy may thus also play a role for these verbs. Thirdly, it can be seen in (10) that PUT appears in the lower half of the object dem/del hierarchy but has a quite high cross-linguistic incorporation score. As mentioned in Section 3.2.3, this finding matches the outcome of the typological survey and corroborates Mithun's (1984: 863) observation that verbs with general semantics are likely to allow incorporation. Fourthly, it has often been argued that body-part nouns are frequently incorporated, and indeed we find that many verbs that often involve body-part nouns, like FEEL PAIN and FEEL COLD, incorporate relatively frequently.

In addition to these morphosyntactic and semantic factors, to a considerable extent the data appear to reflect unmotivated, idiosyncratic lexical restrictions. For instance, in some cases of multiple translational equivalents for a verb meaning in a language, the relevant verbs do not behave the same in terms of their incorporation potential, despite their semantic near-equivalence. This can be illustrated with the three verbs for CUT in Iraqw: *tsaat* 'cut in one movement', *tlaaq* 'cut (pluractional)', and *siik* 'cut in sections'. The first two of these do allow incorporation (although to slightly different degrees; see further below), but the third one does not, according to our data. This example thus shows that verb-based restrictions on noun incorporation in a particular language can be lexically based without any obvious underlying semantic or syntactic regularity.

Interestingly, information about the relative frequency with which verbs appear in incorporation cross-linguistically also reflects some of these additional morphosyntactic, semantic and lexical factors. So far, we have considered the ordering of the verbs in our questionnaire based on their absolute incorporation potential across the sample languages, i.e. on whether or not they are found to allow incorporation, irrespective of the relative frequency with which the verbs actually appear in incorporation constructions in the corpora. Yet, we may also focus on the question how often verbs that allow incorporation actually appear with incorporated nouns. For each verb meaning, we added up all attested occurrences of each translational equivalent that allows noun incorporation across languages, and all instances of incorporation constructions with these translational equivalents across languages. A cross-linguistic incorporation ratio can be calculated for each verb meaning by dividing the frequency of the incorporation constructions by the total frequency of these verbs (see also Appendix 5).²⁹ While it

²⁹ Note that the non-incorporated constructions form an internally rather heterogeneous group. For instance, for Kalamang we found that, besides "regular" transitive constructions with morphologically marked independent object nouns, many of the non-incorporating occurrences of the verb *na* "eat"/"drink" involved constructions with an unexpressed (but pragmatically implied) P-argument.

is hard to interpret a verb scale based on these ratios, since some of the ratios are based on far fewer data points than others, some general observations still seem worthwhile.

Some verbs, such as *EAT* and *KILL*, allow noun incorporation in many of the sample languages, as shown by their cross-linguistic incorporation scores presented in Table 4, have high transitivity prominence scores, and appear high on the object dem/del hierarchy, but have quite low cross-linguistic incorporation ratios. These verbs thus allow incorporation in many sample languages but show incorporation in these languages only rarely. By contrast, other verbs that show incorporation in many languages, like *WASH* and *CUT*, also have relatively high actual incorporation ratios. Possibly, the combination with particular (types of) nouns plays a role here. For example, *WASH* may be prone to incorporate body-part nouns, while *KILL* typically has an animate patient, which may explain its relative resistance to incorporation. Note also that it is possible that some verbs with low cross-linguistic incorporation ratios but high ranks on the object dem/del hierarchy, such as *EAT*, often omit their P-argument completely rather than incorporate it, because they express actions that always involve a particular P-argument (Wichmann 2015: 167).

Conversely, some verbs that have low cross-linguistic incorporation scores, i.e. that show incorporation in the corpora of few languages, nevertheless have high cross-linguistic incorporation ratios, because in the languages in which their translational equivalents *do* allow incorporation these verbs incorporate frequently. For example, *JUMP* is found to incorporate in Baure and Movima only, but does so with fairly high or even very high frequency: *JUMP* shows incorporation in 14 out of the 30 occurrences in Baure and in all of the 21 occurrences in Movima, yielding a cross-linguistic incorporation ratio of 68.63%. This is different for other verbs, which are low-ranking in both absolute and relative terms. For instance, *PLAY* has a low cross-linguistic incorporation score because it is found to incorporate in Yucatec Maya only, and it has an incorporation ratio of only 10.20% because it shows incorporation in only 5 out of the 49 times it occurs in the corpus.

Finally, there are verb meanings whose translational equivalents seem to display widely varied behaviour across, but also within, languages, suggesting idiosyncratic restrictions. As an example of cross-linguistic variation, consider *FALL*. This verb is found to incorporate in four languages: Baure, Movima, Plains Cree, and Yucatec Maya. In Baure and Movima it does so every time it occurs in the corpus, i.e. 45 out of 45 cases for Baure and 98 out of 98 cases for Movima. By contrast, although *FALL* can incorporate in Plains Cree and Yucatec Maya (as attested on the basis of dictionary data and speaker elicitation), none of the 24 occurrences of this verb in the Plains Cree corpus and none of the 72 occurrences of this verb in the Yucatec Maya corpus actually involves incorporation. As for intra-

linguistic variation, recall that for some verb meanings in some languages we considered multiple translational equivalents. We have already mentioned that such near-synonyms may show differential behaviour in terms of their absolute (in)ability to incorporate. Such differences may also show up in relative terms when two translational equivalents can both incorporate, but do so to different degrees. For instance, in Guarayu the verbs *-nupa* ‘hit (hard)’ and *-ipete* ‘hit’ show incorporation in 1 out of 11 and 6 out of 7 cases, respectively. It must be emphasized again, however, that in many cases the low absolute numbers prevent meaningful comparisons and firm conclusions.

4.2.4 Variation across languages

The previous subsections primarily discussed the attested variation from the perspective of the different verbs, but we may also focus on variation across languages. As Appendix 5 shows, the eight languages can be ordered in terms of their language-specific incorporation score, i.e. the number of verbs attested with incorporation, relative to the total number of verb meanings in that language for which data points are available. These scores lead to the ranking in (11).

- (11) Language ranking based on incorporation scores across verbs.
 Guarayu (0.59) > Movima (0.51) > Kalamang (0.45) > Baure (0.43) >
 Yucatec Maya (0.40) > Ese Eja (0.38) > Iraqw (0.32) > Plains Cree (0.31)

Based on the available frequency information, language-specific incorporation ratios across verbs can be calculated as well, by dividing all instances of incorporation found in the corpus of a language across all incorporating verbs in that language by the total number of occurrences of these verbs taken together. According to these language-specific incorporation ratios, which are also included in Appendix 5, the ordering of the languages is as follows:

- (12) Language ranking based on incorporation ratios across verbs.
 Movima (43.21%) > Baure (19.99%) > Guarayu (13.38%) > Ese Eja
 (13.03%) > Plains Cree (13.00%) > Kalamang (8.16%) > Yucatec Maya
 (4.88%) > Iraqw (3.61%)

Before discussing these two language scales in more detail, it is important to realize that (i) they are obviously based only on information about verbs corresponding to verb meanings in our questionnaire, even though we know that (many) other verbs also allow incorporation in the sample languages; (ii) the percentages mentioned in (12) are sometimes skewed due to small absolute numbers. For example, in Baure there are five verbs that show incorporation in all their occurrences in the corpus. However, for three of these verbs, the total

numbers of occurrence are extremely low (1, 2, and 3 respectively). More generally, verbs that are obligatorily incorporating can distort the overall picture.

Comparing now the scales in (11) and (12), we observe both similarities and differences. In terms of similarities, Guarayu and Movima occupy the higher end in both cases, while Iraqw is at the lower end of both scales, and Ese Eja appears in a middle position on both scales. In contrast, Kalamang and Plains Cree behave quite differently in absolute (11) and relative (12) terms. Kalamang ranks fairly high on the absolute scale, but many of its incorporating verbs do not incorporate very frequently. Conversely, in Plains Cree relatively few of the verbs included in the questionnaire show noun incorporation, but those verbs that *can* incorporate do so quite frequently. In general, these results highlight the fact that there are different ways to measure the pervasiveness or productivity of incorporation in a given language and to compare these measurements between languages.

Incidentally, the data also suggest that there is no clear effect of what may be called “degree of synthesis”. Much research considers noun incorporation as a characteristic of polysynthetic languages (see Genee 2018: 243 for an overview). Nevertheless, the sample of noun-incorporating languages used for the questionnaire-based case studies includes both languages that are generally considered polysynthetic, such as Plains Cree (Bakker 2006: 5; Hirose 2003: 8), and languages that are rather analytic, such as Kalamang. Moreover, these two languages do not show any principled differences in terms of their number of incorporating verbs nor in the incorporation frequency relative to the number of incorporating verbs.

5 Discussion and conclusion

This study aimed to find out which verbs are most (un)likely to incorporate nouns across languages, and to shed light on the extent to which verb-based restrictions on incorporation are determined by transitivity, potentially in combination with other factors. This section summarizes our main empirical findings and interprets them in a wider theoretical context.

First of all, the results of the two sub-parts of the study, i.e. the typological survey and the questionnaire-based case studies, show considerable overlap and largely confirm the predictions based on earlier studies reviewed in Section 2.2.2: Morphosyntactically highly transitive verbs and – to a lesser extent – patientive intransitive verbs are cross-linguistically most likely to incorporate, while agentive intransitive verbs are unlikely to do so.

Furthermore, although in the questionnaire-based part of the study both the sample size and the size of the individual language corpora is relatively restricted, the ranking of the verbs based on their likelihood to incorporate is found to

correlate moderately with Haspelmath's (2015) transitivity prominence scale, as well as with Malchukov's (2015) and Wichmann's (2015) object dem/del hierarchy. These correlations further indicate that morphosyntactic transitivity affects the likelihood that a verb is able to incorporate nouns. The fact that our verb ranking does not reach significance in the sense of having implicational power is likely due to the above-mentioned methodological limitations of this study. More generally, however, implicational generalizations regarding the verbs' incorporation potential appear to be modulated by lexical idiosyncrasies (see further below).

In addition to morphosyntactic transitivity, a number of other factors are shown to influence verbal restrictions on incorporation, partly corroborating earlier claims to this effect. In particular, promoting factors for incorporation include high generality of verbal semantics and habituality of the action denoted by a verb plus its incorporated noun. Inhibiting factors, on the other hand, are certain subcategorization properties, namely for animate patients and for clausal complements. Together, these factors point to a tight link between the lexical meaning of a verb, its degree of morphosyntactic transitivity, and the properties of its patient argument, rather than its agent argument. Along the same lines as has been argued for other voice and valency-related constructions such as the anti-passive (cf. *Say in prep.*), the importance of the patient argument for the interpretation of the entire event may explain why a verb's general transitivity and potentially other patient-related semantic properties co-regulate its ability to participate in incorporation alternations across languages.

As discussed in the introduction to this paper, research on noun incorporation has often focused on the question if noun incorporation is a lexical or a syntactic process. Our study shows that, while there may certainly be languages in which incorporation is fully productive and hence arguably syntactic in nature, in general verbs vary in whether or not they allow incorporation and, if they do, in how often they occur in incorporation constructions. In terms of the interaction between a verb's lexical semantics, its morphosyntactic transitivity (which presumably reflects its semantic transitivity), and its incorporation potential there clearly are recurrent patterns, as summarized and tentatively explained above. However, we also see that there are lexical idiosyncrasies, i.e. cases where semantically quite similar verbs, possibly within one language, differ in their incorporation behaviour either in absolute (ability) or in relative (frequency) terms. Moreover, languages differ in terms of the size and composition of their class of incorporating verbs, as well as in the mean frequency with which these verbs actually occur in incorporation constructions. Our corpus-based method allows us to discern these different perspectives on the verb-based degree of productivity of incorporation within and across languages.

In sum, this study provides evidence both for idiosyncratic effects of individual verbs and for higher-level semantico-syntactic transitivity effects on the cross-linguistic incorporation potential of verbal meanings. This conclusion ties in with the more general idea, advanced by corpus-based studies taking a usage-based approach, that argument-structure constructions involve a combination of rather abstract, schematic representations (Goldberg 1995) *and* pervasive “mini-constructions” (Boas 2003: 22), i.e. verb-specific patterns, which may also include statistical “preferences” to occur in one or another variant of a valency or voice alternation (Diessel 2019: 119–121). Thus, our study highlights the importance of lexically fine-grained analysis across multiple and structurally diverse languages, using spoken language corpora. In the future, we would like to further investigate cross-linguistic variation in argument-structure constructions, in particular the relation between lexically specific patterns, semantically-based generalizations, and absolute syntactic rules. In order to do this, larger data sets, both in terms of sample languages and in terms of corpus size, will be indispensable.

Abbreviations in glosses

1 = first person; 3 = third person; A = agent-like argument of canonical transitive verb; ABS = absolutive; E = epenthesis; EMPH = emphatic; ERG = ergative; IND = indicative; INF = infinitive; P = patient-like argument of canonical transitive verb; PL = plural; PST = PAST; S = single argument of canonical intransitive verb; SG = singular.

Acknowledgments: We thank the members of the ACLC research group Language Description and Typology, as well as the audience of Syntax of the World’s Languages 8 (Paris, September 2018), Søren Wichmann, Jan Don and Kees Hengeveld for valuable comments on earlier versions of this paper. Olthof’s contribution to this work is funded by the research programme PhDs in the Humanities, project number 322-70-011 of the Netherlands Organisation for Scientific Research (NWO).

Appendix 1: Language sample and sources typological survey

This table includes the names, glottocodes, and family classifications as presented in Glottolog (Hammarström et al. 2017), and data sources of the sample languages. Alternative names for the languages used in the data sources for the particular languages are included in square brackets.

Language	Glottocode	Family	Source(s)
Alamblak	alam1246	Sepik	Bruce (1984: 170–176, 370)
Atsugewi	atsu1245	Palaihnihan	Palancar (1999: 155, 162)
Bininj Kun-Wok [Bininj Gun-wok]	gunw1252	Gunwinyguan	Evans (2003: 323–336, 450–485)
Caddo	cadd1256	Caddoan	Melnar (1998)
Chimalapa Zoque [San Miguel Chimalapa Zoque]	chim1300	Mixe-Zoque	Johnson (2000: 268–278)
Chukchi	chuk1273	Chukotko-Kamchatkan	Spencer (1995), Dunn (1999: 221–238)
Crow	crow1244	Siouan	Graczyk (2007: 277–299)
Eastern Ojibwa [Central Ojibwa, Ojibwe]	east2542	Algic	Rhodes (1976: 260–285), Barrie and Mathieu (2016: 12–23)
Ese Ejja	esee1248	Paco-Tacanan	Vuillermet (2012: 369, 514–519, 562–570; 2014: 114–138)
Guahibo [Sikuani]	guah1255	Guahibo	Queixalós (2008)
Haida ³⁰	haid1248	Haida	Enrico (2003: 1263–1267)
Halkomelem [Musqueam]	halk1245	Salishan	Gerdts (2003), Suttles (2004: 269–278, 289–290), Gerdts and Hukari (2008)
Hokkaido Ainu [Ainu, Southern Hokkaido Ainu]	ainu1240	Ainu	Shibatani (1990: 61–64), Bugaeva (2012: 489–492; 2017: 897–903)
Iraqw	iraq1241	Afro-Asiatic	Mous (1993: 200–202, 257–259), Kooij and Mous (2002: 633–637)
Kalaallisut [West Greenlandic]	kala1399	Eskimo-Aleut	Fortescue (1980: 274–275)
Karajá	kara1500	Nuclear-Macro-Je	Rivail Ribeiro (2001: 239–242)
Ket	kett1243	Yeniseian	Drossard (2002: 235–238), Maksunova (2003), Vajda (2017: 911–922)
Malayo [Damana]	mala1522	Chibchan	Quesada (1999: 248)
Mamaindé [Mamaindé]	mama1278	Nambiquaran	Eberhard (2009: 357–358, 394, 397–403)
Mandinka	mand1436	Mande	Creissels (2008: 82–87)
Mapudungun [Mapuche]	mapu1245	Araucanian	Harmelink (1992), Zúñiga (2000: 40–41, 2017: 702–705), Baker et al. (2005), Smeets (2008: 318–319)
Marithiel [Marrihiyel]	mari1424	Western Daly	Green (1989: 221–279)
Mohawk [Akwesasne Mohawk]	moha1258	Iroquoian	Mithun (1984: 867–872, 2000), Baker (1988: 76–146, 1996: 279–337), Bonvillain (1989a, 1989b)

³⁰ Haida is considered a language family rather than a single language in Glottolog (Hammarström et al. 2017).

(continued)

Language	Glottocode	Family	Source(s)
Movima	movi1243	Movima (Isolate)	Haude (2006: 367–392)
Murui Huitoto [Uitoto]	murui1274	Huitotoan	Petersen de Piñeros (2007: 405–407)
Nadëb	nade1244	Nadahup	Weir (1990: 324–327)
Nisga'a [Nisgha]	nisg1240	Tsimshian	Tarpen (1987: 784, 791–799)
Niuean	niue1239	Austronesian	Seiter (1980: 69–78), Massam (2001)
Northeast Maidu [Maidu]	nort2952	Maiduan	Palancar (1999: 162)
Northern Gumuz	gumu1244	Gumuz	Ahland (2012: 245–265)
Nuu-chah-nulth [Nuuchahnulth]	nuuc1236	Wakashan	Stonham (2004: 213–234, 2008), Wojdak (2005: 105–156)
Palikúr [Palikur]	pali1279	Arawakan	Aikhenvald and Green (1998: 451– 456)
Panamint [Tümpisa (Panamint) Shoshone]	pana1305	Uto-Aztecan	Dayley (1989: 40–41, 90–103)
Panare	enap1235	Cariban	Payne (1995), Payne and Payne (2013: 91–96, 178–179, 330–333)
Paraguayan Guaraní [Guaraní]	para1311	Tupian	Velázquez-Castillo (1995)
Sora	sora1254	Austroasiatic	Anderson (2007: 186–195, 2017: 937–946), Anderson and Harrison (2008: 351–357)
South Slavey [Slave]	sout2959	Athapaskan- Eyak-Tlingit	Rice (1991: 54–59, 2008: 385–388)
Southern Tiwa	sout2961	Kiowa-Tanoan	Allen, Gardiner and Frantz (1984)
Takelma	take1257	Takelma (Isolate)	Palancar (1999: 160)
Tanimuca-Retuarã [Retuarã]	tani1257	Tucanoan	Strom (1992: 100–101)
Tiwi	tiwi1244	Tiwi (Isolate)	Osborne (1974: 47–50)
Ute-Southern Paiute [Southern Paiute, Ute]	utes1238	Uto-Aztecan	Sapir (1930: 84–87), Givón (2011: 57, 159–161)
Warembori	ware1253	Austronesian	Donohue (1999: 43–46)
Washo	wash1253	Washo (Isolate)	Bochnak and Rhomieux (2013)
Western Frisian [Frisian]	west2354	Indo-European	Dijk (1997: 1–164)
Western Highland Chatino [Yaitepec Chatino]	west2644	Otomanguean	Rasch (2002: 129–135)
Yele	yele1255	Yele (Isolate)	Henderson (1995: 26–28)
Yimas	yima1243		Foley (1991: 319–321, 2017)

(continued)

Language	Glottocode	Family	Source(s)
		Lower Sepik- Ramu	
Yucatec Maya	yuca1254	Mayan	Sullivan (1984), Lehmann and Verhoeven (2005: 143–172)
Zuni	zuni1245	Zuni (Isolate)	Miner (1986: 251–253)

Full references to sources listed in Appendix 1

- Ahland, Colleen. 2012. *A grammar of Northern and Southern Gumuz*. Eugene, OR: University of Oregon dissertation.
- Aikhenvald, Alexandra Y. & Diana Green. 1998. Palikur and the typology of classifiers. *Anthropological Linguistics* 40(3). 429–480.
- Allen, Barbara J., Donna B. Gardiner & Donald G. Frantz. 1984. Noun incorporation in Southern Tiwa. *International Journal of American Linguistics* 50(3). 292–311.
- Anderson, Gregory D. S. 2007. *The Munda verb: Typological perspectives* (Trends in Linguistics: Studies and Monographs 174). Berlin: Mouton de Gruyter.
- Anderson, Gregory D. S. 2017. Polysynthesis in Sora (Munda) with special reference to noun incorporation. In Michael Fortescue, Marianne Mithun & Nicholas Evans (eds.), *The Oxford handbook of polysynthesis* (Oxford Handbooks in Linguistics), 930–947. Oxford: Oxford University Press.
- Anderson, Gregory D. S. & K. David Harrison. 2008. Sora. In Gregory D. S. Anderson (ed.), *The Munda languages* (Routledge Language Family Series), 299–380. London: Routledge.
- Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago, IL: The University of Chicago Press.
- Baker, Mark C. 1996. *The polysynthesis parameter* (Oxford Studies in Comparative Syntax). New York, NY: Oxford University Press.
- Baker, Mark C., Roberto Aranovich & Lucía A. Golluscio. 2005. Two types of syntactic noun incorporation: Noun incorporation in Mapudungun and its typological implications. *Language* 81(1). 138–176.
- Barrie, Michael & Eric Mathieu. 2016. Noun incorporation and phrasal movement. *Natural Language & Linguistic Theory* 34(1). 1–51.
- Bochnak, M. Ryan & Alice Rhomieux. 2013. Limited noun incorporation in Washo. *International Journal of American Linguistics* 79(2). 253–281.
- Bonvillain, Nancy. 1989a. Noun incorporation and metaphor: Semantic process in Akwesasne Mohawk. *Anthropological Linguistics* 31(3/4). 173–194.
- Bonvillain, Nancy. 1989b. Body, mind, and idea: Semantics of noun incorporation in Akwesasne Mohawk. *International Journal of American Linguistics* 55(3). 341–358.
- Bruce, Les. 1984. *The Alambalak language of Papua New Guinea (East Sepik)* (Pacific Linguistics: C81). Canberra: Research School of Pacific Studies, The Australian National University.
- Bugaeva, Anna. 2012. Southern Hokkaido Ainu. In Nicolas Tranter (ed.), *The languages of Japan and Korea* (Routledge Language Family Series), 461–509. London: Routledge.
- Bugaeva, Anna. 2017. Polysynthesis in Ainu. In Michael Fortescue, Marianne Mithun & Nicholas Evans (eds.), *The Oxford handbook of polysynthesis* (Oxford Handbooks in Linguistics), 882–905. Oxford: Oxford University Press.

- Creissels, Denis. 2008. L'incorporation en Mandinka. In Dany Amiot (ed.), *La composition dans une perspective typologique* (Études Linguistiques). Arras: Artois Presses Université.
- Dayley, Jon P. 1989. *Tümpisa (Panamint) Shoshone grammar* (University of California Publications in Linguistics 115). Berkeley, CA: University of California Press.
- Dijk, Siebren. 1997. *Noun incorporation in Frisian*. Groningen: Rijksuniversiteit Groningen dissertation.
- Donohue, Mark. 1999. *Warembori* (Languages of the World/Materials 341). München: LINCOM Europa.
- Drossard, Werner. 2002. Ket as a polysynthetic language, with special reference to complex verbs. In Nicholas Evans & Hans-Jürgen Sasse (eds.), *Problems of polysynthesis* (Studia Typologica 4), 223–256. Berlin: Akademie Verlag.
- Dunn, Michael J. 1999. *A grammar of Chukchi*. Canberra: Australian National University dissertation.
- Eberhard, David M. 2009. *Mamaïndè grammar: A Northern Nambikwara language and its cultural context*. Utrecht: LOT Publications.
- Enrico, John. 2003. *Haida syntax*, vol. 2 (Studies in the Anthropology of North American Indians). Lincoln, NE: University of Nebraska Press.
- Evans, Nicholas. 2003. *Bininj Gun-wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune* (Pacific Linguistics 541). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- Foley, William A. 1991. *The Yimas language of New Guinea*. Stanford, CA: Stanford University Press.
- Foley, William A. 2017. The polysynthetic profile of Yimas, a language of New Guinea. In Michael Fortescue, Marianne Mithun & Nicholas Evans (eds.), *The Oxford handbook of polysynthesis* (Oxford Handbooks in Linguistics), 808–829. Oxford: Oxford University Press.
- Fortescue, Michael D. 1980. Affix ordering in West Greenlandic derivational processes. *International Journal of American Linguistics* 46(4). 259–278.
- Gerdts, Donna B. 2003. The morphosyntax of Halkomelem lexical suffixes. *International Journal of American Linguistics* 69(4). 345–356.
- Gerdts, Donna B. & Thomas E. Hukari. 2008. Halkomelem denominal verb constructions. *International Journal of American Linguistics* 74(4). 489–510.
- Givón, T. 2011. *Ute reference grammar* (Culture and Language Use: Studies in Anthropological Linguistics 3). Amsterdam: John Benjamins Publishing Company.
- Graczyk, Randolph. 2007. *A grammar of Crow: Apsáalooke Aliláau* (Studies in the Native Languages of the Americas). Lincoln, NE: University of Nebraska Press.
- Green, Ian. 1989. *Marrithiyel: A language of the Daly River region of Australia's Northern Territory*. Canberra: Australian National University dissertation.
- Harmelink, Bryan L. 1992. La incorporación nominal en el mapudungun. *Lenguas Modernas* 19. 129–138.
- Haude, Katharina. 2006. *A grammar of Movima*. Nijmegen: Radboud Universiteit Nijmegen dissertation.
- Henderson, James. 1995. *Phonology and grammar of Yele, Papua New Guinea* (Pacific Linguistics B112). Canberra: Department of Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- Johnson, Heidi A. 2000. *A grammar of San Miguel Chimalapa Zoque*. Austin, TX: University of Texas dissertation.
- Kooij, Johannes G. & Maarten Mous. 2002. Incorporation: A comparison between Iraqw and Dutch. *Linguistics* 40(3). 629–645.

- Lehmann, Christian & Elisabeth Verhoeven. 2005. Noun incorporation and participation: A typological study on participant association with particular reference to Yucatec Maya. In Christian Lehmann (ed.), *Typological studies in participation* (Studia Typologica 7), 105–188. Berlin: Akademie Verlag.
- Maksunova, Zoya. 2003. Compounding and incorporation in Ket (central dialect). *Sprachtypologie und Universalienforschung* 56(1/2). 123–132.
- Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language & Linguistic Theory* 19(1). 153–197.
- Melnar, Lynette. R. 1998. *Caddo verb morphology*. Chicago, IL: The University of Chicago dissertation.
- Miner, Kenneth L. 1986. Noun stripping and loose incorporation in Zuni. *International Journal of American Linguistics* 52(3). 242–254.
- Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4). 847–894.
- Mithun, Marianne. 2000. Incorporation. In Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung*, vol. 1 (Handbücher zur Sprach- und Kommunikationswissenschaft 17(1)), 916–928. Berlin: Walter de Gruyter.
- Mous, Maarten. 1993. *A grammar of Iraqw*. Hamburg: Helmut Buske.
- Osborne, Charles R. 1974. *The Tiwi language: Grammar, myths and dictionary of the Tiwi language spoken on Melville and Bathurst Islands, northern Australia* (Australian Aboriginal Studies 55). Canberra: Australian Institute of Aboriginal Studies.
- Palancar, Enrique L. 1999. Instrumental prefixes in Amerindian languages: An overview to their meanings, origin, and functions. *Sprachtypologie und Universalienforschung* 52(2). 151–166.
- Payne, Thomas E. 1995. Object incorporation in Panare. *International Journal of American Linguistics* 61(3). 295–311.
- Payne, Thomas E. & Doris L. Payne. 2013. *A typological grammar of Panare: A Cariban language of Venezuela* (Brill's Studies in the Indigenous Languages of the Americas 5). Leiden: Brill.
- Petersen de Piñeros, Gabriele. 2007. Nominal classification in Uitoto. *International Journal of American Linguistics* 73(4). 389–409.
- Queixalós, Francesc. 2008. Incorporation nominale en sikuani et en katukina-kanamari. *Amerindia* 31. 61–86.
- Quesada, J. Diego. 1999. Chibchan, with special reference to participant-highlighting. *Linguistic Typology* 3(2). 209–258.
- Rasch, Jeffrey W. 2002. *The basic morpho-syntax of Yaitepec Chatino*. Houston, TX: University of Houston dissertation.
- Rhodes, Richard A. 1976. *The morphosyntax of the Central Ojibwa verb*. Ann Arbor, MI: University of Michigan dissertation.
- Rice, Keren. 1991. Intransitives in Slave (Northern Athapaskan): Arguments for unaccusatives. *International Journal of American Linguistics* 57(1). 51–69.
- Rice, Keren. 2008. On incorporation in Athapaskan languages: Aspects of language change. In Thórhallur Eythórsson (ed.), *Grammatical change and linguistic theory: The Rosendal papers* (Linguistik Aktuell 113), 375–409. Amsterdam: John Benjamins Publishing Company.
- Rivail Ribeiro, Eduardo. 2001. Valence, voice and noun incorporation in Karajá. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 27. 229–243.
- Sapir, Edward. 1930. Southern Paiute, a Shoshonean language. *Proceedings of the American Academy of Arts and Sciences* 65(1). 1–296.
- Seiter, William J. 1980. *Studies in Niuean syntax* (Outstanding Dissertations in Linguistics). New York, NY: Garland Publishing.

- Shibatani, Masayoshi. 1990. *The languages of Japan* (Cambridge Language Surveys). Cambridge: Cambridge University Press.
- Smeets, Ineke. 2008. *A grammar of Mapuche* (Mouton Grammar Library 41). Berlin: Mouton de Gruyter.
- Spencer, Andrew. 1995. Incorporation in Chukchi. *Language* 71(3). 439–489.
- Stonham, John. 2004. *Linguistic theory and complex words: Nuuchahnulth word formation*. Basingstoke: Palgrave Macmillan.
- Stonham, John. 2008. Verb-argument composition in Nuuchahnulth: A special case of denominal verb formation. *International Journal of American Linguistics* 74(4). 511–528.
- Strom, Clay. 1992. *Retuarā syntax* (Studies in the Languages of Colombia 3). Arlington, TX: Summer Institute of Linguistics and University of Texas.
- Sullivan, Paul R. 1984. Noun incorporation in Yucatec Maya. *Anthropological Linguistics* 26(2). 138–160.
- Suttles, Wayne. 2004. *Musqueam reference grammar* (First Nations Languages). Vancouver: University of British Columbia Press.
- Tarpent, Marie-Lucie. 1987. *A grammar of the Nisgha language*. Victoria: University of Victoria dissertation.
- Vajda, Edward. 2017. Polysynthesis in Ket. In Michael Fortescue, Marianne Mithun & Nicholas Evans (eds.), *The Oxford handbook of polysynthesis* (Oxford Handbooks in Linguistics), 906–929. Oxford: Oxford University Press.
- Velázquez-Castillo, Maura. 1995. Noun incorporation in Guaraní: A functional analysis. *Linguistics* 33(4). 673–709.
- Vuillermet, Marine. 2012. *A grammar of Ese Eja, a Takanan language of the Bolivian Amazon*. Lyon: Lumière University Lyon 2 dissertation.
- Vuillermet, Marine. 2014. Two types of incorporation in Ese Eja (Takanan). In Swintha Danielsen, Katja Hannss & Fernando Zúñiga (eds.), *Word formation in South American languages* (Studies in Language Companion Series 163), 113–142. Amsterdam: John Benjamins Publishing Company.
- Weir, E. M. Helen. 1990. Incorporation in Nadëb. In Doris L. Payne (ed.), *Amazonian linguistics: Studies in Lowland South American languages* (Texas Linguistics Series), 321–363. Austin, TX: University of Texas Press.
- Wojdak, Rachel. 2005. *The linearization of affixes: Evidence from Nuuchahnulth*. Vancouver: University of British Columbia dissertation.
- Zúñiga, Fernando. 2000. *Mapudungun* (Languages of the World/Materials 376). München: LINCOM Europa.
- Zúñiga, Fernando. 2017. Mapudungun. In Michael Fortescue, Marianne Mithun & Nicholas Evans (eds.), *The Oxford handbook of polysynthesis* (Oxford Handbooks in Linguistics), 696–712. Oxford: Oxford University Press.

References

- Aikhenvald, Alexandra Y. 2007. Typological distinctions in word-formation. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 3, *Grammatical categories and the lexicon*, 2nd edn., 1–65. Cambridge: Cambridge University Press.
- Aldai, Gontzal & Søren Wichmann. 2018. Statistical observations on hierarchies of transitivity. *Folia Linguistica* 52(2). 249–281.
- Allen, Barbara J., Donna B. Gardiner & Donald G. Frantz. 1984. Noun incorporation in Southern Tiwa. *International Journal of American Linguistics* 50(3). 292–311.

- Anderson, Stephen R. 2000. Lexicalism, incorporated (or incorporation, lexicalized). *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 36(2). 13–34.
- Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago, IL: The University of Chicago Press.
- Bakker, Peter. 2006. Algonquian verb structure: Plains Cree. In Grażyna J. Rowicka & Eithne B. Carlin (eds.), *What's in a verb? Studies in the verbal morphology of the languages of the Americas* (LOT Occasional Series 5), 3–27. Utrecht: LOT.
- Barrie, Michael & Eric Mathieu. 2016. Noun incorporation and phrasal movement. *Natural Language & Linguistic Theory* 34(1). 1–51.
- Boas, Hans C. 2003. *A constructional approach to resultatives* (Stanford Monographs in Linguistics). Stanford, CA: Center for the Study of Language and Information Publications.
- Bochnak, M. Ryan & Alice Rhomieux. 2013. Limited noun incorporation in Washo. *International Journal of American Linguistics* 79(2). 253–281.
- Caballero, Gabriela, Michael J. Houser, Nicole Marcus, Teresa McFarland, Anne Pycha, Maziar Toosarvandani & Johanna Nichols. 2008. Nonsyntactic ordering effects in noun incorporation. *Linguistic Typology* 12(3). 383–421.
- Danielsen, Swintha. 2007. *Baure: An Arawak language of Bolivia*. Nijmegen: Radboud Universiteit Nijmegen dissertation.
- Diessel, Holger. 2019. *The grammar network. How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.
- Dijk, Siebren. 1997. *Noun incorporation in Frisian*. Groningen: Rijksuniversiteit Groningen dissertation.
- Dixon, R. M. W. 2004. Adjective classes in typological perspective. In R. M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Adjective classes: A cross-linguistic typology* (Explorations in Linguistic Typology 1), 1–49. Oxford: Oxford University Press.
- Fortescue, Michael D. 1980. Affix ordering in West Greenlandic derivational processes. *International Journal of American Linguistics* 46(4). 259–278.
- Fortescue, Michael. 1984. *West Greenlandic* (Croom Helm Descriptive Grammars). London: Croom Helm.
- Genee, Inge. 2018. Measuring polysynthesis: A Functional Discourse Grammar approach. In Evelien Keizer & Hella Olbertz (eds.), *Recent developments in Functional Discourse Grammar* (Studies in Language Companion Series 205), 233–272. Amsterdam: John Benjamins Publishing Company.
- Gerds, Donna B. 1998. Incorporation. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology* (Blackwell Handbooks in Linguistics), 84–100. Oxford: Blackwell Publishers.
- Gerds, Donna B. 2003. The morphosyntax of Halkomelem lexical suffixes. *International Journal of American Linguistics* 69(4). 345–356.
- Goldberg, Adele. 1995. *Constructions: A Construction Grammar approach to argument structure* (Cognitive Theory of Language and Culture). Chicago, IL: The University of Chicago Press.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath. 2017. *Glottolog 3.0*. Jena: Max Planck Institute for the Science of Human History. <https://glottolog.org> (accessed 17 May 2017).
- Hartmann, Iren, Martin Haspelmath & Bradley Taylor (eds.). 2013. *Valency patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://valpal.info>.
- Haspelmath, Martin. 2015. Transitivity prominence. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, vol. 1, *Introducing the framework, and case studies from Africa and Eurasia* (Comparative Handbooks of Linguistics 1(1)), 131–147. Berlin: De Gruyter Mouton.

- Haspelmath, Martin. 2018. The last word on polysynthesis: A review article. *Linguistic Typology* 22(2). 307–326.
- Haspelmath, Martin & Iren Hartmann. 2015. Comparing verbal valency across languages. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, vol. 1, *Introducing the framework, and case studies from Africa and Eurasia* (Comparative Handbooks of Linguistics 1(1)), 41–71. Berlin: De Gruyter Mouton.
- Haude, Katharina. 2006. *A grammar of Movima*. Nijmegen: Radboud Universiteit Nijmegen dissertation.
- Haugen, Jason D. 2015. Incorporation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe*, vol. 1, (*Handbücher zur Sprach- und Kommunikationswissenschaft* 40(1)), 413–434. Berlin: De Gruyter Mouton.
- Heaton, Raina. 2017. *A typology of antipassives, with special reference to Mayan*. Honolulu, HI: University of Hawai'i dissertation.
- Hirose, Tomio. 2003. *Origins of predicates: Evidence from Plains Cree* (Outstanding Dissertations in Linguistics). New York, NY: Routledge.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.
- Huson, Daniel & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2). 254–267.
- Johns, Alana. 2017. Noun incorporation. In Martin Everaert & Henk C. van Riemsdijk (eds.), *The Wiley Blackwell companion to syntax*, 2nd edn. Hoboken, NJ: John Wiley & Sons Inc.
- Kemmer, Suzanne. 1993. *The middle voice* (Typological Studies in Language 23). Amsterdam: John Benjamins Publishing Company.
- Kurebito, Tokusu. 2001. On lexical affixes in Chukchi. In Osahito Miyaoka & Fubito Endo (eds.), *Languages of the North Pacific Rim*, vol. 6, (Endangered languages of the Pacific Rim), 65–84. Kyoto: Nakanishi.
- Levin, Beth. 2015. Verb classes within and across languages. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, vol. 2: *Case studies from Austronesia, the Pacific, the Americas, and theoretical outlook* (Comparative Handbooks of Linguistics 1(2)), 1627–1670. Berlin: De Gruyter Mouton.
- Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity: At the syntax-lexical semantics interface* (Linguistic Inquiry Monographs 26). Cambridge, MA: The Massachusetts Institute of Technology Press.
- Lier, Eva van. 2017. The typology of property words in Oceanic languages. *Linguistics* 55(6). 1237–1280.
- Malchukov, Andrej L. 2005. Case pattern splits, verb types and construction competition. In Mengistu Amberber & Helen de Hoop (eds.), *Competition and variation in natural languages: The case for case* (Perspectives on Cognitive Science), 73–117. Amsterdam: Elsevier.
- Malchukov, Andrej. 2015. Valency classes and alternations: Parameters of variation. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, vol. 1, *Introducing the framework, and case studies from Africa and Eurasia* (Comparative Handbooks of Linguistics 1(1)), 73–130. Berlin: De Gruyter Mouton.
- Massam, Diane. 2009. Noun incorporation: Essentials and extensions. *Language and Linguistics Compass* 3(4). 1076–1096.

- Massam, Diane. 2017. Incorporation and pseudo-incorporation in syntax. In *Oxford research encyclopedia of linguistics* (Oxford Research Encyclopedias). Oxford: Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.190> (accessed 28 March 2019).
- Mathieu, Eric. 2013. Denominal verbs in Ojibwe. *International Journal of American Linguistics* 79(1). 97–132.
- Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4). 847–894.
- Mithun, Marianne. 2000. Incorporation. In Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung*, vol. 1 (Handbücher zur Sprach- und Kommunikationswissenschaft 17(1)), 916–928. Berlin: Walter de Gruyter.
- Mithun, Marianne. 2010. Constraints on compounding and incorporation. In Sergio Scalise & Irene Vogel (eds.), *Cross-disciplinary issues in compounding* (Current Issues in Linguistic Theory 311), 37–56. Amsterdam: John Benjamins Publishing Company.
- Mosel, Ulrike. in prep. Cut-verbs of the Oceanic language Teop, a critical study of collecting and analyzing lexical data in a language documentation project. In Helen Bromhead & Zhengdao Ye (eds.), *Meaning, life and culture: In conversation with Anna Wierzbicka*.
- Mous, Maarten & Martha Qorro. 2010. The syntactic function of a scope marking suffix in Iraqw. *Journal of African Languages and Linguistics* 31(1). 47–78.
- Næss, Åshild. 2007. *Prototypical transitivity* (Typological Studies in Language 72). Amsterdam: John Benjamins Publishing Company.
- Olthof, Marieke. 2020. Formal variation in incorporation: A typological study and a unified approach. *Linguistics* 58(1). 131–205.
- Olthof, Marieke & Eva van Lier. 2018. Verb-based restrictions on incorporation: A typological survey. (Paper presented at Syntax of the World's Languages 8, Paris, 3–5 September 2018.).
- Palancar, Enrique L. 1999. Instrumental prefixes in Amerindian languages: An overview to their meanings, origin, and functions. *Sprachtypologie und Universalienforschung* 52(2). 151–166.
- Polinsky, Maria. 2013. Antipassive constructions. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/109>.
- Rosen, Sara T. 1989. Two types of noun incorporation: A lexical analysis. *Language* 65(2). 294–317.
- Say, Sergey. in prep. *The antipassive derivation and the lexical meaning of the verb*.
- Seiter, William J. 1980. *Studies in Niuean syntax* (Outstanding Dissertations in Linguistics). New York, NY: Garland Publishing.
- Tsunoda, Tasaku. 1981. Split case-marking patterns in verb-types and tense/aspect/mood. *Linguistics* 19(5/6). 389–438.
- Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of Linguistics* 21(2). 385–396.
- Vajda, Edward. 2017. Polysynthesis in Ket. In Michael Fortescue, Marianne Mithun & Nicholas Evans (eds.), *The Oxford handbook of polysynthesis* (Oxford Handbooks in Linguistics), 906–929. Oxford: Oxford University Press.
- Velupillai, Viveka. 2012. Legend for feature map 6.1 (p. 120). Online materials for *An introduction to linguistic typology*. Amsterdam: John Benjamins Publishing Company. <https://benjamins.com/series/z/176/6.1.pdf> (accessed 16 September 2019).
- Vigus, Meagan. 2018. Antipassive constructions: Correlations of form and function across languages. *Linguistic Typology* 22(3). 339–384. <https://doi.org/10.1515/lingty-2018-0013>.
- Vuillermet, Marine. 2012. *A grammar of Ese Eja, a Takanan language of the Bolivian Amazon*. Lyon: Lumière University Lyon 2 dissertation.

- Wichmann, Søren. 2015. Statistical observations on implicational (verb) hierarchies. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, vol. 1, *Introducing the framework, and case studies from Africa and Eurasia* (Comparative Handbooks of Linguistics 1(1)), 155–181. Berlin: De Gruyter Mouton.
- Wichmann, Søren. 2016. Quantitative tests of implicational verb hierarchies. In Taro Kageyama & Wesley M. Jacobsen (eds.), *Transitivity and valency alternations: Studies on Japanese and beyond* (Trends in Linguistics: Studies and Monographs 297), 423–444. Berlin: De Gruyter Mouton.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/lingty-2020-2054>).