Formal variation in incorporation: A typological study and a unified approach

Olthof, M.

DOI
10.1515/ling-2019-0036

Publication date
2020

Document Version
Final published version

Published in
Linguistics

License
CC BY

Citation for published version (APA):
Abstract: This study investigates the formal variation in elements involved in incorporation structures. Although it has traditionally been assumed that only stems can be incorporated, several languages also show incorporation of formally more complex elements. Most theoretical approaches to incorporation are, however, limited to incorporated stems and the few that do include more complex forms consider the incorporation of simple and complex elements separate processes. The present study, by contrast, recognizes the many shared characteristics between incorporation structures with simple and complex elements and adopts the approach to incorporation advocated in Functional Discourse Grammar, which proposes a unified account of incorporated simple stems, derived stems, inflected words, phrases and clauses. In addition, the study hypothesizes that these forms constitute an implicational hierarchy, i.e., that more complex incorporated elements only occur in languages that also allow the incorporation of all simpler forms. Results from a typological study of incorporated elements in a variety sample of 30 incorporating languages show that all forms except clauses are indeed found incorporated and that the hypothesized implicational pattern holds. The study thus demonstrates that the incorporation of simple and complex elements is interrelated, supporting a unified treatment of incorporated elements of different degrees of complexity.

Keywords: Functional Discourse Grammar, incorporation, polysynthetic languages, typology

1 Introduction

This paper addresses the often-unnoticed formal variation in elements involved in incorporation structures. Incorporation can be described as the inclusion of one lexical element in another lexical element such that they together constitute a single word (Mithun 1994: 5024; Gerdts 1998: 84; Haugen 2015: 414). While this
process is relatively rare in most well-known European languages, it is applied productively in various other languages, many of which are generally considered polysynthetic (Haspelmath and Sims 2010: 138; Murasugi 2014: 283–284) and most of which are spoken in North and South America, Northern Australia, Austronesia and Siberia (Mithun 1994: 5024; Velupillai 2012a: 120). The most widely investigated type of incorporation is noun incorporation (Gerdts 1998: 84; Iturrioz Leza 2001: 714), in which a nominal argument, typically an object, or modifier of a verb is incorporated into this verb (Mithun 2000: 917; Haugen 2015: 414–415). An example of such an incorporation construction in Chukchi is shown in (1b).

(1) Incorporation of a nominal stem into a verb in Chukchi

a. ʔə tt-e piri-nin-Ø melota-lyən
   dog-ERG catch-3SG > 3SG-PST hare-ABS.SG
   ‘The dog caught the hare.’

b. ʔə tt-ə-n milute-piri-ɣʔi-Ø
   dog-E-ABS.SG hare-catch-3SG.S-PST
   ‘The dog caught a hare.’

(Kurebito 2012: 181)

Example (1a) contains a clause consisting of a subject noun in the ergative case ʔə tt-e ‘dog’, a verb with the stem piri ‘catch’ and a direct object noun in the

---

1 Incorporated elements in the examples are displayed in bold. Glosses in the examples follow the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/resources/glossing-rules.php). Abbreviations used: 1 = first person, 2 = second person, 3 = third person, I = gender class I, IV = noun class IV, A = agent-like argument of canonical transitive verb, AB = absolutive, ABS = absolute, ACSBJ = subject of active verb, ADVZ = adverbializer, ALL = allative, APPROX = approximative, ASP = aspect, AU = augmented, AUG = augmentative, AUX = auxiliary, BE = bound element, COMPL = complete, DDEIX = distal deixis, DECL = declarative, DEF = definite, DET = determiner, DINC = dependent incomplete, DISTR = distributive, DR = bivalent direct, DS = different-subject marker, DU = dual, E = epenthesis, EMPH = emphatic, ERG = ergative, FUT = future, GOAL = goal postposition, HAB = habitual, IMP = imperative, IND = indicative, INDF = indefinite, INS = instrumental, INTR = intransitive, IVF = incorporating verb form, LNDEIX = location/negation deixis, M = masculine, MIR = mirative, N = neuter, NPST = nonpast, OBJ = object, P = patient-like argument of canonical transitive verb, PERAMB = perambulative, PFV = perfective, PL = plural, POSS = possessive, PRO = emphatic or contrastive proform, PROG = progressive, PROS = prosecutive case, PRS = present, PST = past, PTCP = participle, RC = relative case, REFII = reflexive, REL = relative, RMP = remote past, S = single argument of canonical intransitive verb, SBJ = subject, SBIV = subjunctive, SG = singular, SIM = simultaneous, SS = same-subject marker, SUB = subordinator, TC = thematic consonant.

2 The vowel differences between melota in (1a) and milute in (1b) are due to a vowel harmony rule (Kurebito 2012: 188, n. 3).
absolutive case *melota-lyan* ‘hare’. In (1b), the stem of the direct object noun is incorporated into the verb and the nominal and verbal stem together form a single, complex verbal stem. Here the nominal stem *milute* can be called an incorporated element, whereas the verb with the stem *piri* is the host of the incorporation process. In this example, the incorporation process detransitivizes the verbal stem, such that the subject noun *ʔət-t-ə-n* appears in the absolutive case in (1b). In addition, *milute* in (1b) is interpreted as an indefinite noun, in contrast to the freestanding noun *melota-lyan* in (1a). Note also that while Example (1b) shows a case of noun incorporation, crosslinguistically not only nouns but also other lexical parts of speech, i.e., verbs, adjectives, adverbs and adpositions, can be incorporated into verbs (Baker 1988: 147, 229; Gerdts 1998: 84; Massam 2009: 1077; Štekauer et al. 2012: 63–64).

Incorporation constructions have been the topic of a large body of literature. While descriptive studies have examined the various characteristics of incorporation structures in diverse languages, theoretical work has concentrated primarily on the question whether incorporation, or, more specifically, noun incorporation, should be considered a morphological or a syntactic process (Mithun 2000: 923–925; Massam 2009: 1083–1086; Haugen 2015: 414–421). One important issue that has been addressed is the question whether incorporated nouns have a non-referential function, resembling compounded nouns, or a referential function, like independent nouns. In addition, theoretical studies have focused on the phenomena of modifier stranding and external possession, in which apparent modifiers and possessors of incorporated nouns, respectively, appear next to the relevant incorporation constructions. An example of modifier stranding from Southern Tiwa is shown in Example (2), in which *wisi* ‘two’ may be considered a modifier of the incorporated *musa* ‘cat’, whereas Example (3) from Chukchi exemplifies the presence of a supposed external possessor, i.e., *nenenə* ‘child’ can be analyzed as the possessor of *many* ‘hand’.

(2) Incorporation of a nominal stem into a verb in combination with a stranded numeral modifier in Southern Tiwa

> Wisi *ibi*-**musa-tuwi-ban.**
> two 1.PL >1.PL -cat-buy-PST
> ‘They bought two cats.’

(Allen et al. 1984: 297)

(3) Incorporation of a nominal stem into a verb in combination with an external possessor in Chukchi

> *t-ə-many-ə-ytak-w?an-Ø nenena-Ø*
> 1SG.A-E-hand-E-wipe-3SG.P-PST child-ABS.SG
‘I wiped a child’s hands.’
(Kurebito 2012: 182)

If incorporated nouns can function referentially and are assumed to show a
syntactic relationship to stranded modifiers and external possessors, they may
be considered similar to independent nouns occurring in noun phrases. In
contrast, if incorporated nouns are non-referential and stranded modifiers and
external possessors are not interpreted as being directly related to incorporated
nouns, incorporated nouns are simply like prototypical compound members.
Thus, incorporation has been an important topic in research on the similarities
and differences between word formation and the construction of phrases and
clauses (Massam 2009: 1081).

Due to large crosslinguistic variation in incorporation structures, proposed
definitions of incorporation diverge greatly. One characteristic that is never-
theless argued to be shared by many incorporation structures and often even
considered one of the distinctive properties of incorporation is that the incorpo-
rated element has the form of a stem (Mithun 2000: 917; Mattissen 2003: 178;
Štekauer et al. 2012: 43), or, more specifically, a simple stem, consisting of a
1998: 85; Haugen 2015: 414). In the morphological literature, such a mono-
morphemic stem is often called a root (e.g., Payne 1997: 24; Haspelmath and
Sims 2010: 21). However, as the term root will be used exclusively for bound
lexical morphemes in this paper (see Section 2.3), here the term ‘simple stem’
will be used to refer to unbound, mono-morphemic stems.

An incorporated element is thus typically a simple stem, without any deriva-
tional or inflectional morphology, such as the incorporated noun in Example (1b),
which does not show the marking for case and number present in (1a). However,
recently it has been shown that not all incorporated elements are simple stems.
Several studies have presented incorporated nouns that are derivationally com-
plex or compounded (Muro 2009: 130–133; Mithun 2010: 45; Barrie and Mathieu
2016), such as the Chimalapa Zoque deverbal noun can-kuyʔ ‘seat’, which is
derived from the verb can ‘to sit’ by means of the instrumental suffix -kuyʔ and
is incorporated into the verb ciʔ ‘give’ in (4) (Johnson 2000: 185, 276).

(4) Incorporation of a nominal derived stem, consisting of a stem and a
grammatical affix, into a verb in Chimalapa Zoque

\[
\begin{array}{l}
?an = \text{can-kuyʔ}-ciʔ-\text{šuk-\text{-wə}}
\end{array}
\]

1.\text{ERG} = \text{sit-INS-give-2/3PL-COMPL} \quad \text{2/3PL-COMPL}

‘I gave them seats and they sat down.’
(Johnson 2000: 276)
In addition, some languages allow the incorporation of inflected words (Iturrioz Leza 2001: 721; Hengeveld and Mackenzie 2008: 414; Muro 2009: 144; Barrie and Mathieu 2016). An example of a Kalaallisut nominal stem that is incorporated together with its inflectional morphology marking possession is displayed in (5).

(5) Incorporation of a nominal inflected word into a verb in Kalaallisut

\[ \text{illu-mi-niip-puq} \]

house-REFL.Poss-be.in-3SG.IND

‘He is in his (own) house.’

(Fortescue 1984: 300–301)

Moreover, some languages even show incorporation structures in which a full phrase is incorporated (Aikhenvald 2007: 13–14; Hengeveld and Mackenzie 2008: 415; Muro 2009: 140; Barrie and Mathieu 2016). For instance, the Mapudungun noun phrase in (6a), consisting of the noun \textit{mansun} ‘ox’ and the modifier \textit{ngilla-n} ‘bought’, which is again modified by the adverb \textit{we} ‘newly’, can as a whole be incorporated into the verb with the stem \textit{adkintu} ‘watch’ (Zúñiga 2017: 705), as demonstrated in (6b).

(6) Incorporation of a noun phrase into a verb in Mapudungun\(^3\)

\textbf{a. Adkintu-yaw-i we ngilla-n mansun.}

watch-PERAMB.IND newly buy-PTCP ox

\textbf{b. Adkintu-we-ngilla-n-mansun-kiyaw-i.}

watch-newly-buy-PTCP-ox-PERAMB.IND

‘He is (going around) looking after a recently bought ox.’

(Harmelink 1992: 133; Zúñiga 2006: 181; translation from Spanish and glosses based on Zúñiga 2017)

Although incorporated elements thus appear to vary in their forms, the incorporation of elements consisting of more material than just a simple stem has received relatively little attention in the literature on incorporation (but see Iturrioz Leza 2001; Aikhenvald 2007; Muro 2009; Barrie and Mathieu 2016). A comprehensive investigation of the crosslinguistic formal variation in incorporated elements is lacking. Moreover, so far, no single, comprehensive account of all forms of incorporated elements that appear to exist has been proposed (see Section 2.1). The general focus in theoretical research on incorporation has been

\(^3\) The change in the form of the perambulative suffix is due to allomorphy: when the suffix follows a vowel it takes the form \textit{-yaw}, as in (6a), while in all other cases it is \textit{-kiyaw}, as in (6b) (Zúñiga 2000: 50).
on the incorporation of stems (Mithun 1984, Mithun 1986; Baker 1988, Baker 1996, Baker 2003, Baker 2009). A few studies such as Muro (2009) and Barrie and Mathieu (2016) have attempted to account for the incorporation of larger elements such as inflected words and phrases, but their approaches exclude non-referential incorporated nouns, thus restricting their domain of applicability. In addition, in these studies the incorporation of simple and more complex forms are considered distinct processes. Consequently, no unified account is available that is able to explore the full range of formal variation in all types of incorporated elements. However, I believe that such an account is highly desirable since incorporation structures, regardless of the form of their incorporated elements, are highly similar in appearance and share a number of characteristics, such as their ability to combine with stranded modifiers and external possessors (Mithun 1984: 856–859; Baker 1988: 92–105; Rosen 1989: 298–301; Barrie and Mathieu 2016). In addition, as the present research will show, such structures are interrelated in terms of their distribution.

The present paper therefore investigates the range of variation in the forms of incorporated elements and their crosslinguistic distribution, proposing a unified treatment of these forms. To this end, I present a typological study of the forms of incorporated elements, taking a Functional Discourse Grammar (FDG) approach to incorporation. FDG is a functional linguistic theory that attempts to explain formal characteristics of languages on the basis of their communicative function (Hengeveld and Mackenzie 2008). In correspondence with Baker (1988, 1996, 2009), Muro (2009) and Barrie and Mathieu (2016), FDG considers incorporation a syntactic or grammatical process rather than a lexical one. In contrast to the abovementioned theories, however, FDG includes both referential and non-referential incorporated nouns and both formally simple and formally more complex incorporated elements, which makes it suitable for the broad explorative study into the forms of incorporated elements of the current research. Moreover, FDG proposes that in these constructions a single phenomenon is at work, allowing a unified account of the incorporation of simple elements and more complex elements.

In addition to identifying the varied forms of incorporated elements, the study investigates the crosslinguistic distribution of these forms, hypothesizing that a pattern can be found. More specifically, it is predicted that the forms of incorporated elements constitute an implicational hierarchy, ranging from the most simple and frequent forms of incorporated elements to the rarer and more complex incorporated elements. Such a distributional pattern would provide support for a unified account of the incorporation of formally simple and formally complex elements, as it would show that the occurrence of incorporated elements of different forms is interdependent in a specific way. Using data
from a sample of 30 incorporating languages with genealogically, geographi-
cally, and typologically diverse backgrounds, the present research examines the
range of variation in the forms of incorporated elements, based on FDG’s unified
approach towards incorporated elements of different forms, and the distribution
of the different forms of incorporated elements.

The outline of the paper is as follows. Section 2 discusses the theoretical
background for the research, addressing previous theoretical work on incorpo-
ration and outlining the FDG approach to incorporation taken in the study.
Section 3 offers the hypotheses and predictions regarding the forms and cross-
linguistic distribution of incorporated elements investigated. After a discussion
of the research method in Section 4, Section 5 presents the results. Finally,
Section 6 provides the conclusions about the formal variation of incorporated
elements, the crosslinguistic distribution of the different elements, and the
suitability of a unified treatment of the different forms.

2 Theoretical background

2.1 Previous accounts of incorporation

As indicated above, the aim of this paper is to examine the full range of
variation in the forms of incorporated elements and to propose a unified account
of the many forms that incorporated elements may take. Previous accounts do
not seem suitable to achieve this, as they are all restricted to particular forms
and types of incorporation. I will discuss them briefly in this section.

There are two primary theoretical perspectives on incorporation, differing
from each other in where in the linguistic system incorporation constructions
are assumed to be formed (Mithun 1994: 5025; Štekauer et al. 2012: 45; Haugen
2015). Some researchers claim that incorporation is a morphological process that
operates in the lexicon, while others argue that the constructions are created in
the syntax and that their behavior is consistent with general syntactic principles

In the morphological or lexical approach, incorporation is considered a
word-formation process similar or identical to compounding (Caballero et al.
this morphological account of incorporation are an early paper by Sapir (1911)
and the research by Mithun (1984, 1986), Di Sciullo and Williams (1987), Rosen
(1989) and Anderson (2000), who all discuss characteristics that incorporation
shares with other types of word formation. For instance, Mithun (1984: 889)
emphasizes that incorporation, unlike syntactic mechanisms but in correspondence with other morphological processes, has limited productivity, in that most languages do not allow all nouns and verbs to be involved in noun incorporation. She also proposes that incorporated nouns are non-referential, just like compounded nouns (Mithun 1984: 849) and argues that the phonological and semantic idiosyncrasies of many incorporation structures are evidence for their lexical status (Mithun 1984: 889–890). Rosen (1989) adds to these word-formation-like properties the different valency effects of incorporation. In one type of noun incorporation, which she calls “compound NI [noun incorporation]”, the verb detransitivizes, whereas in another type, “classifier NI”, the valency of the verb remains unaffected (Rosen 1989: 295–296). According to Rosen (1989: 313–314), these different valency effects can best be accounted for by claiming that they are specified lexically. Finally, Anderson (2000: 16) highlights that incorporation is typically restricted to nouns with particular thematic roles: incorporated nouns are themes or sometimes locatives or instrumentals. Many other word-formation processes are sensitive to such restrictions as well.

The syntactic approach, by contrast, argues that incorporation is a process that adheres to regular syntactic principles (Štekauer et al. 2012: 45). The incorporated noun is assumed to be an independent syntactic constituent with normal referential status. Within the syntactic approach, two strands can again be identified. In the first, incorporation is analyzed in terms of head movement, as first proposed by Baker (1988, 1996, 2009). In Baker’s influential account, incorporation involves the movement of a head noun from its regular syntactic position to the verb. According to Baker (1988: 95), this movement analysis explains the modifier stranding possibilities of many incorporated nouns: when a noun undergoes head movement it may leave its modifier behind. Further support for the head-movement analysis can be found in the observation that many languages only allow direct objects and subjects of unaccusative verbs to be incorporated, which would indicate that incorporation is restricted to movement from particular positions in the syntactic structure, in the same way as other syntactic movement operations (Baker 1988: 81–92).

In the second type of syntactic approach, incorporation is considered phrasal movement. Barrie and Mathieu (2016) propose that incorporation involves the merging of an XP with a verb. They adopt the Distributed Morphology assumption that roots are acategorical abstract items that only become categorized by

---

4 Baker (1988: Ch. 4 and 5) also addresses verb and adverb incorporation, in which verbs and adpositions undergo head movement.
5 Barrie and Mathieu’s (2016: 5–6) main motivation for proposing a phrasal-movement analysis is that Baker’s head movement analysis does not match the present-day Minimalist approach.
merging them with functional heads (Barrie and Mathieu 2016: 4–5), which also give them referential status. Following this assumption, they argue that incorporation targets the following phrases: “nP (categorized/nominalized stems), dP (modified N-stem), DP (possessor DPs, demonstratives), KP (case-marked nominals), and CP (relative clauses)” (Barrie and Mathieu 2016: 9). In this way, Barrie and Mathieu’s analysis (2016) can account for various formally complex incorporated elements such as derived stems, inflected words, and phrases. Other studies presenting a phrasal-movement analysis of incorporated elements are Allen (1988), in which a KP-movement analysis for the incorporation of case-marked nouns in Kalaallisut is proposed, and Muro (2009), which distinguishes between a phrasal-movement analysis for complex incorporated elements and Baker’s head movement account for simple forms.

Each of these approaches is valuable in being able to explain diverse characteristics of incorporation. However, whereas the present study investigates the whole range of variation in the forms of incorporated elements, the three approaches just discussed are restricted to certain forms of incorporated elements and to elements with certain referential characteristics. The morphological compounding approach and Baker’s head-movement analysis are limited to the incorporation of stems. Barrie and Mathieu’s phrasal account, by contrast, states that noun incorporation minimally involves an nP. Correspondingly, they exclude particular simple nominal stems, i.e., those that are considered uncategorized roots in DM, such as bound nominal forms and nouns that are morphologically reduced when they are part of a verbal word (Barrie and Mathieu 2016: 23). Similarly, Muro (2009) considers the incorporation of simple forms as distinct from the incorporation of more complex forms.

In addition, with respect to noun incorporation, the morphological approach is limited to the incorporation of non-referential or non-specific elements, while the two syntactic approaches are restricted to the incorporation of fully referential elements. However, in different languages both referential and non-referential incorporated nouns have been attested and constructions with referential and non-referential incorporated nouns share many characteristics such as the possibility to strand a modifier of the incorporated noun (Rosen 1989: 298–301; Baker 1988: 92–96), the possibility to combine the construction with an external possessor of the incorporated noun (Mithun 1984: 856; Baker 1988: 96–105) and the possibility to double the incorporated noun in an external noun phrase (Mithun 1984: 863–864; Rosen 1989: 297, 302–304; Baker 1988: 144–145). As my aim is to provide a broad approach to incorporation that enables a unified

---

6 Wiltschko (2009) also proposes a Distributed Morphology-based account of incorporation, but limits incorporation to bare roots and nominalized roots.
treatment of all forms and types of incorporation, a more flexible approach than the ones discussed in this section is needed. FDG appears to provide this flexibility. Its approach to incorporation is discussed in the next section.

2.2 Incorporation in Functional Discourse Grammar

FDG is a functional linguistic theory that investigates linguistic forms in terms of their communicative functions (Hengeveld and Mackenzie 2008: 26–27). The Grammatical Component of the framework contains four independent but interacting linguistic levels (Hengeveld and Mackenzie 2008: 5, 23). The first two levels, the Interpersonal Level (IL) and the Representational Level (RL), take care of the Formulation of an utterance, which means that conceptual representations that a speaker wants to express are translated into pragmatic and semantic representations (Hengeveld and Mackenzie 2008: 12). At IL the pragmatic units of the utterance are selected, while the semantic units of the utterance are obtained at RL (Hengeveld and Mackenzie 2008: 12). Subsequently, Encoding of the utterance takes place at the Morphosyntactic Level (ML) and the Phonological Level (PL), where the pragmatic and semantic representations are converted into morphosyntactic and phonological representations respectively (Hengeveld and Mackenzie 2008: 12). The levels consist of several hierarchically ordered layers which contain units that are relevant at these levels, i.e., pragmatic units at IL, semantic units at RL, morphosyntactic units at ML, and phonological units at PL (Hengeveld and Mackenzie 2008: 14).

Here I use the FDG approach to incorporation, in which the process of incorporation is argued to take place in the Grammatical Component, i.e., in the grammar rather than in the lexicon, and incorporation is defined on the basis of semantic and morphosyntactic characteristics, i.e., at RL and ML. At RL, the incorporated element and its host correspond to two semantic units that are in a dependency relation. FDG distinguishes two types of dependency relations (Hengeveld and Mackenzie 2008: 305–306). On the one hand, two units may be in a head-modifier relation, where the modifier is an optional dependent of the head. An example of an incorporation construction in which the incorporated element and its host form a head-modifier relation is shown in (7), in which the adverb toyko ‘thoroughly’ is a modifier of the verbal predicate kikkik ‘beat’. 

7 Technical terms as applied in FDG are capitalized (see Hengeveld and Mackenzie 2008: 43).
Incorporation of an adverb (modifier) into a verb (head) in Hokkaido Ainu
\textit{A-toyko-kikkik}.

\begin{tabular}{l}
\text{INDF.A-thoroughly-beat} \\
'I beat (him) up thoroughly.'
\end{tabular} \\
\text{(Shibatani 1990: 71–72)}

A second example is the incorporation of an adjunct into (the predicate of) the predication frame that it modifies, as in (8), where the instrument \textit{sal} ‘tobacco’ is incorporated into the verbal predicate \textit{kit} ‘rub’ of the predication frame ‘the hunter rubbed the dog’.

\begin{tabular}{l}
\text{Incorporation of a nominal adjunct (modifier) into a verb (head) in Ket} \\
\text{\textit{assano ke’d tib d = sal-a-t-a-kit}} \\
\text{hunting person dog 3.SBJ = tobacco-3SG.M.SG-TC-PRS-rub} \\
'The hunter “tobaccoed” the dog (to rid it of fleas).'
\end{tabular} \\
\text{(Vajda 2017: 916)}

On the other hand, dependency relations may have the form of nucleus-dependent relations, which hold between a predicate and an argument, which are both obligatory parts of the head of a semantic unit (Hengeveld and Mackenzie 2008: 305). An example of an incorporation construction in which the incorporated element and its host form a nucleus-dependent relation is the structure in (9). Here, the nominal argument \textit{shut} ‘shirt’ is incorporated into its verbal predicate \textit{pe} ‘make’, and \textit{shut} and \textit{pe} together form the semantic head of the predication.

\begin{tabular}{l}
\text{Incorporation of a nominal argument (dependent) into a predicate (nucleus) in Southern Tiwa} \\
\text{\textit{Ti-shut-pe-ban}.} \\
\text{1SG > 1.SG-shirt-make-PST} \\
'I made the/a shirt.'
\end{tabular} \\
\text{(Allen et al. 1984: 293)}

At RL, incorporation constructions may thus involve either a head-modifier or a predicate-argument relation, such that semantically diverse incorporation structures are allowed.

At ML, incorporation constructions are characterized as showing “lexically realized equipollent units” at the word layer, i.e., within a single morphosyntactic word (Hengeveld and Mackenzie 2008: 404). Many different definitions of the morphosyntactic or grammatical word have been proposed, but here I will
assume that such a word is a fixed combination of elements that consistently occur together (Dixon and Aikhenvald 2002: 19; Haspelmath and Sims 2010: 193; Aronoff and Fudeman 2011: 38), that appear in a fixed order, i.e., the meaning of a word typically changes if the order of the elements is altered (Dixon and Aikhenvald 2002: 19; Aronoff and Fudeman 2011: 38), and that together form a unit that takes inflectional marking (Haspelmath and Sims 2010: 193, see also Dixon and Aikhenvald 2002: 22). In FDG, morphosyntactic words are formed on the basis of word templates, which determine which units can occur inside a word. Word templates may in principle consist of one or more morphemes (Xm), other words (Xw), phrases (Xp) and clauses (Cl), such that the maximal template for words is the one shown in (10) (Hengeveld and Mackenzie 2008: 400).

(10) Maximally elaborated morphosyntactic word template

\[(Xw_1: [(Xm) (Xw) (Xp) (Cl)] (Xw_1))] \]

(Hengeveld and Mackenzie 2008: 400)

Each language makes use of a language-specific inventory of word templates that are based on this template in (10) and thus maximally consist of a number of morphemes, words, phrases and clauses. All words in a language, including, in the case of an incorporating language, incorporation structures, correspond to one of the language’s word templates. The square brackets around (Xm), (Xw), (Xp) and (Cl) in (10) indicate that these units are in an equipollent relation within the word (Hengeveld and Mackenzie 2008: 14–15). Thus, at ML, an incorporation structure is a word that contains two or more lexical morphemes, words, phrases and/or clauses.8

Taking the characterizations of incorporation at RL and ML together, the definition of an incorporation construction can be formulated as follows: an incorporation construction is a morphosyntactic word containing two (or more) equipollent lexical morphosyntactic units that are semantically in a head-modifier or a nucleus-dependent relation. Note that this definition describes the domain of incorporation crosslinguistically and that each particular language may have restrictions on incorporation structures within this domain, both in terms of semantics and in terms of morphosyntax. For instance, in some languages incorporation may be limited to the incorporation of arguments into predicates, i.e., incorporation of modifiers into heads is then impossible. Similarly, depending on their inventory of word templates, some incorporating

8 Definitions of lexical morphemes, words, phrases and/or clauses used in FDG follow in Section 3.1.
languages may, for example, only allow incorporated elements in the form of morphemes or words, not showing incorporated phrases and clauses. FDG thus allows for crosslinguistic variation in the semantics, i.e., the type of dependency relation between the incorporated element and its host, and in the morphosyntax of incorporation structures. Finally, many languages do not show incorporation at all. These languages are assumed not to have word templates that match incorporation structures.

2.3 Implications of the FDG definition of incorporation

The definition of incorporation given above allows for a relatively broad perspective on incorporation, which is very useful in the present study as it aims to explore formal variation in incorporated elements in all types of incorporation constructions. This perspective has a number of implications for the exact types of structures that are included in the study. More specifically, the FDG approach is comprehensive with respect to the possible pragmatic characteristics of incorporated nouns, the phonological characteristics that incorporated elements may have, the parts of speech that can be involved in incorporation, the possible bound status of incorporated elements and their hosts, and the relation between incorporation and serial verb constructions and compounding. I will address each of these properties of the approach in turn.

In the first place, FDG does not pose any pragmatic restrictions on incorporation structures, as it does not specify which pragmatic units at IL can be involved in incorporation. Thus, the approach does not exclude constructions with or without particular pragmatic functions, such as reference or ascription. Importantly, this means that FDG does not limit the domain of noun incorporation in terms of the referentiality characteristics of the incorporated noun, in contrast to many other accounts.

Secondly, the FDG definition of incorporation does not require any phonological characteristics. As a consequence, not only incorporation constructions that form single phonological words but also constructions in which the incorporated element and its host remain phonologically independent of each other are included. Such constructions, which have been described as juxtaposition (Mithun 1984), loose incorporation (Miner 1986: 252) and pseudo-incorporation (Massam 2001, Massam 2009: 1087), are especially common in isolating languages such as the Oceanic languages (Mithun 1984: 849; Margetts 2008), which do generally not allow more than one morpheme per phonological word. An example from Niuean appears in (11).
In (11), the noun *ika* ‘fish’ is morphosyntactically incorporated into the verb *takafaga* ‘hunt’, as evidenced by the post-verbal clitics *tūmau* and *nī*, which here attach to the noun *ika* rather than directly to the verb *takafaga* (Seiter 1980: 69). In addition, the subject *ia* ‘he’ combines with the absolutive marker *a*, which indicates that the verb is intransitive and *ika* cannot be a verb-external direct object (Seiter 1980: 70). However, as their appearance as independent orthographic words shows, the noun and the verb in (11) remain separate phonological words (Massam 2001: 192). Nevertheless, in correspondence with studies as Mithun (1984: 849–854) and Aikhenvald (2007: 14), examples such as (11) are considered incorporation structures in FDG on the basis of their semantic and morphosyntactic characteristics.

The absence of phonological requirements for incorporation structures also entails that incorporated elements may show phonological alternations compared to their corresponding freestanding forms (see also Mithun 1984: 875–876). For instance, in the Munda languages the form of incorporated nouns, called the “combining form”, is typically a short version of the “full form” used for unincorporated nouns (Anderson 2007: 175–182). An example from Sora is the noun meaning ‘banana’, which has the full form *knte* but takes the combining form *-te* when it is incorporated (Anderson and Harrison 2008: 351), as shown in (12).

(12) Incorporation of a nominal stem with stem alternation into a verb in Sora

\[
\text{ɲen} \text{jum-te-ti-n-ai}
\]

‘I am eating a banana.’

(Anderson and Harrison 2008: 351)

While Example (12) from Sora indicates that an incorporated simple stem may have an alternate form, phonological alternations may also occur in more complex incorporated elements. Thus, in Crow the inflected noun *b-ashtá* ‘my eye(s)’ can be incorporated, as shown in (13), which includes the nominal stem *íshta* ‘eye’ that shows a vowel alternation (*i > á*) when it combines with the first person singular possessive prefix *b* (Graczyk 2007: 54–55).
Thus, stem alternations can occur in incorporated stems but also in more complex incorporated elements such as inflected words. Note also that the alternations do not have to be limited to the context of incorporation, as in the Sora example in (12), but may also be independent of the incorporation process, as in the Crow example in (13), in which the alternation is due to the presence of the first person singular possessive prefix b-. Alternations such as those in Sora and Crow that depend on the position of an element in the morphosyntactic structure, i.e., in an incorporation construction or in a particular possessive construction, are accounted for at PL (Hengeveld and Mackenzie 2008: 21), and are thus independent of the distinctive characteristics of incorporation which only pertain to RL and ML.

A third way in which the FDG perspective on incorporation is broad concerns the parts of speech that can be involved in incorporation. In FDG, the lexical morphemes, words and phrases that can occur in a word template may be nominal, verbal, adjectival, adverbial and adpositional (Hengeveld and Mackenzie 2008: 376, 401, 404). Consequently, incorporated elements and the hosts of incorporation, which are also part of word templates, may also belong to the classes of nouns, verbs, adjectives, adverbs and adpositions.9 Regarding the hosts of incorporation, however, most research focuses on incorporation into verbs (e.g., Baker 1988; Payne 1997: 231–233; Matthews 2007: 188). Although some studies show examples of incorporation of elements into nouns as well (Spencer 1995: 440 [for Chukchi]; Givón 2011: 194–196, 199–200 [for Ute-Southern Paiute]; Štekauer et al. 2012: 64), the present research follows the general trend in incorporation studies in considering incorporation into verbs only. With respect to incorporated elements, many studies address only or primarily nouns, i.e., they concentrate on noun incorporation (Gerdts 1998; Mithun 2000; Iturrioz Leza 2001; Massam 2009). Nevertheless, there are also several studies that discuss

9 Clauses do not correspond to a part of speech so the division between nouns, verbs, adjectives, adverbs and adpositions is not relevant for them.
constructions with incorporated elements of other parts of speech, such as incorporation of adpositions into verbs (Baker 1988: 229; Gerdt 1998: 84; Velupillai 2012a: 120–121), adjectives into verbs (Štekauer et al. 2012: 64), verbs into verbs (e.g., Baker 1988: 147; Payne 1997: 232; Dunn 1999: 231 [for Chukchi]; Evans 2003: 319 [for Bininj Kun-Wok]; Graczyk 2007: 300 [for Crow]) and adverbs into verbs (Rivero 1992 [for Greek]; Mithun 1994: 5024; Gerdt 1998: 84; Barrie and Mathieu 2016: 38 [for Algonquian languages]). Taking an FDG approach to incorporation, all these constructions can indeed be considered instantiations of incorporation. By contrast, the study does not include constructions that have been argued to involve the incorporation of pronouns or pronominal affixes (Mithun 1994: 5025, Mithun 2000: 922; Gerdt 1998: 84). Such constructions are unlike incorporation structures in involving the inclusion of a grammatical rather than a lexical element and can be considered verbs with cross-reference marking.

Fourthly, the FDG approach does not restrict incorporation to unbound elements, which is relevant for the issue of so-called denominal verb constructions and lexical affix constructions. These constructions closely resemble prototypical incorporation structures but, in contrast to most incorporation constructions, involve a bound morpheme (Mithun 1997, Mithun 1998, Mithun 1999: 48–56; Gerdt 1998: 94–98; Gerdt and Marlett 2008). In denominal verb constructions the host of the incorporation is a bound morpheme with a verb-like meaning, and in lexical affix constructions the incorporated element, called a lexical affix, is a bound noun-like morpheme. Denominal verb constructions occur mainly in North American and Eskimo languages (Mithun 1998, Mithun 1999: 54, Mithun 2009; Haugen 2007, 2008; Stonham 2008) and also in Chukchi (Kurebito 2001), as shown in (14). Lexical affix constructions, which are characteristic for a few northwestern American language families such as Salishan, Chimakuan and Wakashan (Mithun 1999: 54; Gerdt 1998: 94; Kinkade 1998: 266; Bischoff 2011: 1), are structures such as the one in (15) from Halkomelem.

(14) Denominal verb construction in Chukchi
t-irʔə-tw-ə-rkɐn
1SG.S-skin.coat-E-take.off-E-PRS
‘I am taking off my skin coat.’
(Kurebito 2001: 73)

(15) Lexical affix construction in Halkomelem
niʔ skʷ-ayał ɪə Mary.
AUX bathe-baby DET Mary
‘Mary bathed the/a baby.’
(Gerdt 2003: 347)
In (14), the element -tv ‘to take off’ has a verbal meaning but is a bound morpheme that obligatorily combines with an incorporated element. In (15), -ayəɫ ‘baby’ is a bound element with a noun-like meaning that necessarily attaches to another lexical morpheme.

Several researchers have argued that denominal verb constructions and lexical affix constructions are not incorporation structures but instead involve derivation, because the bound morphemes in these constructions show some similarities to derivational affixes (Sapir 1911; Mithun 1986, Mithun 1997: 364, Mithun 1999: 49–50, 54, 68–69; Stonham 2008: 513–514; Bischoff 2011: 15). However, others have emphasized the constructions’ many correspondences to incorporation structures with unbound morphemes such as their lexical function of creating new words for name-worthy activities (Mithun 1997: 364–365, Mithun 1999: 50–54, Mithun 2009: 11–12), their discourse function of backgrounding information that is already known (Mithun 1997: 364–365, Mithun 1999: 51–54, Mithun 2009: 12; Mathieu 2013: 117–118), and syntactic characteristics such as the possibility to strand modifiers of the incorporated element (Haugen 2007: 150, Haugen 2008: 439, 442; Muro 2008: 18; Mathieu 2013: 124–126) and the possibility to double the incorporated element in an external noun phrase (Czaykowska-Higgins et al. 1996: 33; Mithun 1997: 365; Haugen 2007: 150, Haugen 2008: 439, 445). Moreover, the bound morphemes in denominal verb constructions and lexical affix constructions are unlike derivational affixes in that they are typically quite concrete, i.e., lexical, in meaning (Czaykowska-Higgins et al. 1996: 29; Mithun 1997: 364, Mithun 1998: 63, 65, Mithun 1999: 48–49; Gerdt 2001: 94; Kurebito 2001: 54; Gerdts 2000: 63; Stonham 2008: 514) and very numerous (Czaykowska-Higgins et al. 1996: 29; Mithun 1997: 204, Mithun 1999: 54; Gerdt 2001: 94; Gerdt 2003: 346; Kurebito 2000; Stonham 2008: 514). In addition, incorporation constructions, on the one hand, and denominal verb constructions and lexical affix constructions, on the other hand, are often diachronically related. Incorporation constructions may develop into denominal verb constructions or lexical affix constructions, i.e., one of the lexical elements may become bound (Carlson 1990: 78–81). Similarly, denominal verb constructions may over time change into incorporation constructions (Jacques 2012: 1230).

Importantly, in several approaches, including FDG, a distinction is made between two types of lexical morphemes, i.e., unbound lexical morphemes and bound lexical morphemes. The latter obligatorily attach to other lexical morphemes. The latter obligatorily attach to other lexical morphemes. The latter obligatorily attach to other lexical

\[10\] Doubling is not possible in all languages with denominal verb constructions (Mathieu 2013: 127–128 [for Eastern Ojibwa]), but for some types of incorporation and in some languages with incorporation doubling is not possible either (i.e., for “compound NI” [Rosen 1989] or “type I incorporation” [Mithun 1984]) (Gerdt 1998: 95–96).
morphemes such that they necessarily occur in incorporation structures or other compounds (Hengeveld and Mackenzie 2008: 404; Delahunty and Garvey 2010: 132). In FDG, unbound lexical morphemes are called stems, while bound ones are termed roots. The recognition of bound lexical morphemes makes it possible to classify the bound morphemes with a lexical meaning in denominal verb constructions and lexical affix constructions as lexical morphemes. In this way, denominal verb constructions and lexical affix constructions can be analyzed as involving two lexical morphemes, rather than a lexical and a derivational morpheme, and do as such fulfil the morphosyntactic requirements for incorporation in FDG (Hengeveld and Mackenzie 2008: 414; Genee 2016: 1094). The present study does thus not exclude constructions that are like incorporation in containing two lexical-like elements but in which one of the two morphemes is bound.

A further characteristic of FDG’s broad approach towards incorporation is that some serial verb constructions are considered to involve incorporation as well, as the FDG definition also includes incorporation of a verb into another verb, which in some cases can also be described as verb serialization. Serial verbs may be defined as constructions in which two or more verbs combine in a single clause without the one being overtly subordinated to or coordinated with the other (Foley and Olson 1985: 18; Muysken and Veenstra 1994: 290; Crowley 2002: 10–11; Ansaldo 2006: 260–261; Velupillai 2012a: 332–333). The verbs typically share at least one argument and have the same tense, aspect, and mood values (Foley and Van Valin 1984: 189; Muysken and Veenstra 1994: 290; Durie 1997: 291; Ansaldo 2006: 261; Velupillai 2012a: 331). Although in many languages the two verbs in a serial verb construction are separate words, in some languages they combine to form a single morphosyntactic word (Foley and Olson 1985: 22–23; Crowley 2002: 15–16; Aikhenvald 2006: 37–38). When two serialized verbs form a single morphosyntactic word and additionally show a head-modifier or a nucleus-dependent relation at RL, they cannot be distinguished from incorporation structures. This is the case in Example (16) from Kalaallisut, in which kati ‘get married’ functions as an argument of ssamaar ‘plan’.

(16) Verb serialization/incorporation of a verbal stem into a verb in Kalaallisut

\textbf{kati-ssamaar-put}

get.married-plan-3PL.IND

‘They are planning to get married.’

(Fortescue 1984: 325)

When serialized verbs remain separate morphosyntactic words and/or show an equipollence relation rather than a dependency relation at RL, they can however clearly be distinguished from incorporation structures.
Finally, it should be mentioned that incorporation is here considered a type of compounding. In FDG, a distinction is made between lexical and grammatical compounding (Hengeveld and Mackenzie 2016: 1150).\textsuperscript{11} Compounding processes that take place in the lexicon have restricted productivity, may have idiosyncratic meanings, and their components cannot be modified separately or used referentially (Hengeveld and Mackenzie 2016: 1152–1153). As these compounds are formed in the lexicon, i.e., outside the grammar proper, they appear in the Grammatical Component as single items. They are thus different from incorporation constructions in that they consist of only one unit at RL and one unit at ML, and not of a combination of two. By contrast, compounding processes taking place in the grammar sometimes equal incorporation. Grammatical compounding processes are productive, create regular, compositional meanings and their components can be modified separately and might be referential (Hengeveld and Mackenzie 2016: 1151–1152). Three types of grammatical compounds are identified: head-modifier compounds, such as bookcase, in which book modifies case; predicate-argument compounds, such as truck driver, in which truck is an argument of drive; and conjunct-conjunct compounds, such as singer-composer (Hengeveld and Mackenzie 2016: 1151).\textsuperscript{12} The first two types of compounds are incorporation constructions, as they are not only morphosyntactic words consisting of two lexical units at ML, but also show a dependency relation at RL. Conjunct-conjunct compounds, by contrast, consist of two components that are in an equipollence relation at RL. As such, they are distinguished from incorporation on the basis of their RL properties. The position of incorporation in the classification of compounds in FDG can be presented schematically as in Figure 1.

In FDG, incorporation is thus differentiated from other types of compounding in two ways. Firstly, incorporation is different from lexical compounding in that it is a productive, semantically predictable process. Secondly, incorporation is different from conjunct-conjunct grammatical compounding in that incorporation involves a dependency relation rather than an equipollence relation between the host and the incorporated element.

\textsuperscript{11} This distinction between lexical and grammatical compounding is here related to two different engines, i.e., the lexicon and the grammar. However, it can also be found in single-engine approaches such as Distributed Morphology, in which the two types of compounds correspond to different syntactic structures (e.g., Harðarson 2018: 88–89; Steddy forthcoming.).

\textsuperscript{12} Note that of each of these three compounding types, i.e., head-modifier, predicate-argument and conjunct-conjunct, both endocentric and exocentric examples can be found (Scalise and Bisetto 2009; Hengeveld and Mackenzie 2016: 1153). For instance, a possessive compound like loudmouth is considered an exocentric compound of the head-modifier type.
3 Hypotheses

3.1 The possible forms of incorporated elements

The FDG approach to incorporation provides an important hypothesis about the crosslinguistic variation in the forms of incorporated elements, based on the maximal morphosyntactic word template presented in (10), repeated here in (17).

\[(Xw_1): [(Xm) (Xw) (Xp) (Cl)] (Xw_j)]\]

(Hengeveld and Mackenzie 2008: 400)

As this template specifies that words, including incorporation structures, may crosslinguistically include morphemes (Xm), words (Xw), phrases (Xp) and clauses (Cl), incorporated elements are predicted to be able to take any of these forms.

The simplest form of incorporation involves the incorporation of a single lexical morpheme. As discussed in Section 2.3, such a morpheme can either be an unbound simple stem, which is the most commonly investigated type of incorporated element, or a single bound root. FDG also recognizes grammatical morphemes, but these are never considered as incorporated morphemes because incorporated elements always contain lexical material.

A second type of incorporated element that is predicted to occur may be called the derived stem. A derived stem is like a single lexical morpheme in that...
inflectional affixes can become attached to it, but it consists of a lexical morpheme and another, lexical or grammatical, morpheme. This type of incorporated element is not included as a separate unit in the maximal template in (17), but it nevertheless follows from this template, as units in the template may occur more than once in a single word. It is therefore possible for an incorporation structure to include an element that has the form of two morphemes.

Derived stems can be divided into two subtypes. On the one hand, a derived stem may contain a lexical morpheme and a grammatical, derivational affix that changes the word class or some other morphosyntactic property of the lexical morpheme (Hengeveld and Mackenzie 2008: 228–229, Hengeveld and Mackenzie 2016: 1149–1150). Examples of such affixes are nominalizing affixes and verbal valency-changing affixes. The lexical morpheme and the derivational affix are combined at ML (Hengeveld and Mackenzie 2008: 228, 413), where they form a derived stem consisting of two morphemes. On the other hand, two lexical morphemes can combine to constitute a derived stem in the form of a grammatical compound, i.e., an incorporation structure or a conjunct-conjunct compound (see Section 2.3).

In addition to the incorporation of lexical morphemes and derived stems, lexical words, phrases and clauses are expected to occur as incorporated elements. A lexical word generally contains a simple or derived stem and one or more inflectional affixes, i.e., it is an inflected word. Of course, as follows from the template in (17), words can be more complex as well. At the same time, words may also be simpler, as in many languages not all parts of speech take inflectional marking, and in some languages inflection does not occur at all. A phrase can be described as a configuration of morphosyntactic words, which may be lexical and/or grammatical, other phrases and/or embedded clauses that typically appear next to each other in a clause (Hengeveld and Mackenzie 2008: 376). Finally, it is predicted that incorporated elements may be clauses, i.e., groupings of lexical and/or grammatical words, phrases and/or other clauses that typically have a fixed, organized order (Hengeveld and Mackenzie 2008: 293, 310). Clauses express nucleus-dependent relations, i.e., they contain at least a predicate, usually a verb, and an argument, typically a noun (Hengeveld and Mackenzie 2008: 294, 310).

13 FDG also recognizes derivational processes in the lexicon, which do not affect the word class or morphosyntactic characteristics of a lexical morpheme but only add semantic content. However, just like the compounds created in the lexicon discussed in Section 2.3, lexically derived stems are considered single morphemes and not incorporation structures at ML (Hengeveld and Mackenzie 2008: 229–230, 2016: 1150).
Note that in FDG all these different forms of incorporated elements belong to a single phenomenon of incorporation, which involves the occurrence of more than one lexical unit in a single morphosyntactic word. Thus, FDG proposes a unified account of the different forms of incorporated elements.

3.2 The crosslinguistic distribution of the various forms of incorporated elements

While FDG predicts that incorporated elements may crosslinguistically take many different forms, it is not expected that all of these forms occur in every language. Whether or not incorporation exists in a language and, if it does, which forms incorporated elements may take, depends on the inventory of word templates available in the language (see Section 2.2). The present study, therefore, also investigates how the different forms of incorporated elements are distributed over the languages of the world, and hypothesizes that this distribution can be described by an implicational hierarchy.

Crosslinguistic generalizations concerning the occurrence of linguistic features typically form implicational statements, which express that a particular feature A only exists in languages that also show feature B (Comrie 1989: 17; Croft 2003: 53; Hengeveld and Mackenzie 2008: 32; Velupillai 2012a: 33). When several features are in such an implicational relationship to each other, an implicational hierarchy may be formulated (Croft 2003: 122; Hengeveld and Mackenzie 2008: 33; Corbett 2011: 191; Aikhenvald and Dixon 2017: 8). Features higher on an implicational hierarchy occur only in languages that also show all the features lower on that hierarchy, and this distributional pattern may then be explained on the basis of a communicative or cognitive preference for the lower ordered elements over the higher ordered ones (Comrie 1989: 25–27; Hengeveld and Mackenzie 2008: 35). Even though implicational statements and hierarchies usually reflect tendencies rather than absolute universal patterns, they are highly valuable in clearly demonstrating which patterns are preferred (Comrie 1989: 19–20; Croft 2003: 51–52).

Implicational hierarchies are often related to the concept of markedness (Croft 2003: 87–121; Hengeveld and Mackenzie 2008: 34; Song 2018: 143–144). Elements that are ranked low on the hierarchies and thus seem to be favored in communication are then considered to be less marked than higher-ranked elements. These less marked, low-ranked elements are generally characterized by a high degree of cognitive or physical simplicity, such as a short, simple form (Croft 2003: 92; Haspelmath 2008: 213; Hengeveld and Mackenzie 2008: 35;
In addition, the less marked elements are typically more frequent than the more marked elements, both intra- and inter-linguistically (Croft 2003: 110; Hengeveld and Mackenzie 2008: 34–35; Bybee 2011: 134; Moravcsik 2013: 54).

Importantly, the forms of incorporated elements that the FDG word template predicts to occur in languages can clearly be ordered on the basis of their degree of complexity: single lexical morphemes, derived stems, inflected words, phrases and clauses may be argued to be increasingly complex based on their length and structure. At the same time, the different forms of incorporated elements appear to differ in frequency and, interestingly, forms with increasing complexity seem to show decreasing frequency. It is well known that the incorporation of simple stems, i.e., single lexical morphemes, is the most common type of incorporation (Smit 2005: 94). Incorporated inflected words are not infrequent either, but the incorporation of phrases occurs less often (Aikhenvald 2007: 12–13).

With respect to the crosslinguistic distribution of the different forms of incorporated elements, it may therefore be hypothesized that the forms constitute the implicational hierarchy presented in (18).

(18) Hypothesized implicational hierarchy of incorporated elements

clause
→
phrase
→
inflected word
→
derived stem
→
lexical morpheme

The incorporated forms highest on the hierarchy in (18), which are the most complex and the least frequent, are predicted to occur in a language only if the incorporated forms lower on the hierarchy, which are the simplest and most frequently incorporated forms, appear in that language as well. Importantly, the hierarchy suggests that the incorporation of more complex forms is related to the incorporation of simpler forms, which would support the claim that they are instances of the same phenomenon. Whether or not this hierarchy holds is thus an important test for the unified approach towards incorporation structures with incorporated elements of different forms that FDG proposes.
4 Method

4.1 Sampling procedure

The present research investigates the hypotheses about the forms of incorporated elements and their distribution on the basis of a typological study of 30 languages. For this investigation, only languages that show incorporation are relevant, and therefore the first step of the sampling procedure involved compiling a list of incorporating languages, from which the sample could be drawn. The starting point for this list was a survey of languages with noun incorporation provided by Velupillai (2012b), which is primarily based on typological studies and review articles on incorporation such as Mithun (1984), Gerdts (1998), and Aikhenvald (2007). Subsequently, I added to this list on the basis of other well-known studies of incorporation structures (Sapir 1911; Sadock 1980, Sadock 1985, Sadock 1986; Baker 1988, Baker 1996; Rosen 1989; Anderson 2000), a number of overview articles on incorporation (Mithun 1994, Mithun 2010; Iturrioz Leza 2001; Anderson 2007; Massam 2009), and a few crosslinguistic studies on incorporation (Caballero et al. 2008; Štekauer et al. 2012; Barrie and Mathieu 2016). The list was then further extended based on articles about incorporation structures that appear in the Linguistic Bibliography (Bobyleva et al. n.d.) and the Modern Language Association International Bibliography. This procedure yielded a list of 248 incorporating languages, which is included in Appendix 1. From this list I drew a so-called variety sample that is suitable to explore the variation that exists with respect to the forms of incorporated elements (Rijkhoff and Bakker 1998: 265; Croft 2003: 21; Bakker 2011: 104; Velupillai 2012a: 50). In order to capture the whole range of variation in incorporation structures, the sample is aimed to be representative for the existing genealogical, geographic, and typological diversity (Rijkhoff and Bakker 1998: 267–268; Croft 2003: 21; Bakker 2011: 104–5; Velupillai 2012a: 50).

To account for the genealogical variation, I followed the classification presented in Glottolog (Hammarström et al. 2017). As my list of incorporating languages contains languages from 69 different language families, plus 10 language isolates, not all language families and isolates in the list could be represented in the sample. The requirement of genealogical diversity was therefore easily met by simply selecting 30 languages that do not belong to the same family, such that correspondences between languages in the forms of their incorporated elements cannot be due to common ancestry.

In order to guarantee a representative geographical distribution, I calculated the proportion of language families in the list of incorporating languages from
each macro-area distinguished in Glottolog (Africa, Australia, Eurasia, North America, Papunesia and South America), and I selected the sample languages in such a way that the sample reflects these proportions, as shown in Appendix 2. Consequently, macro-areas including many families with incorporating languages are represented by more languages than macro-areas that contain only few such families. In addition, the location of the languages as indicated in Glottolog was taken into account in the sampling procedure in order to avoid selecting languages spoken in contiguous regions.

For this study, typological diversity means that all possible sets of forms of incorporated elements should be sufficiently represented. In order to test the hypothesized hierarchy, it is crucial that languages with forms high on this hierarchy, i.e., incorporating phrases and clauses, are included in the study. A few languages that have been claimed to show such incorporated elements, i.e., Bininj Kun-Wok (Hengeveld and Mackenzie 2008: 415), Crow (Barrie and Mathieu 2016: 33–34), Eastern Ojibwa (Barrie and Mathieu 2016: 17–18), and Chukchi (Hengeveld and Mackenzie 2008: 415–416), were therefore deliberately included in the sample.

Finally, the amount of available data was a point of consideration too. The evaluation of the proposed implicational hierarchy requires information about the incorporation structures in the languages that is as complete as possible. Therefore, languages for which extensive documentation about their incorporation structures exists were given precedence in the sampling procedure. In most cases, this documentation consists of papers on incorporation structures in the particular language and/or a reference grammar. The data sources used for the languages in the sample are included in Appendix 3. The sample of 30 languages used in the study is presented in Table 1.

### 4.2 Data analysis

In order to investigate which forms incorporated elements in the sample languages can take, the incorporation structures in the available data from each language were analyzed. In this analysis, I used the definitions of the forms given in Section 3.1. I also needed a few additional principles with respect to the use of the data in order to enable a consistent investigation of the incorporation structures.

In the first place, I assumed that a language shows a particular form of incorporation when at least one example of this type of incorporation or a statement about its existence could be found in the literature. The frequency of the type of incorporation was thus not taken into consideration.
Table 1: Languages included in the sample. The names of the languages, their family classifications, macro-areas and countries are based on Glottolog (Hammarström et al. 2017). Alternative names for the languages used in the data sources for the particular languages are included in square brackets.

<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bininj Kun-Wok [Bininj Gun-Wok, Gunwinggu, Mayali]</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Chimalapa Zoque [San Miguel Chimalapa Zoque]</td>
<td>Mixe-Zoque</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Chukchi [Chukchee]</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Crow</td>
<td>Siouan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Eastern Ojibwa [Central Ojibwa, Nishnaabemwin, Ojibwa, Ojibwe]</td>
<td>Algic</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Halkomelem [Halkomelem Salish, Musqueam]</td>
<td>Salishan</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Hokkaido Ainu [Ainu, Southern Hokkaido Ainu]</td>
<td>Ainu</td>
<td>Eurasia</td>
<td>Japan</td>
</tr>
<tr>
<td>Iraqw</td>
<td>Afro-Asiatic</td>
<td>Africa</td>
<td>Tanzania, United Republic of</td>
</tr>
<tr>
<td>Kalaallisut [Esperanto, Greenlandic, West Greenlandic]</td>
<td>Eskimo-Aleut</td>
<td>Eurasia</td>
<td>Greenland</td>
</tr>
<tr>
<td>Ket</td>
<td>Yeniseian</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Mapudungun [Mapuche, Mapudungan]</td>
<td>Araucanian</td>
<td>South America</td>
<td>Argentina; Chile</td>
</tr>
<tr>
<td>Mohawk [Akwesasne Mohawk]</td>
<td>Iroquoian</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Movima</td>
<td>Movima (Isolate)</td>
<td>South America</td>
<td>Bolivia, Plurinational State</td>
</tr>
<tr>
<td>Nadëb</td>
<td>Nadahup</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Niuean [Niue]</td>
<td>Austronesian</td>
<td>South America</td>
<td>Niue</td>
</tr>
<tr>
<td>Northern Gumuz</td>
<td>Gumuz</td>
<td>Africa</td>
<td>Ethiopia; Sudan</td>
</tr>
<tr>
<td>Nuu-chah-nulth [Kyuquot, Nootka, Nuuchahnulth]</td>
<td>Wakashan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Palikür [Palikur]</td>
<td>Arawakan</td>
<td>South America</td>
<td>Brazil; French Guiana</td>
</tr>
<tr>
<td>Panare</td>
<td>Cariban</td>
<td>South America</td>
<td>Venezuela, Bolivarian</td>
</tr>
<tr>
<td>Paraguayan Guaraní [Guaraní, Paraguayan Guarani]</td>
<td>Tupian</td>
<td>South America</td>
<td>Argentina; Paraguay</td>
</tr>
</tbody>
</table>

(continued)
Secondly, when I was unable to find examples of incorporated elements of some forms and an explicit statement about the existence of incorporated elements of that form could not be found either, I assumed that this language does not allow this form to be incorporated. This principle is of course not ideal, because the absence of incorporated elements of a particular form in one data source is no conclusive evidence for the non-existence of incorporated elements of that form in the language. Therefore, I aimed to make use of as much data of each language as possible, which is the main reason that comprehensive documentation of the process of incorporation in each of the sample languages was required. In addition, where possible I consulted experts on the relevant languages in order to verify my assumptions about the impossibility to incorporate particular forms.

Thirdly, some extra principles were needed in the analysis of incorporated words, phrases and clauses. As described in Section 2.3, a lexical word typically contains one or more inflectional affixes, but not all words in all languages show inflection. Nevertheless, in the present study inflectional marking was seen as necessary in the identification of lexical words, in order to make it possible to consistently distinguish the incorporation of words from the incorporation of simpler forms, i.e., single lexical morphemes and derived stems. Similarly, although a phrase may consist of a single word, in the analysis only those phrases that contain at least two words or other units were considered phrases, such that I
could clearly differentiate between phrases and words. Also for incorporated clauses it was required that they contain at least two words. In many polysynthetic languages, a clause may consist of a single verbal inflected word, as the arguments may be represented by pronominal affixes (Evans and Sasse 2002: 2; Aikhenvald 2007: 5–6; Murasugi 2014: 293). Although such words can fulfil the same functions as multi-word clauses, they were considered inflected words rather than clauses in the analysis, as it was impossible to prove that they were incorporated as full clauses and not simply as inflected words.

Fourthly, some languages simply lack particular forms altogether, and not only in incorporation. For instance, there are languages without inflectional morphology. It makes little sense to investigate whether such languages allow the incorporation of inflected words. In such cases, I marked the form as irrelevant in the evaluation of the hypothesized implicational hierarchy for these languages. Some languages do show all forms but not for the parts of speech for which complex forms can be incorporated. For these languages, I applied the same strategy: simpler forms that the relevant parts of speech lack were considered irrelevant in the investigation of the hierarchy.

Finally, with respect to the occurrence of incorporated inflected words a distinction has to be made between syntactically active and frozen inflectional marking. In the case of frozen inflection, the inflectional marking is not meaningful synchronically and typically only expresses a default value. Such inflection can be found in Example (19) from Kalaallisut.

(19) Incorporation of a nominal stem into a verb including frozen allative case marking in Kalaallisut

\[juulip\ kingurn-a-gut\ Nuum-mukar-puq\]

July-RC.SG after-3SG.POSS-PROS Nuuk-go.to-3SG.IND

‘After July he went to Nuuk.’

(Fortescue 1984: 245)

According to Fortescue (1984: 245, 300), Example (19) contains a verbal suffix -mukar, which in FDG is considered a verbal bound root. This morpheme -mukar probably originated as a combination of two morphemes (Fortescue 2017: p.c.): the allative singular suffix -mut, which loses its -t when it is followed by a morpheme that starts with k (Fortescue 1984: 351), and a morpheme -kar meaning ‘go’. This suggests that a noun with allative case marking could be incorporated into -kar. However, synchronically -mu(t) and -kar are recognized as a single morpheme -mukar, i.e., it is not possible to use -kar without -mu(t). Therefore, -mu- in (19) may be considered a frozen inflectional suffix. I did not consider such frozen affixes as evidence for the possibility to incorporate...
inflected words, and Example (19) was therefore not analyzed as involving incorporation of an inflected word *Nuum-mu*. Instead, *Nuum* was regarded as an incorporated simple stem, of which many other examples are found in the language as well, with the verb *-mukar* as its host.

By contrast, the reflexive possessive inflection used on the incorporated noun in (5), repeated here as (20), is syntactically active and was therefore considered evidence for the incorporation of inflected nouns in Kalaallisut.

(20) Incorporation of a nominal inflected word into a verb in Kalaallisut

```
illu-mi-niip-puq
```

house-REFL.POSS-be.in-3SG.IND

‘He is in his (own) house.’

(Fortescue 1984: 300–301)

Nouns with other, non-reflexive possessive marking can appear incorporated in *-miit* ‘be in’ as well, as shown in (21).\(^\text{14}\)

(21) Incorporation of a nominal inflected word into a verb in Kalaallisut

```
Anta-p Aani-l = lu irnr-an-niip-puq
```

Anta-RC.SG Aani-RC.SG = and son-3PL.POSS-be.at-3SG.IND

‘She is at the house of Anta and Aani’s son.’

(Fortescue 1984: 135)

Moreover, the translations show that the possessive marking on the incorporated nouns is meaningful. The possessive marking on the incorporated nouns in (20) and (21) thus provided evidence for the possibility to incorporate inflected nouns in Kalaallisut.

## 5 Results

The hypotheses concerning the variation in the forms of incorporated elements and the crosslinguistic distribution of these forms were studied on the basis of the 30 sample languages, and the results of this study are presented in Table 2. The full set of data on which this table is based can be found on [http://dx.doi.org/10.21942/uva.6834188](http://dx.doi.org/10.21942/uva.6834188).

---

\(^\text{14}\) The verbal bound root *-miit* takes the form *-niip* when it follows a possessed form and precedes the inflectional suffix *-puq* (Fortescue 1984: 78, 334), as in (20) and (21).
The possible forms of incorporated elements

Table 2 shows clearly that most of the forms that were hypothesized to occur as forms of incorporated elements are indeed found as such in the sample languages: incorporated lexical morphemes, derived stems, inflected words, and
phrases are all attested in several languages. As expected, the languages differ in which of these forms their incorporated elements may take, in correspondence with the hypothesis that the inventories of word templates and thus the sets of forms of incorporated elements that languages allow are language-specific. However, in contrast to the prediction, no examples of incorporated clauses are attested in the data.

With respect to each of the forms of incorporated elements investigated in the sample a few observations can be made. First of all, Table 2 shows that all languages in the sample allow single lexical morphemes to be incorporated. The sample languages display nominal, verbal, adjectival, adverbial, as well as adpositional incorporated morphemes, i.e., incorporated morphemes of all lexical parts of speech. For instance, Example (22) shows the incorporation of an adjectival stem in Chimalapa Zoque, whereas Yucatec Maya allows the incorporation of an adverbial stem, as in (23).

(22) Incorporation of an adjectival stem into a verb in Chimalapa Zoque

\[
\text{nmma } \bar{\text{?ay}} = \bar{\text{w\thetaa}}-\bar{\text{c\k-suk-w\theta}}
\]

\text{PROG 3.ERG = good-do-2/3PL-DINC}

‘They were repairing it.’

(Johnson 2000: 278)

(23) Incorporation of an adverbial stem into a verb in Yucatec Maya

\[
\text{Le’ } t’eel-o’ \quad [...] \quad \text{ken } \bar{\text{uy-u(’u)b a-t’aan-e’}}
\]

\text{DEF hen-DDEIX SUB.FUT 3.SBJ-hear 2.SBJ-speak-LNDEIX}

\text{k-u = chen-t’aan}

\text{HAB-3.SBJ = just-speak}

‘The rooster [...] when he hears you speaking, he just speaks.’


In (22) \text{w\thetaa} ‘good’ is a modifier of the head \text{c\k} ‘do’, while in (23) \text{chen} ‘just’ modifies \text{t’aan} ‘speak’.

Incorporated lexical morphemes in the sample languages are unbound stems, as in (22) and (23), as well as bound roots, as exemplified for Movima in (24). Here, the incorporated element \text{-mo} ‘bird’ is a nominal bound morpheme (Haude 2006: 220), functioning as an argument of the verb \text{yok} ‘catch’.

---

15 The incorporated element \text{w\thetaa} ‘good’ is an adjective (Johnson 2000: 68, 278), although it functions as an adverb here, modifying the verb (Johnson 2000: 278).
(24) Incorporation of a nominal bound root into a verb in Movima

\[
\text{asko yok-a-mo-na} = \text{ney} = \text{s kara’ di’ sere:re}
\]

3SG.N.AB catch-DR-BE.bird-DR = PL.AB here = DET red.macaw REL wild

‘That [forest isle] (was) where they caught those wild red macaws.’

(Haude 2006: 326)

Especially incorporated unbound simple stems appear to be common, as they occur in all sample languages. In this way, the study confirms that stem incorporation is the prototypical type of incorporation, as is argued in many incorporation studies.

Most languages also display formally more complex incorporated elements. In 22 languages incorporated derived stems are found. Both stems derived with grammatical derivational affixes (25) and stems in the form of grammatical compounds (26) occur in incorporation structures.

(25) Incorporation of a nominal derived stem, consisting of a stem and a grammatical affix, into a verb in Crow

\[
\text{baa-luúsh-chiili}
\]

INDF-eat-look.for

‘look for food’

(Graczyk 2007: 281)

(26) Incorporation of a verbal derived stem, in the form of a grammatical compound, into a verb in Chimalapa Zoque

\[
də = \text{kahwe-ʔuk-tuk-wə}
\]

1.ABS = coffee-drink-finish-COMPL

‘I finished drinking coffee.’

(Johnson 2000: 221)

The incorporated element in the Crow example in (25) consists of the verb luúsh ‘to eat’ and a nominalizing indefinite prefix baa- (Graczyk 2007: 48, 50) and is an argument of its host, chiili ‘look for’ (Graczyk 2007: 297). The Chimalapa Zoque verbal compound kahwe-ʔuk ‘to drink coffee’ in (26) is an incorporation structure, i.e., a grammatical compound (Johnson 2000: 275), which is again incorporated into the verb tuk ‘finish’ as an argument of this verb.

Most derived stems occurring in incorporation structures in the sample languages are nominal or verbal, but Yimas shows incorporated derived adverbs. Thus, the Yimas derived adverb mampi ‘again’, consisting of the adjective ma ‘other’ and the adverbializing suffix -mpi (Foley 1991: 343), is incorporated in (27), in which it modifies the predicate wampunjka ‘angry’.
Incorporation of a adverbial derived stem into a verb in Yimas

\textit{na-n\-\textit{ma-mpi}-ira-wampun\textit{kra-ntut}}

3SG.P-3SG.A-other-ADVZ-ALL-angry-RMP

‘He was angry with her again.’

(Foley 1991: 336)

Several languages also show incorporated inflected words in the form of nouns with number, noun class agreement, possessive or reflexive marking and verbs marked for tense, aspect or person and number of their arguments. Incorporated nouns with number marking can be found in Ket. Example (28) shows the plural argument \textit{don'\-a\-ŋ} ‘knives’ incorporated into its predicate, the verb \textit{vet} ‘make’.

(28) Incorporation of a nominal inflected word in Ket

\textit{d-\textit{don'\-a\-ŋ}s'-i-\textit{vet}}

1SG.SBJ-knife-PL-PRS-E-make

‘I’m making knives.’

(Drossard 2002: 235)

The incorporation of a verb with aspect marking is exemplified in (29) from Kalaallisut. Here the verb \textit{nillir-sima} ‘had been cold’, part of the dependent \textit{imiq nillir-sima} ‘the water had been cold’, is incorporated into the nucleus \textit{nirar} ‘say that’.

(29) Incorporation of a verbal inflected word in Kalaallisut

\textit{imiq \textit{nillir-sima}-nirar-paa}

water be.cold-PFV-say.that-3SG \rightarrow 3SG.IND

‘He said the water had been cold (e. g., the day before).’

(Fortescue 1984: 274)

In addition, in South Slavey, adpositions with a pronominal object affix can be incorporated (Rice 1989: 741), such as the postposition \textit{wq} ‘to’ with the first person singular affix \textit{se-} (Rice 1989: 269), modifying \textit{tlah} ‘go’, in (30).

(30) Incorporation of an adpositional inflected word into a verb in South Slavey

\textit{se-wq-e-\textit{tlah}}

1SG.OBJ-to-ASP.go.SG/DU

‘S/he came to me.’

(Rice 1989: 766)

Interestingly, most incorporated inflected words contain inherent inflection, which represents relatively concrete information that is independent of the syntactic context in which the words occur, such as nominal number and verbal tense and
aspect (Booij 1996: 2; Haspelmath and Sims 2010: 100–101). The only examples of incorporated inflected words with contextual inflection, expressing values that are dependent on other words in the context (Booij 1996: 2; Haspelmath and Sims 2010: 100–101), are found in Bininj Kun-Wok and Yimas. In Bininj Kun-Wok, incorporated verbs appear in a special gerundive form (Evans 2003: 536). An example is shown in (31), which includes the incorporated element wayini-hmi ‘singing’, a modifier of the host re ‘go’, in which -hmi is the gerundive suffix, glossed by Evans (2003) as IVF ‘incorporating verb form’. Note that the gerundive suffix has different allomorphs for each of the conjugation classes in the language (Evans 2003: 538).

(31) Incorporation of a verbal inflected word into a verb in Bininj Kun-Wok

Ga-wayini-hmi-re.
3-sing-IVF-go.NPST

‘He goes along singing all the way.’
(Evans 2003: 543)

It thus turns out that incorporated inflected words with contextual inflection are rare in the sample languages. However, only about half of the sample languages show contextual inflection on the parts of speech that they allow as incorporated elements at all, such that for many sample languages investigating the incorporation of contextually inflected words is simply irrelevant.

The most complex incorporated elements found in the study are phrases, which occur in seven sample languages. Five of them allow incorporated noun phrases, exemplified in (32), and three show incorporated adposition phrases, illustrated in (33).

(32) Incorporation of a noun phrase into a verb in Niuean

 [...] ke kumi mena ke nonofo = ai a lautolu.
SBJV seek thing SBJV settle = there ABS they

‘[...] that they would seek a place to settle.’

---

16 Incorporated elements and their hosts remain independent phonological words in Niuean. Evidence for incorporation can be found in the word order and case marking. Firstly, because Niuean is a VSO language (Massam 2001: 155) and in (32) the subject is a lautolu ‘they’, the preceding mena ke nonofo = ai ‘a place to settle’ cannot be an unincorporated object. Secondly, the absolutive case-marking a on the subject lautolu shows that the clause is intransitive, which also supports the analysis of mena ke nonofo = ai as incorporated element rather than morphosyntactically independent object phrase. Thirdly, mena ke nonofo = ai itself does not show case marking, whereas independent noun phrases in Niuean are typically required to combine with such marking (Massam 2001: 157).
Incorporation of an adposition phrase into a verb in Crow

(33) \( ash\text{-}bachee\text{\textemdash}itchi-m \) Apsáalooke \( kuxshi\text{-}ka\text{\textemdash}ak \)

lodge\text{-}chief\text{-}SIM Crows help\text{-}AUG\text{-}SS

\( ak = baa\text{\textemdash}iláp\text{\textemdash}xisaakh\text{\textemdash}kuua = ss\text{-}dee\text{-}sh \) héelee\text{-}la\text{-}k

REL = Washington = GOAL\text{-}go\text{-}DET among\text{-}be\text{-}at\text{-}DECL

‘As a reservation chief he really helped the Crows, he was among those who went to Washington [as tribal delegates].’

(Graczyk 2007: 412)

The incorporated phrase in the Niuean example in (32) contains a noun, \textit{mena} ‘thing’, and a subjunctive relative clause \( ke\ text{nonofo} = \text{ai} ‘\text{to settle there}’ \) (Massam 2001: 161). The Crow incorporated element in (33) is considered a phrase because the postposition = \( ss \) ‘goal’ is a clitic, i.e., an independent morphosyntactic word (Payne 1997: 22; Hengeveld and Mackenzie 2008: 332; Haspelmath and Sims 2010: 196), which here forms a phrase with \( baa\text{iláp\textemdash}xisaakh\text{kuua} \) ‘Washington’ (Graczyk 2007: 372). This phrase is again preceded by the relativizer \( ak = \), which is a verbal proclitic (Graczyk 2007: 258).

In addition to these seven languages with clear examples of phrase incorporation, one other sample language, Chukchi, shows examples of incorporated elements for which it has not been possible to verify whether they are phrases or derived stems. Chukchi shows both incorporation constructions and phrases consisting of an adjective and a noun (Dunn 1999: 159). Consequently, the incorporated element in Example (34), \( tor-ta\text{-}ŋ-pəlwəntə-jo\text{\textemdash}gə \) ‘good, new, metal spear’ could be an incorporated phrase consisting of three adjectives and a noun, but it could also be a noun in which three adjectives are incorporated.\(^{17}\)

(34) Incorporation of a nominal derived stem, in the form of a grammatical compound into a verb/incorporation of a noun phrase into a verb in Chukchi

\( tə\text{-}tə\text{-}tor-ta\text{-}ŋ-pəlwəntə-jo\text{\textemdash}gə-pəla-rkən \)

1SG.S\text{-}new\text{-}good\text{-}metal\text{-}spear\text{-}leave\text{-}PRS

‘I am leaving a good, new, metal spear.’


\(^{17}\) Note that generally only absolutive-marked nouns can form phrases containing an adjectival modifier in Chukchi; in non-absolutive noun phrases the modifiers have to be incorporated into the noun (Dunn 1999: 159). However, incorporated elements never carry case marking in this language (Muravyova 1998: 522; Dunn 2017: p.c.). It is thus not unexpected that the incorporated element in (34) does not show absolutive case-marking, even though it functions as a direct object. The absence of absolutive case-marking does thus not exclude the possibility that the incorporated element is really a phrase.
As it is not possible to identify the incorporated element *tor-taŋ-palwənta-pojaŋ* ‘good, new, metal spear’ in (34) either as an incorporation structure or as a phrase, this element is not considered as evidence for either derived stem or phrase incorporation. Thus, my analysis is cautious and strict in order not to assume that complex incorporated elements are allowed in a language without having unambiguous evidence for the existence of such forms.

Most of the expected forms of incorporated elements are thus found in the sample. However, in contrast to the prediction, none of the sample languages show indisputable examples of incorporated clauses, even though a few languages were deliberately included in the sample because earlier literature had described them as clause-incorporating languages (see Section 4.1). Several explanations for this finding may be proposed. First of all, it is possible that clauses indeed cannot be incorporated, such that the FDG maximal word template described in Section 2.2 and 3.1 is somewhat too broad with respect to the forms of incorporated elements. A second possibility is that clause incorporation is not found in the data because it is very rare. This explanation is plausible because clauses would be a highly complex type of incorporated element and were therefore expected to be infrequent. Consequently, incorporated clauses possibly only occur in a few languages, which happen not to be included in the sample of the present study. Alternatively, the incorporation of clauses is very infrequent even in languages that do allow it, i.e., language-internally, such that examples of clause incorporation may simply not have been included in the consulted literature. In the third place, the present study only focuses on the incorporation of elements into verbs. It cannot be excluded that clauses can be incorporated into other parts of speech.

Most importantly, however, it should be noted that it is very difficult to find decisive evidence for clause incorporation, as possible cases of clause incorporation often have an alternative analysis as well. As was already mentioned in Section 3.1, in this study incorporated clauses consisting of a single verbal word with referential person affixes are not considered evidence for clause incorporation, as these can also simply be seen as inflected words. Thus, Example (35) from Crow is not included as a case of clause incorporation.

(35) Incorporation of a verbal inflected word into a verb in Crow

```
[...] “d-iłu-ń dii-lii-wah-kúnnaa-wuu-o-k”
```

```
2.ACSBJ-stand-IMP 2.PRO-2.OBJ-1.ACSBJ-fetch-1.ACSBJ.PL.come-1.PL-DECL
huu-k
say.PL-DECL

‘[...] “stand up, we have come to fetch you” they said.’
```

(Graczyk 2007: 313)
In this example the incorporated element *dii-lii-wah-kúnnaa* ‘we fetch you’ contains a verb with a first person subject prefix *wah-* and a second person object prefix *lii-* combined with the bound emphatic second person proform *dii-* (Graczyk 2007: 61). Because this incorporated element contains a verbal predicate, a subject and an object, it seems to function as a full subordinate clause. Moreover, in Crow, clauses often consist of only a verb with subject and object prefixes (Graczyk 2009: 269). These are good reasons to consider Example (35) to involve clause incorporation. At the same time, however, the incorporated element is formally a single inflected word, i.e., a verb with a subject prefix and an object prefix. Considering cases like (35) as incorporated clauses would thus make it impossible to distinguish between incorporated verbal inflected words and incorporated clauses. Consequently, this example is not considered a case of clause incorporation in the present study.

Similarly, some languages show examples of the incorporation of a verbal predicate and a nominal argument, which are possible cases of clause incorporation. However, here it is often not possible to determine whether an example involves the incorporation of a clause or rather of a verbal predicate that itself has incorporated its argument. Interestingly, this type of construction is found in Bininj Kun-Wok and Chukchi, both of which have been claimed to show clause incorporation in other studies and were included in the sample for that reason (see Section 4.1). Consider, for instance, Example (36) from Bininj Kun-Wok.

(36) Incorporation of a verbal derived stem, in the form of a grammatical compound into a verb/incorporation of a clause into a verb in Bininj Kun-Wok

*Ga-ganj-ngu-nihmi-re.*

3-meat-eat-IVF-go.NPST

‘He goes along eating meat.’

(Evans 2003: 536)

Evans (2003: 536) describes this example as an incorporation construction in which the verb *re* ‘go’ has incorporated the verb *ngu* ‘eat’ with its gerundive suffix *-nihmi*, which itself has incorporated its direct object *ganj* ‘meat’. This description suggests that Example (36) should be analyzed as a case of incorporation of an incorporation structure, i.e., a derived stem. However, the elements that are incorporated can also form independent words in a clause outside the incorporation structure, as evidenced by Example (37), such that the incorporated element in (36) could also be analyzed as a clause consisting of a verbal word, the predicate, and a nominal word, the direct object, without their normal morphological marking.
As both analyses are plausible, Example (36) is not regarded as evidence for clause incorporation, in order not to ascribe more complex incorporated elements to languages than they actually have.

In the same way, Example (38) from Chukchi cannot be used to show that clause incorporation is possible in this language.

Example (38) can be analyzed in various ways. First of all, because the incorporated adverb *mac* ‘almost’ semantically modifies the host of the incorporation construction, *platko* ‘finish’, rather than the incorporated verb *garke* ‘hunt’ (Spencer 1995: 461), it seems likely that this construction involves two incorporated elements, i.e., the adverb *mac* ‘almost’ and the combination of the verbal predicate *garke* ‘hunt’ and its nominal argument *qora* ‘deer’. Indeed, Spencer (1995: 459) discusses this example as a case of “multiple incorporation”. The adverb *mac* ‘almost’ thus seems to be incorporated as a single adverbial morpheme. However, because *garke* ‘hunt’ and *qora* ‘deer’ are in a predicate-argument relation, these two morphemes should be considered to form a single incorporated element. This element could then either be an incorporation construction in which *qora* is incorporated into *garke* or a clause consisting of a nominal word *qora* and a verbal word *garke*, in which case it has to be assumed that the inflection on both words is left out. This assumption matches the general observation that incorporated elements in Chukchi do not show inflectional morphology (Muravyova 1998: 522, 535; Dunn 2017: p.c.). Just as in the

---

18 Note that Dunn (1999: 141) actually describes *mac* as an approximative verbal prefix rather than as a lexical adverb. According to that analysis, the combination of *qora* ‘deer’ and *garke* ‘hunt’ is the only incorporated element in Example (38).
case of Example (36) from Bininj Kun-Wok, there is no way to determine whether the incorporated element *qora-gərke* is an incorporation structure, i.e., a derived stem, or a clause, such that Example (38) is not regarded as evidence for clause incorporation in Chukchi.

The identification of clause incorporation is thus complicated because potential examples can be analyzed in different ways. For this reason, even in the sample languages that were specifically selected because they were expected to show clause incorporation, i.e., Bininj Kun-Wok and Chukchi, incorporated clauses could not easily be identified. This observation leads to the question what an unambiguous case of clause incorporation would look like. In theory, several types of constructions could provide clear evidence for clause incorporation. Firstly, restrictions on the forms of elements when incorporated by themselves could be informative. For example, if a language does not allow incorporated nouns to take case marking, but does allow the incorporation of a noun with case marking together with its verbal predicate, such a case-marked noun must be incorporated as part of a clause. Similarly, if a noun shows an alternation when used as an incorporated stem, as most nouns in Sora do for example (see Section 2.3), but this noun does not show this alternation when it is incorporated with a verb that functions as its predicate, the resulting construction must be clause incorporation. If the noun had been part of an incorporated incorporation construction, it would have shown the alternation.

Secondly, the presence of clitics within the incorporated element may be decisive in the analysis of an incorporation construction. Because clitics are phonologically dependent on their host but, at the same time, morphosyntactically independent (Payne 1997: 22; Hengeveld and Mackenzie 2008: 332; Haspelmath and Sims 2010: 196), they can mark the boundaries of morphosyntactic words. For instance, if a language makes use of proclitics on verbs, such proclitics clearly mark the left boundary of the morphosyntactic verb. If an incorporated element contains a noun, followed by a verb with a proclitic, this incorporated element must be an incorporated clause, because the proclitic shows that the noun cannot be analyzed as being incorporated into the verb with the proclitic. If that were the case, the proclitic would show up to the left of the incorporated noun.

Finally, it should be mentioned that despite the absence of direct evidence for the incorporation of clauses, the data do provide an indication that incorporated elements as complex as clauses exist. Consider for instance Example (32) from Niuean above. In this example the incorporated element *mena ke nonofo = ai* ‘a place to settle’ is a noun phrase consisting of a head noun *mena* ‘thing’ and a relative clause *ke nonofo = ai* ‘to settle’. This example thus contains an incorporated clause. However, in the same way as an incorporated
phrase including an inflected noun would only be regarded as evidence for phrase incorporation and not for incorporation of an inflected word, Example (32) counts as an example of phrase incorporation rather than of clause incorporation, because the incorporated element as a whole is a phrase.

5.2 The crosslinguistic distribution of the various forms of incorporated elements

The second hypothesis presented in Section 3 was that the various forms that incorporated elements may take constitute the implicational hierarchy presented in (18), repeated here as (39).

(39) Hypothesized implicational hierarchy of incorporated elements

clause
→ phrase
→ inflected word
→ derived stem
→ lexical morpheme

Table 2 above, in which the languages are ordered on the basis of the allowed complexity of the forms of their incorporated elements, shows that this hypothesis regarding the distribution of the forms of incorporated elements is completely borne out. All sample languages show a contiguous area on the proposed hierarchy, i.e., languages that show more complex incorporated elements also show incorporated elements of all less complex forms. A few languages do appear to show gaps on the hierarchy, but for these languages these gaps are no counterexamples to the hierarchy. For instance, Nadëb and Niuean do not allow the incorporation of inflected words while they do use incorporated phrases. However, because these languages, being highly isolating, lack inflection in general (Weir 1990; Massam 2005: 227), the absence of incorporated inflected words is expected. In the same way, the absence of incorporated inflected words in Mapudungun is not problematic. In Mapudungun, the more complex incorporated elements are noun phrases, and subject and object nouns, which are the nouns that can be incorporated in this language (Smeets 2008: 318–319), do not take inflectional morphology in Mapudungun (Smeets 2008: 61), such that for
Mapudungun inflected words are simply not relevant in the evaluation of the hierarchy. Similarly, the lack of incorporated derived stems and incorporated inflected words in Northern Gumuz does not affect the evaluation of the hierarchy. In Northern Gumuz incorporated phrases are all adposition phrases, and adpositions in this language do not take derivational and/or inflectional morphology, which makes the absence of incorporated derived stems and inflected words in Northern Gumuz irrelevant with respect to the hierarchy. Finally, as discussed in Section 5.1, none of the sample languages decisively show incorporated clauses. However, as the clause is the most complex form that was predicted to be incorporated, it occurs furthest to the right on the hierarchy, and the lack of incorporated clauses in the data cannot provide counterexamples to the hierarchy. The data thus provide strong evidence for the proposed implicational hierarchy.

6 Conclusion

The aim of the present study has been to survey the forms that incorporated elements may take crosslinguistically and to present a unified account of these different forms. The research has adopted the FDG approach to incorporation, in which the incorporation of elements of different forms is considered a single phenomenon (Hengeveld and Mackenzie 2008). FDG hypothesizes that incorporated elements may take the following forms, in order of increasing formal complexity: single lexical morphemes, derived stems, inflected words, phrases and clauses. In addition, it is hypothesized in the study that these forms constitute an implicational hierarchy, i.e., that the incorporation of more complex forms only occurs in languages that also allow the incorporation of all simpler forms. These hypotheses about the forms of incorporated elements and their distribution have been tested on the basis of data from a variety sample of 30 incorporating languages.

The study has shown a wide range of diversity in the forms of incorporated elements, including incorporated single lexical morphemes, derived stems, inflected words and phrases. This finding is largely in correspondence with the hypothesis about the possible forms of incorporated elements. The only deviation from this hypothesis concerns the incorporation of clauses, for which the present research has not found any conclusive evidence. In addition, the research demonstrates that the different forms of incorporated elements are not randomly distributed over languages but instead show a pattern corresponding to the proposed implicational hierarchy. The incorporation of a particular
complex form occurs only in languages that also allow the incorporation of all less complex forms. Languages differ in the number and types of forms that they allow for incorporated elements, but all sample languages are in agreement with the hierarchy.

These findings about the forms of incorporated elements and their distribution have important theoretical consequences. Firstly, they show that incorporated elements may take highly varied forms and that incorporation is not limited to simple stems, even though traditional theoretical approaches have claimed that the simple stem is the only possible form of incorporated elements (e. g., Mithun 1984; Baker 1988). Importantly, the data from the 30 languages demonstrate that incorporated elements with different degrees of complexity occur in various languages and that they thus are relevant crosslinguistically. In this way, the study adds to previous studies that discuss incorporated elements of different forms like Aikhenvald (2007), Muro (2009) and Barrie and Mathieu (2016). Secondly, the clear distributional pattern of incorporated elements of different forms found in the study provides important support for a unified treatment of the incorporation of these formally diverse elements. It had already been shown that incorporation constructions with simpler and more complex forms are highly alike in appearance and that they are similar in that they both allow modifier stranding and the presence of an external possessor. However, the present study additionally demonstrates that the incorporation of simple and complex forms is interrelated in that formally more complex incorporated elements only occur in languages that also allow simpler incorporated forms. By revealing this pattern, the study supports the unified treatment of incorporated elements of different forms, as proposed in FDG.

An issue that remains for further research is that of clause incorporation. The lack of clear examples of incorporated clauses in the present study may indicate that morphosyntactic words do not allow as much complexity as predicted. On the other hand, several examples for which one of the possible analyses is clause incorporation are found. In addition, the attestation of incorporated phrases which include relative clauses suggests that incorporated elements as complex as clauses do exist. Moreover, it is possible that incorporated clauses are only found in languages not included in the sample or in constructions with non-verbal hosts that have not been investigated in this study. Thus, further research, especially on other incorporating languages and/or on incorporation constructions with non-verbal hosts, may show whether incorporated clauses perhaps nevertheless occur.

Acknowledgements: I would like to thank Jan Don, Kees Hengeveld, Frank Seifart and two anonymous Linguistics reviewers for their constructive feedback.
I am indebted to Diane Massam, Donald Frantz, Colleen Ahland, Michael Dunn, Thomas Payne, Fernando Zúñiga, William Foley, Keren Rice, Stefan Georg, Gregory Anderson, Michael Fortescue, Katharina Haude, Anna Bugaeva, Christian Lehmann and Tom Givón for their help on their languages of expertise. I am also grateful to Ulrika Klomp for commenting on the pre-final version of this paper. This work is part of the research program PhDs in the Humanities, project number 322-70-011, which is financed by the Netherlands Organization for Scientific Research (NWO).

**Appendix 1: List of incorporating languages**

Incorporating languages, identified as described in Section 4.1. The names of the languages, their family classifications, macro-areas and countries are based on Glottolog (Hammarström et al. 2017). “*” indicates that a particular language is considered a language (sub)family rather than a single language in Glottolog.

<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alagwa</td>
<td>Afro-Asiatic</td>
<td>Africa</td>
<td>Tanzania, United Republic of Ethiopia</td>
</tr>
<tr>
<td>Gawwada</td>
<td>Afro-Asiatic</td>
<td>Africa</td>
<td>Ethiopia</td>
</tr>
<tr>
<td>Iraqw</td>
<td>Afro-Asiatic</td>
<td>Africa</td>
<td>Tanzania, United Republic of Djibouti; Ethiopia; Kenya; Somalia</td>
</tr>
<tr>
<td>Somali</td>
<td>Afro-Asiatic</td>
<td>Africa</td>
<td>Djibouti; Ethiopia; Kenya; Somalia</td>
</tr>
<tr>
<td>Kinyarwanda</td>
<td>Atlantic-Congo</td>
<td>Africa</td>
<td>Burundi; Congo, The Democratic Republic of the; Rwanda, Tanzania, United Republic of; Uganda</td>
</tr>
<tr>
<td>Konni</td>
<td>Atlantic-Congo</td>
<td>Africa</td>
<td>Ghana</td>
</tr>
<tr>
<td>Yoruba</td>
<td>Atlantic-Congo</td>
<td>Africa</td>
<td>Benin; Nigeria</td>
</tr>
<tr>
<td>Plateau Malagasy</td>
<td>Austronesian</td>
<td>Africa</td>
<td>Madagascar</td>
</tr>
<tr>
<td>Northern Gumuz</td>
<td>Gumuz</td>
<td>Africa</td>
<td>Ethiopia; Sudan</td>
</tr>
<tr>
<td>//Ani</td>
<td>Khoe-Kwadi</td>
<td>Africa</td>
<td>Botswana</td>
</tr>
<tr>
<td>Kxoe</td>
<td>Khoe-Kwadi</td>
<td>Africa</td>
<td>Angola; Botswana; Namibia; Zambia</td>
</tr>
<tr>
<td>Mandinka</td>
<td>Mande</td>
<td>Africa</td>
<td>Gambia; Guinea; Guinea-Bissau; Senegal</td>
</tr>
<tr>
<td>Anindilyakwa</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Bininj Kun-Wok</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Ngalakan</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Ngalkbun</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Ngandi</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Rembarrnga</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
</tbody>
</table>

(continued)
(continued)

<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warray</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Wubuy</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Gurinji</td>
<td>Pama-Nyungan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Guugu Yimidhirr</td>
<td>Pama-Nyungan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Pintupi-Luritja</td>
<td>Pama-Nyungan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Pitta Pitta</td>
<td>Pama-Nyungan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Walmajarri</td>
<td>Pama-Nyungan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Yir-Yoront</td>
<td>Pama-Nyungan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Murrinh-Patha</td>
<td>Southern Daly</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Nangikurrunggurr</td>
<td>Southern Daly</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Tiwi</td>
<td>Tiwi (Isolate)</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Marithiel</td>
<td>Western Daly</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Ubykh</td>
<td>Abkhaz-Adyge</td>
<td>Eurasia</td>
<td>Turkey</td>
</tr>
<tr>
<td>Modern Hebrew</td>
<td>Afro-Asiatic</td>
<td>Eurasia</td>
<td>Israel; Jordan; Lebanon;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Palestine, State of; Syrian</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Arab Republic</td>
</tr>
<tr>
<td>Hokkaido Ainu</td>
<td>Ainu</td>
<td>Eurasia</td>
<td>Japan</td>
</tr>
<tr>
<td>Bahnar</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>Viet Nam</td>
</tr>
<tr>
<td>Bodo Gadaba</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Bondo</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Bugan</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>China</td>
</tr>
<tr>
<td>Car Nicobarese</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Central Nicobarese</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Gata'</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Gorum-Parenga</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Ho</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Juang</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Juray</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Kharia</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India; Nepal</td>
</tr>
<tr>
<td>Khasi</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>Bangladesh; India</td>
</tr>
<tr>
<td>Mundari</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>Bangladesh; India; Nepal</td>
</tr>
<tr>
<td>Sora</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Temiar</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>Malaysia</td>
</tr>
<tr>
<td>Alutor</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Chukchi</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Itelmen</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Koryak</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Tamil</td>
<td>Dravidian</td>
<td>Eurasia</td>
<td>India; Sri Lanka</td>
</tr>
<tr>
<td>Telugu</td>
<td>Dravidian</td>
<td>Eurasia</td>
<td>India</td>
</tr>
<tr>
<td>Central Siberian</td>
<td>Eskimo-Aleut</td>
<td>Eurasia</td>
<td>Russian Federation; United</td>
</tr>
<tr>
<td>Yupik</td>
<td>Eskimo-Aleut</td>
<td>Eurasia</td>
<td>States</td>
</tr>
<tr>
<td>Kalaallisut</td>
<td>Eskimo-Aleut</td>
<td>Eurasia</td>
<td>Greenland</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Andorra; France; Italy; Spain</td>
</tr>
<tr>
<td>Danish</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Denmark; Germany</td>
</tr>
<tr>
<td>Dutch</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Belgium; Brazil; French Guiana; Germany; Guyana; Netherlands; Suriname</td>
</tr>
<tr>
<td>English</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Australia; Belize; Bermuda; Brazil; Canada; Cayman Islands; Cook Islands; Cuba; Dominican Republic; Falkland Islands (Malvinas); Gibraltar; Guadeloupe; Guatemala; Guyana; Honduras; Ireland; Liberia; Mexico; Norfolk Island; Pitcairn; Saint Helena, Ascension and Tristan da Cunha; South Africa; Suriname; United Kingdom; United States; Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>French</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Andorra; Belgium; Canada; France; French Guiana; Italy; Luxembourg; Monaco; Saint Pierre and Miquelon; Switzerland; United States Austria; Belgium; Czechia; Denmark; France; Germany; Hungary; Italy; Liechtenstein; Luxembourg; Netherlands; Poland; Romania; Slovakia; Slovenia; Switzerland</td>
</tr>
<tr>
<td>German</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Austria; Belgium; Czechia; Denmark; France; Germany; Hungary; Italy; Liechtenstein; Luxembourg; Netherlands; Poland; Romania; Slovakia; Slovenia; Switzerland</td>
</tr>
<tr>
<td>Hindi</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Bangladesh; Bhutan; India; Nepal</td>
</tr>
<tr>
<td>Lower Sorbian</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Germany</td>
</tr>
<tr>
<td>Modern Greek</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Albania; Bulgaria; Cyprus; Egypt; Greece; Italy; Macedonia, Republic of; Romania; Turkey; Ukraine Arabic; Azerbaijan; Iran; Islamic Republic of; Iraq; Jordan; Kuwait; Syrian Arab Republic; Turkey; Turkmenistan</td>
</tr>
<tr>
<td>Northern Kurdish</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Armenia; Azerbaijan; Iran; Islamic Republic of; Iraq; Jordan; Kuwait; Syrian Arab Republic; Turkey; Turkmenistan</td>
</tr>
<tr>
<td>Language</td>
<td>Language family</td>
<td>Macro-area</td>
<td>Country</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>------------</td>
<td>---------</td>
</tr>
<tr>
<td>Russian</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Belarus; China; Estonia; Finland; Georgia; Kazakhstan; Korea, Democratic People's Republic of; Kyrgyzstan; Latvia; Lithuania; Moldova, Republic of; Mongolia; Norway; Poland; Romania; Russian Federation; Turkmenistan; Ukraine; Uzbekistan</td>
</tr>
<tr>
<td>Spanish</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Andorra; Argentina; Belize; Bolivia, Plurinational State of; Brazil; Chile; Colombia; Costa Rica; Cuba; Dominican Republic; Ecuador; El Salvador; France; Gibraltar; Guatemala; Guyana; Haiti; Honduras; Mexico; Morocco; Nicaragua; Panama; Paraguay; Peru; Portugal; Puerto Rico; Spain; United States; Uruguay; Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>Upper Sorbian</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Germany</td>
</tr>
<tr>
<td>Western Farsi</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Afghanistan; Azerbaijan; Iran, Islamic Republic of; Iraq; Turkey; Turkmenistan</td>
</tr>
<tr>
<td>Western Frisian</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Japanese</td>
<td>Japonic</td>
<td>Eurasia</td>
<td>Japan</td>
</tr>
<tr>
<td>Korean</td>
<td>Koranic</td>
<td>Eurasia</td>
<td>China; Korea, Democratic People's Republic of; Korea, Republic of; Russian Federation</td>
</tr>
<tr>
<td>Nivkh</td>
<td>Nivkh (Isolate)</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Israeli Sign Language</td>
<td>Sign Language</td>
<td>Eurasia</td>
<td>Israel</td>
</tr>
<tr>
<td>Japanese Sign Language</td>
<td>Sign Language</td>
<td>Eurasia</td>
<td>Japan</td>
</tr>
<tr>
<td>Japhug</td>
<td>Sino-Tibetan</td>
<td>Eurasia</td>
<td>China</td>
</tr>
<tr>
<td>Lahu</td>
<td>Sino-Tibetan</td>
<td>Eurasia</td>
<td>China; Lao People’s Democratic Republic; Myanmar; Thailand; Viet Nam</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin Chinese</td>
<td>Sino-Tibetan</td>
<td>Eurasia</td>
<td>China; Korea, Democratic People's Republic of; Lao People's Democratic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Republic; Mongolia; Myanmar; Russian Federation; Taiwan, Province of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>China; Viet Nam</td>
</tr>
<tr>
<td>Tangut</td>
<td>Sino-Tibetan</td>
<td>Eurasia</td>
<td>China</td>
</tr>
<tr>
<td>Tibetan</td>
<td>Sino-Tibetan</td>
<td>Eurasia</td>
<td>Bhutan; China; India; Nepal</td>
</tr>
<tr>
<td>Udihe</td>
<td>Tungusic</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Turkish</td>
<td>Turkic</td>
<td>Eurasia</td>
<td>Albania; Armenia; Azerbaijan; Bulgaria; Cyprus; Georgia; Greece; Iran,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Islamic Republic of; Iraq; Macedonia; Republic of; Syrian Arab</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Republic; Turkey</td>
</tr>
<tr>
<td>Yakut</td>
<td>Turkic</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Estonian</td>
<td>Uralic</td>
<td>Eurasia</td>
<td>Estonia; Latvia; Russian Federation</td>
</tr>
<tr>
<td>Hungarian</td>
<td>Uralic</td>
<td>Eurasia</td>
<td>Austria; Croatia; Hungary; Romania; Serbia; Slovakia; Slovenia; Ukraine</td>
</tr>
<tr>
<td>Ket</td>
<td>Yeniseian</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Kott-Assan</td>
<td>Yeniseian</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Yugh</td>
<td>Yeniseian</td>
<td>Eurasia</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>Eastern Ojibwa</td>
<td>Algic</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Meskwaki</td>
<td>Algic</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Montagnais</td>
<td>Algic</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Plains Cree</td>
<td>Algic</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Shawnee</td>
<td>Algic</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Sikiska</td>
<td>Algic</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Ahtena</td>
<td>Athapaskan-Eyak-Tlingit</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Carrier</td>
<td>Athapaskan-Eyak-Tlingit</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Chipewyan</td>
<td>Athapaskan-Eyak-Tlingit</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Koyukon</td>
<td>Athapaskan-Eyak-Tlingit</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Navajo</td>
<td>Athapaskan-Eyak-Tlingit</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Sarsi</td>
<td>Athapaskan-Eyak-Tlingit</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>South Slavey</td>
<td>Athapaskan-Eyak-Tlingit</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Tanaina</td>
<td>Athapaskan-Eyak-Tlingit</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Tlingit</td>
<td>Athapaskan-Eyak-Tlingit</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Arikara</td>
<td>Caddoan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Caddo</td>
<td>Caddoan</td>
<td>North America</td>
<td>United States</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pawnee</td>
<td>Caddoan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Wichita</td>
<td>Caddoan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Bribri</td>
<td>Chibchan</td>
<td>North America</td>
<td>Costa Rica; Panama</td>
</tr>
<tr>
<td>Quileute</td>
<td>Chimakuan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Quechuan</td>
<td>Cochimi-Yuman</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Aleut</td>
<td>Eskimo-Aleut</td>
<td>North America</td>
<td>Russian Federation; United States</td>
</tr>
<tr>
<td>Central Yupik</td>
<td>Eskimo-Aleut</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Eastern Canadian</td>
<td>Eskimo-Aleut</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Inuktitut</td>
<td>Eskimo-Aleut</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Haida*</td>
<td>Haida</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Cayuga</td>
<td>Iroquoian</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Cherokee</td>
<td>Iroquoian</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Mohawk</td>
<td>Iroquoian</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Nottoway</td>
<td>Iroquoian</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Oneida</td>
<td>Iroquoian</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Onondaga</td>
<td>Iroquoian</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Seneca</td>
<td>Iroquoian</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Tuscarora</td>
<td>Iroquoian</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Wyandot</td>
<td>Iroquoian</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Kiowa</td>
<td>Kiowa-Tanoan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Southern Tiwa</td>
<td>Kiowa-Tanoan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Towa</td>
<td>Kiowa-Tanoan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Nisenan</td>
<td>Maiduan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Northeast Maidu</td>
<td>Maiduan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Northwest Maidu</td>
<td>Maiduan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Akateko</td>
<td>Mayan</td>
<td>North America</td>
<td>Guatemala; Mexico</td>
</tr>
<tr>
<td>Chuj</td>
<td>Mayan</td>
<td>North America</td>
<td>Guatemala; Mexico</td>
</tr>
<tr>
<td>Mam</td>
<td>Mayan</td>
<td>North America</td>
<td>Guatemala; Mexico</td>
</tr>
<tr>
<td>Tzotzil</td>
<td>Mayan</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Yucatec Maya</td>
<td>Mayan</td>
<td>North America</td>
<td>Belize; Guatemala; Mexico</td>
</tr>
<tr>
<td>Miskito</td>
<td>Misumalpan</td>
<td>North America</td>
<td>Honduras; Nicaragua</td>
</tr>
<tr>
<td>Chimalapa Zoque</td>
<td>Mixe-Zoque</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Highland Popoluca</td>
<td>Mixe-Zoque</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Oluta Popoluca</td>
<td>Mixe-Zoque</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Choctaw</td>
<td>Muskogean</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Creek</td>
<td>Muskogean</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Koasati</td>
<td>Muskogean</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Natchez</td>
<td>Natchez (Isolate)</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Acatepec Me'phaa</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Chichahuaxta Triqui</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Copala Triqui</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico; United States</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Jerónimo</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Tecóatl Maztec</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Western Highland Chatino</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Achumawi</td>
<td>Palaihnihan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Central Pomo</td>
<td>Pomoan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Clallam</td>
<td>Salishan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Coeur d’Alene</td>
<td>Salishan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Comox</td>
<td>Salishan</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Halkomelem</td>
<td>Salishan</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Lillooet</td>
<td>Salishan</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Shasta</td>
<td>Shastan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>American Sign Language</td>
<td>Sign Language</td>
<td>North America</td>
<td>Barbados; Benin; Bolivia, Plurinational State of; Burkina Faso; Canada; Central African Republic; Chad; China; Congo, The Democratic Republic of the; Côte d’Ivoire; Gabon; Ghana; Jamaica; Kenya; Madagascar; Mauritania; Nigeria; Philippines; Singapore; Togo; United States; Zimbabwe</td>
</tr>
<tr>
<td>Assiniboine</td>
<td>Siouan</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Biloxi</td>
<td>Siouan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Crow</td>
<td>Siouan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Hidatsa</td>
<td>Siouan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Ho-Chunk</td>
<td>Siouan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Lakota</td>
<td>Siouan</td>
<td>North America</td>
<td>United States; Canada; United States</td>
</tr>
<tr>
<td>Mandan</td>
<td>Siouan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Takelma</td>
<td>Takelma (Isolate)</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Purepecha</td>
<td>Tarascan</td>
<td>North America</td>
<td>Mexico; United States</td>
</tr>
<tr>
<td>Tonkawa</td>
<td>Tonkawa (Isolate)</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Upper Necaxa</td>
<td>Totonac</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Totonac</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gitxsan</td>
<td>Tsimshian</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Nisga’a</td>
<td>Tsimshian</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Southern-Coastal Tsimshian</td>
<td>Tsimshian</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Tarahumaran*</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Classical Nahuatl</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Comanche</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>United States</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Huasteca Nahuatl</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Hopi</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Huichol</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Panamint</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Pima Bajo</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Pipil</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>El Salvador</td>
</tr>
<tr>
<td>Shoshoni</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Tlaxcala-Puebla-Central</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Ute-Southern Paiute</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Yaqui</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>United States; Mexico</td>
</tr>
<tr>
<td>Kwak’wala</td>
<td>Wakashan</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Nuu-chah-nulth</td>
<td>Wakashan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Washo</td>
<td>Washo (Isolate)</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Yana</td>
<td>Yana (Isolate)</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Zuni</td>
<td>Zuni (Isolate)</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Aneltyum</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Vanuatu</td>
</tr>
<tr>
<td>Chamorro</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Guam; Northern Mariana Islands</td>
</tr>
<tr>
<td>Chuukese</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Micronesia, Federated States of</td>
</tr>
<tr>
<td>Dehu</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>New Caledonia</td>
</tr>
<tr>
<td>East Futuna</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Wallis and Futuna</td>
</tr>
<tr>
<td>Fijian</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Fiji</td>
</tr>
<tr>
<td>Iaai</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>New Caledonia</td>
</tr>
<tr>
<td>Indonesian</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Indonesia; Netherlands; Philippines; Saudi Arabia; Singapore; United States</td>
</tr>
<tr>
<td>Kambera</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Indonesia</td>
</tr>
<tr>
<td>Kosraean</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Micronesia, Federated States of</td>
</tr>
<tr>
<td>Kumak</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>New Caledonia</td>
</tr>
<tr>
<td>Manam</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td>Maori</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>New Zealand</td>
</tr>
<tr>
<td>Marshallese</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Marshall Islands</td>
</tr>
<tr>
<td>Mokilese</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Micronesia, Federated States of</td>
</tr>
<tr>
<td>Niuean</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Niue</td>
</tr>
<tr>
<td>Paiwan</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Taiwan, Province of China</td>
</tr>
<tr>
<td>Pampanga</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Philippines</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pohnpeian</td>
<td>Austronesian</td>
<td>Micronesia</td>
<td>Federated States of</td>
</tr>
<tr>
<td>Sakao</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Vanuatu</td>
</tr>
<tr>
<td>Saliba</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td>Samoan</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>American Samoa; Samoa</td>
</tr>
<tr>
<td>To’abaita</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Solomon Islands</td>
</tr>
<tr>
<td>Tonga (Tonga Islands)</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Tonga</td>
</tr>
<tr>
<td>Tukang Besi*</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Indonesia</td>
</tr>
<tr>
<td>Warembori</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Indonesia</td>
</tr>
<tr>
<td>Yapese</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Micronesia; Federated States of</td>
</tr>
<tr>
<td>Yimas</td>
<td>Lower Sepik-Ramu</td>
<td>Papunesia</td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td>Amele</td>
<td>Nuclear Trans New Guinea</td>
<td>Papunesia</td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td>Alamblik</td>
<td>Sepik</td>
<td>Papunesia</td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td>Saweru</td>
<td>Yawa</td>
<td>Papunesia</td>
<td>Indonesia</td>
</tr>
<tr>
<td>Yele</td>
<td>Yele (Isolate)</td>
<td>Papunesia</td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td>Mapudungun</td>
<td>Araucanian</td>
<td>South America</td>
<td>Argentina; Chile</td>
</tr>
<tr>
<td>Palikúr</td>
<td>Arawakan</td>
<td>South America</td>
<td>Brazil; French Guiana; Guyana; Suriname; Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>Wayuu</td>
<td>Arawakan</td>
<td>South America</td>
<td>Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>Paumarí</td>
<td>Arawan</td>
<td>South America</td>
<td>Brazil; French Guiana; Guyana; Suriname; Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>Galibi Carib</td>
<td>Cariban</td>
<td>South America</td>
<td>Brazil; French Guiana; Guyana; Suriname; Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>Panare</td>
<td>Cariban</td>
<td>South America</td>
<td>Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>Wari’</td>
<td>Chapacuran</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Maloyo</td>
<td>Chibchan</td>
<td>South America</td>
<td>Colombia</td>
</tr>
<tr>
<td>Guahibo</td>
<td>Guahibo</td>
<td>South America</td>
<td>Colombia; Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>Muruí Huitoto</td>
<td>Huitotoan</td>
<td>South America</td>
<td>Colombia; Peru</td>
</tr>
<tr>
<td>Katukina-Kanamarí</td>
<td>Katukinan</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Wichí*</td>
<td>Matacoan</td>
<td>South America</td>
<td>Argentina; Bolivia, Plurinational State of; Paraguay</td>
</tr>
<tr>
<td>Movima</td>
<td>Movima (Isolate)</td>
<td>South America</td>
<td>Bolivia, Plurinational State of</td>
</tr>
<tr>
<td>Dâw</td>
<td>Nadahup</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Nadēb</td>
<td>Nadahup</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Mamaindé</td>
<td>Nambiquaran</td>
<td>South America</td>
<td>Brazil</td>
</tr>
</tbody>
</table>
## Appendix 2: Proportions of language families

Proportion of language families from each macro-area in the list of incorporating languages and corresponding number of languages from each macro-area in the sample. Note that the Afro-Asiatic, Austronesian, Chibchan, Eskimo-Aleut and Sign Language families are counted twice in this table because the languages from these families included in the list are spread over two different macro-areas.

<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern Nambikuára</td>
<td>Nambiquaran</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Karajá</td>
<td>Nuclear-Macro-Je</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Panará</td>
<td>Nuclear-Macro-Je</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Ese Ejja</td>
<td>Pano-Tacanan</td>
<td>South America</td>
<td>Bolivia, Plurinational State of; Peru</td>
</tr>
<tr>
<td>Tanimuca-Retuarã</td>
<td>Tucanoan</td>
<td>South America</td>
<td>Colombia</td>
</tr>
<tr>
<td>Gavião do Jiparaná</td>
<td>Tupian</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Mundurukú</td>
<td>Tupian</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Paraguayan Guaraní</td>
<td>Tupian</td>
<td>South America</td>
<td>Argentina; Paraguay</td>
</tr>
<tr>
<td>Tapieté</td>
<td>Tupian</td>
<td>South America</td>
<td>Bolivia, Plurinational State of; Paraguay</td>
</tr>
<tr>
<td>Tupinambá</td>
<td>Tupian</td>
<td>South America</td>
<td>Brazil</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Macro-area</th>
<th>Number of families (including isolates)</th>
<th>Proportion of families</th>
<th>Corresponding number of sample languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>6</td>
<td>0.071428571</td>
<td>2.142857143</td>
</tr>
<tr>
<td>Australia</td>
<td>5</td>
<td>0.05952381</td>
<td>1.785714286</td>
</tr>
<tr>
<td>Eurasia</td>
<td>17</td>
<td>0.202380952</td>
<td>6.071428571</td>
</tr>
<tr>
<td>North America</td>
<td>33</td>
<td>0.392857143</td>
<td>11.78571429</td>
</tr>
<tr>
<td>Papunesia</td>
<td>6</td>
<td>0.071428571</td>
<td>2.142857143</td>
</tr>
<tr>
<td>South America</td>
<td>17</td>
<td>0.202380952</td>
<td>6.071428571</td>
</tr>
</tbody>
</table>

| Total          | 84                                      | 1                      | 30 | 30 |

(continued)
Appendix 3: Consulted sources for the sample languages

Bininj Kun-Wok


Chimalapa Zoque

Chukchi


Crow


*Eastern Ojibwa*


**Halkomelem**


Hokkaido Ainu


Iraqw


Kalaallisut
Ket


**Mapudungun**


**Marithiel**


**Mohawk**


**Movima**


**Nadëb**


**Niuean**


*Northern Gumuz*


*Nuu-chah-nulth*


**Palikúr**


**Panare**


**Paraguayan Guaraní**


**Sora**


**South Slavey**


*Southern Tiwa*


approaches to empirical questions (Lecture Notes), 75–95. Stanford, CA: Center for the Study of Language and Information Publications.


Ute-Southern Paiute


Washo


**Western Frisian**


**Western Highland Chatino**


**Yimas**


Foley, William A. 1997. Polysynthesis and complex verb formation: The case of applicatives in Yimas. In Alex Alsina, Joan Bresnan & Peter Sells (eds.), *Complex predicates* (Center for the Study of Language and Information


Yucatec Maya


References


