Incorporation:
Constraints on variation
Cover illustration: Marieke Olthof. The cover of this thesis shows eight of the reference grammars that have been used as data sources for this research.

ISBN: 978-94-6093-360-8
NUR: 616

Copyright © 2020: Marieke Olthof. All rights reserved.
Incorporation:
Constraints on variation

ACADEMISCH PROEFSCHRIFT
ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op vrijdag 4 september 2020, te 10.00 uur
door Marieke Diana Olthof
geboren te Zaanstad
Promotiecommissie:

Promotoren:  dr. J. Don Universiteit van Amsterdam
              prof. dr. P.C. Hengeveld Universiteit van Amsterdam

Overige leden:  dr. S.P. Aalberse Universiteit van Amsterdam
                prof. dr. E.O. Aboh Universiteit van Amsterdam
                prof. dr. M.B.H. Everaert Universiteit Utrecht
                dr. J. Mattissen Universität Köln
                prof. dr. A.P. Versloot Universiteit van Amsterdam

Faculteit der Geesteswetenschappen

This work is part of the research program PhDs in the Humanities with project number 322–70–011, which is financed by the Dutch Research Council (NWO).
# Table of contents

Acknowledgments .................................................................................... ix
Author contributions ................................................................................... xi
List of abbreviations ................................................................................... xiii
List of online materials ................................................................................. xv

## 1 Introduction

1.1 Background ......................................................................................... 1
1.2 Research aim and research topics ....................................................... 5
1.3 Approach and data ............................................................................. 7

## 2 Noun incorporation in Functional Discourse Grammar

2.1 Introduction ......................................................................................... 13
2.2 Functional Discourse Grammar .......................................................... 15
2.3 Defining noun incorporation in FDG ................................................... 16
2.4 The IL-ML interface ........................................................................... 19
   2.4.1 The interpersonal category of the incorporated noun .................. 20
   2.4.2 The head of the incorporated noun ........................................... 22
   2.4.3 The pragmatic operators applying to the incorporated noun ....... 23
   2.4.4 The pragmatic function of the incorporated noun ..................... 24
2.5 The RL-ML interface ........................................................................... 27
   2.5.1 The semantic layer of the incorporated noun ............................... 27
   2.5.2 The semantic function of the incorporated noun ....................... 30
   2.5.3 Type of dependent with respect to the incorporating verb ............ 34
   2.5.4 Morphosyntactic alignment ....................................................... 38
   2.5.5 Relationality ............................................................................. 41
2.6 The ML-PL interface ........................................................................... 42
   2.6.1 The head of the incorporated noun ........................................... 42
   2.6.2 The phonological layer of the incorporated noun ....................... 44
2.7 A worked example ............................................................................. 45
   2.7.1 Introduction ............................................................................. 45
   2.7.2 The IL-ML interface ................................................................ 47
   2.7.3 The RL-ML interface ................................................................ 48
   2.7.4 The ML-PL interface ................................................................ 50
   2.7.5 Mappings ................................................................................ 51
2.8 Discussion and conclusions ................................................................. 51
3 Formal variation in incorporation: A typological study and a unified approach

3.1 Introduction

3.2 Theoretical background

3.2.1 Previous accounts of incorporation

3.2.2 Incorporation in Functional Discourse Grammar

3.2.3 Implications of the FDG definition of incorporation

3.3 Hypotheses

3.3.1 The possible forms of incorporated elements

3.3.2 The cross-linguistic distribution of the various forms of incorporated elements

3.4 Method

3.4.1 Sampling procedure

3.4.2 Data analysis

3.5 Results

3.5.1 The possible forms of incorporated elements

3.5.2 The cross-linguistic distribution of the various forms of incorporated elements

3.6 Conclusion

4 Referentiality and modifiability of incorporated nouns: Cross- and intra-linguistic variation

4.1 Introduction

4.2 Referentiality and modifiability in FDG

4.2.1 Referentiality

4.2.2 Modifiability

4.2.3 Pragmatic-semantic typology of common nouns in FDG

4.3 Method

4.3.1 Sampling procedure

4.3.2 Definition of noun incorporation

4.3.3 Data analysis

4.4 Results

4.5 Discussion

4.6 Conclusion

5 Verb-based restrictions on noun incorporation across languages

5.1 Introduction

5.2 Definition, theoretical background and research questions

5.2.1 Defining noun incorporation

5.2.2 Theoretical background

5.2.2.1 Noun incorporation and semantic transitivity
5.2.2.2 Noun incorporation and morphosyntactic transitivity 133
5.2.3 Research questions 136
5.3 Typological survey 137
5.3.1 Method and data 137
5.3.2 Results and discussion 139
5.3.2.1 Frequently incorporating verbs 139
5.3.2.2 Morphosyntactic transitivity 140
5.3.2.3 Other factors 141
5.3.2.4 Variation across languages 142
5.4 Questionnaire-based case studies 143
5.4.1 Method and data 143
5.4.1.1 Questionnaire design 143
5.4.1.2 Language sample and data 144
5.4.1.3 Method 146
5.4.2 Results and discussion 147
5.4.2.1 Frequently incorporating verbs 147
5.4.2.2 Morphosyntactic transitivity 150
5.4.2.3 Other factors 153
5.4.2.4 Variation across languages 156
5.5 Discussion and conclusion 157

6 Incorporation and linguistic theory
6.1 Theoretical approaches to incorporation 161
6.2 Basic settings and implicational hierarchies 162
6.3 Morphosyntactic forms of incorporated elements 163
6.4 Pragmatic referentiality and semantic modifiability 166
6.5 Verb-based lexical restrictions 171
6.6 Conclusion 172

References 175

Appendix 1 List of incorporating languages 191
Appendix 2 Proportions of language families 199
Appendix 3 Sources consulted for the sample languages in Chapter 3 201
Appendix 4 Sources consulted for the sample languages in Chapter 4 219
Appendix 5 Language sample and sources consulted for the typological
survey in Chapter 5 227

Summary 235
Samenvatting (summary in Dutch) 239
Curriculum Vitae 243
Acknowledgments

A lot of people have contributed to this thesis in many different ways. I am very thankful both for the help that has been essential for the content of the thesis and for the support that has made the four years of research a very positive experience.

First of all, I would like to thank my two supervisors, Kees Hengeveld and Jan Don, for all their advice, their time and the smooth cooperation. Kees, thank you for suggesting that we could write a proposal for the PhDs in the Humanities in the first place. I am also very grateful for your positivity, the insightful discussions and your generous help during the whole project. Jan, your non-FDG theoretical perspective has been extremely useful. Thank you for teaching me so much and for all your careful reading and detailed and constructive comments.

I also wish to thank my co-authors with whom I wrote two of the chapters of this thesis: Kees Hengeveld, for Chapter 2, and Eva van Lier, Tjeu Claessen, Swintha Danielsen, Katharina Haude, Nico Lehmann, Maarten Mous, Elisabeth Verhoeven, Eline Visser, Marine Vuillermet and Arok Wolvengrey, for Chapter 5. Working together in this way was a very valuable experience. A special thanks goes to Eva van Lier. Eva, I really enjoyed all the times we sat together behind one of our desks to work on our project, and I learned a lot from you.

I would like to express my gratitude to all the researchers who helped me analyze the data of their languages of expertise: Colleen Ahland, Alexandra Aikhenvald, Gregory Anderson, Anna Bugaeva, Michael Dunn, Nicholas Evans, William Foley, Michael Fortescue, Donald Frantz, Stefan Georg, Talmy Givón, Katharina Haude, Christian Lehmann, Diane Massam, Marianne Mithun, Thomas Payne, Keren Rice, Tomomi Sato and Fernando Zúñiga.

I am indebted to the participants of the seventh International Workshop on Functional Discourse Grammar for their feedback on Chapter 2, to Frank Seifart and Ulrika Klomp for their suggestions for Chapter 3 and to Sune Gregersen and Søren Wichmann for their comments on Chapter 5. I am also thankful for the feedback that I received from the anonymous reviewers of Chapter 3, 4 and 5. Thanks to Tijmen Klous for proofreading Chapter 1 and 6 and the summary.

In addition, I would like to express my appreciation to Suzanne Aalberse, Enoch Aboh, Martin Everaert, Johanna Mattissen and Arjen Versloot for agreeing to be part of my doctorate committee.

I gratefully acknowledge the financing of the project that resulted in this thesis (project number 322-70-011, research program PhDs in the Humanities) by the Dutch Research Council (NWO).

I am also very thankful to all my (former) colleagues at the P.C. Hooftius for creating such a pleasant atmosphere and fruitful research environment. A big thanks
to Cindy, Dunja, Imme, Iris, Kiki, Klaas, Marloes, Merel, Natalia, Rosa, Sune, Sybren, Tessa, Thom, Tiffany and Ulrika for all the chats and for the lunches on the fifth floor. I am really happy that I got to spend my time as a PhD researcher with you.

Thanks to my office mates Fabienne, Jakko, Nadia, Sune and Thom for making our attic room a nice place to work. Thanks to Andreas, Dunja, Hongmei, Jie, Kiki, Marloes and Ulrika for the open and encouraging talks about our experiences as PhD researchers in our peer coaching group. Thanks to the Language description & typology research group for the interesting and helpful meetings. Of course a special thanks goes to my paranymphs Sune and Ulrika.

I want also to thank my friends and family who were just outside the university for me during the past four years. Thanks Emma, Katja, Sabine, Sofie and Tereza for all the IKC-time. Emma and Tereza, as I think about the past four years I think of all the evenings together at home in Amsterdam. Cas, Christianne and Vicky, we met even together to discuss language and culture at the UvA but I know we will be friends for sure. Marieke, thanks for all the friendship and all the conversations over our experiences as PhD’s. To my whole family Olthof, Kooke, Van Bemmelen and everyone who is part of: thank you so much for all your support and all the happiness during the past four years.

Daniëlle, Lennard, papa and mama, really thank you for always being there. Marieke, is grassmaaien also incorporated?), for all the help when I was so much worrying, for being there and for being such a wonderful person. I know you will always be there.

Lieve Koen, you can never thank you enough. I was always there for you when I was sure I would never get the NWO-grant, for being there with me when my first article was accepted, for coming with me to the conference and the summer school and for making sure that I was happy. But you were already there before I even knew what a PhD was. And now I especially look forward to all our time together that lies ahead.
Author contributions

1 Introduction
This chapter has not been submitted for publication elsewhere.

Marieke Olthof is the sole author of this chapter. She revised the chapter on the basis of valuable feedback provided by Jan Don and Kees Hengeveld.

2 Noun incorporation in Functional Discourse Grammar
This chapter is a slightly adapted version of: Olthof, Marieke & Kees Hengeveld. subm. Noun incorporation in Functional Discourse Grammar.

Marieke Olthof and Kees Hengeveld developed the study and the outline of this paper together. Marieke Olthof wrote most of Section 2.3–2.6, while Section 2.1–2.2 and 2.7–2.8 were written primarily by Kees Hengeveld. After discussing the sections together, both authors made revisions to the paper. They further revised the paper on the basis of valuable feedback provided by Jan Don and the participants of the seventh International Workshop on Functional Discourse Grammar, Oviedo, September 12–13, 2019.

3 Formal variation in incorporation: A typological study and a unified approach
This chapter is a slightly adapted version of: Olthof, Marieke. 2020. Formal variation in incorporation: A typological study and a unified approach. Linguistics 58(1). 131–205.

Marieke Olthof is the sole author of this paper. She revised the paper on the basis of valuable feedback provided by Jan Don, Kees Hengeveld, Frank Seifart and two anonymous reviewers and with the help of useful comments made by Ulrika Klomp.

4 Referentiality and modifiability of incorporated nouns: Cross- and intra-linguistic variation
This chapter is a slightly adapted version of: Olthof, Marieke. forthc. Referentiality and modifiability of incorporated nouns: Cross- and intra-linguistic variation. Sprachtypologie und Universalienforschung.

Marieke Olthof is the sole author of this paper. She revised the paper on the basis of valuable feedback provided by Jan Don, Kees Hengeveld and four anonymous reviewers.
Verb-based restrictions on noun incorporation across languages

This chapter is a slightly adapted version of: Olthof, Marieke, Eva van Lier, Tjeu Claessen, Swintha Danielsen, Katharina Haude, Nico Lehmann, Maarten Mous, Elisabeth Verhoeven, Eline Visser, Marine Vuillermet & Arok Wolvengrey. forthc. Verb-based restrictions on noun incorporation across languages. Linguistic Typology.

Marieke Olthof and Eva van Lier developed the study, collected the data for the typological survey, analyzed the data for the typological survey and designed the questionnaire for the questionnaire-based case studies. Tjeu Claessen, Swintha Danielsen, Katharina Haude, Nico Lehmann, Maarten Mous, Elisabeth Verhoeven, Eline Visser, Marine Vuillermet and Arok Wolvengrey then provided the data for these questionnaire-based case studies as well as general information on incorporation in the languages of their expertise. Subsequently, Marieke Olthof and Eva van Lier analyzed the data for the questionnaire-based case studies and wrote the first version of the paper. Tjeu Claessen, Swintha Danielsen, Katharina Haude, Nico Lehmann, Maarten Mous, Elisabeth Verhoeven, Eline Visser, Marine Vuillermet and Arok Wolvengrey commented on this version. Marieke Olthof and Eva van Lier revised the paper according to their comments and made further revisions on the basis of valuable feedback provided by Jan Don, Sune Gregersen, Kees Hengeveld, Søren Wichmann and three anonymous reviewers.

Incorporation and linguistic theory

This chapter has not been submitted for publication elsewhere.

Marieke Olthof is the sole author of this chapter. She revised the chapter on the basis of valuable feedback provided by Jan Don and Kees Hengeveld.
## List of abbreviations

<table>
<thead>
<tr>
<th></th>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>first person</td>
<td>DET</td>
</tr>
<tr>
<td>2</td>
<td>second person</td>
<td>DINC</td>
</tr>
<tr>
<td>3</td>
<td>third person</td>
<td>DISC</td>
</tr>
<tr>
<td>I</td>
<td>gender class I</td>
<td>DISTR</td>
</tr>
<tr>
<td>II</td>
<td>noun class II</td>
<td>DOBJ</td>
</tr>
<tr>
<td>IV</td>
<td>noun class IV</td>
<td>DR</td>
</tr>
<tr>
<td>VII</td>
<td>noun class VII</td>
<td>DS</td>
</tr>
<tr>
<td>A</td>
<td>agent-like argument of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>canonical transitive verb</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>absential</td>
<td></td>
</tr>
<tr>
<td>ABL</td>
<td>ablative</td>
<td></td>
</tr>
<tr>
<td>ABS</td>
<td>absolutive</td>
<td></td>
</tr>
<tr>
<td>ABSS</td>
<td>absolute state</td>
<td></td>
</tr>
<tr>
<td>ACSBJ</td>
<td>subject of active verb</td>
<td></td>
</tr>
<tr>
<td>ADVZ</td>
<td>adverbializer</td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>allative</td>
<td></td>
</tr>
<tr>
<td>AN</td>
<td>animate</td>
<td></td>
</tr>
<tr>
<td>APPL</td>
<td>applicative</td>
<td></td>
</tr>
<tr>
<td>APPROX</td>
<td>approximative</td>
<td></td>
</tr>
<tr>
<td>ART</td>
<td>article</td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>aspect</td>
<td></td>
</tr>
<tr>
<td>AU</td>
<td>augmented</td>
<td></td>
</tr>
<tr>
<td>AUG</td>
<td>augmentative</td>
<td></td>
</tr>
<tr>
<td>AUX</td>
<td>auxiliary</td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>bound element</td>
<td></td>
</tr>
<tr>
<td>BR</td>
<td>bound root</td>
<td></td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
<td></td>
</tr>
<tr>
<td>COLL</td>
<td>collective</td>
<td></td>
</tr>
<tr>
<td>COMPL</td>
<td>completive</td>
<td></td>
</tr>
<tr>
<td>COND</td>
<td>conditional</td>
<td></td>
</tr>
<tr>
<td>CONT</td>
<td>continuative</td>
<td></td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
<td></td>
</tr>
<tr>
<td>DDEIX</td>
<td>distal deixis</td>
<td></td>
</tr>
<tr>
<td>DECL.</td>
<td>declarative</td>
<td></td>
</tr>
<tr>
<td>DEF</td>
<td>definite</td>
<td></td>
</tr>
<tr>
<td>DEM</td>
<td>demonstrative</td>
<td></td>
</tr>
<tr>
<td>DES</td>
<td>desiderative</td>
<td></td>
</tr>
<tr>
<td>DET</td>
<td>deteminer</td>
<td></td>
</tr>
<tr>
<td>DINC</td>
<td>dependent incomplete</td>
<td></td>
</tr>
<tr>
<td>DISC</td>
<td>discontinuative</td>
<td></td>
</tr>
<tr>
<td>DISTR</td>
<td>distributive</td>
<td></td>
</tr>
<tr>
<td>DOBJ</td>
<td>direct object</td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>bivalent direct</td>
<td></td>
</tr>
<tr>
<td>DS</td>
<td>different-subject marker</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>epenthesis</td>
<td></td>
</tr>
<tr>
<td>EMPH</td>
<td>emphatic</td>
<td></td>
</tr>
<tr>
<td>ERG</td>
<td>ergative</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>feminine</td>
<td></td>
</tr>
<tr>
<td>FUT</td>
<td>future</td>
<td></td>
</tr>
<tr>
<td>GOAL</td>
<td>goal postposition</td>
<td></td>
</tr>
<tr>
<td>HAB</td>
<td>habitual</td>
<td></td>
</tr>
<tr>
<td>IMP</td>
<td>imperative</td>
<td></td>
</tr>
<tr>
<td>INACSBJ</td>
<td>subject of inactive verb</td>
<td></td>
</tr>
<tr>
<td>INCOMPL</td>
<td>incompletive</td>
<td></td>
</tr>
<tr>
<td>IND</td>
<td>indicative</td>
<td></td>
</tr>
<tr>
<td>INDF</td>
<td>indefinite</td>
<td></td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
<td></td>
</tr>
<tr>
<td>INGR</td>
<td>ingressive</td>
<td></td>
</tr>
<tr>
<td>INS</td>
<td>instrumental</td>
<td></td>
</tr>
<tr>
<td>INT</td>
<td>intentional</td>
<td></td>
</tr>
<tr>
<td>INTERR</td>
<td>interrogative mood</td>
<td></td>
</tr>
<tr>
<td>INTR</td>
<td>intransitive</td>
<td></td>
</tr>
<tr>
<td>INTS</td>
<td>intensifier/intensive</td>
<td></td>
</tr>
<tr>
<td>INV</td>
<td>invisible/distal</td>
<td></td>
</tr>
<tr>
<td>INTS</td>
<td>intensifier/intensive</td>
<td></td>
</tr>
<tr>
<td>INV</td>
<td>invisible/distal</td>
<td></td>
</tr>
<tr>
<td>IPV</td>
<td>imperfective</td>
<td></td>
</tr>
<tr>
<td>IRR</td>
<td>irrealis</td>
<td></td>
</tr>
<tr>
<td>IVF</td>
<td>incorporation verb form</td>
<td></td>
</tr>
<tr>
<td>LK</td>
<td>linker</td>
<td></td>
</tr>
<tr>
<td>LNDEIX</td>
<td>location/negation deixis</td>
<td></td>
</tr>
<tr>
<td>LV</td>
<td>linking vowel</td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Meaning</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>masculine</td>
<td></td>
</tr>
<tr>
<td>MIR</td>
<td>mirative</td>
<td></td>
</tr>
<tr>
<td>MULT</td>
<td>multiple</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>neuter</td>
<td></td>
</tr>
<tr>
<td>NEARFUT</td>
<td>near future</td>
<td></td>
</tr>
<tr>
<td>NEG</td>
<td>negation/negative</td>
<td></td>
</tr>
<tr>
<td>NF</td>
<td>nonfeminine</td>
<td></td>
</tr>
<tr>
<td>NMLZ</td>
<td>nominalizer/ nominalization</td>
<td></td>
</tr>
<tr>
<td>NPST</td>
<td>nonpast</td>
<td></td>
</tr>
<tr>
<td>NSFX</td>
<td>noun suffix</td>
<td></td>
</tr>
<tr>
<td>NSPEC</td>
<td>nonspecific aspect</td>
<td></td>
</tr>
<tr>
<td>OBJ</td>
<td>object</td>
<td></td>
</tr>
<tr>
<td>OBL</td>
<td>oblique</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>patient-like argument of canonical transitive verb</td>
<td></td>
</tr>
<tr>
<td>PART</td>
<td>participial mood</td>
<td></td>
</tr>
<tr>
<td>PERAMB</td>
<td>perambulative</td>
<td></td>
</tr>
<tr>
<td>PFV</td>
<td>perfective</td>
<td></td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
<td></td>
</tr>
<tr>
<td>PNCT</td>
<td>punctual</td>
<td></td>
</tr>
<tr>
<td>POSS</td>
<td>possessive</td>
<td></td>
</tr>
<tr>
<td>PRF</td>
<td>perfect</td>
<td></td>
</tr>
<tr>
<td>PRO</td>
<td>emphatic or contrastive proform</td>
<td></td>
</tr>
<tr>
<td>PROG</td>
<td>progressive</td>
<td></td>
</tr>
<tr>
<td>PROS</td>
<td>prosecutive case</td>
<td></td>
</tr>
<tr>
<td>PROX</td>
<td>proximate specifier</td>
<td></td>
</tr>
<tr>
<td>PRS</td>
<td>present</td>
<td></td>
</tr>
<tr>
<td>PRT</td>
<td>particle</td>
<td></td>
</tr>
<tr>
<td>PST</td>
<td>past</td>
<td></td>
</tr>
<tr>
<td>PTCP</td>
<td>participle</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>question particle/question marker</td>
<td></td>
</tr>
<tr>
<td>QU</td>
<td>qualifier</td>
<td></td>
</tr>
<tr>
<td>QUOT</td>
<td>quotative</td>
<td></td>
</tr>
<tr>
<td>RC</td>
<td>relative case</td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>repetitive</td>
<td></td>
</tr>
<tr>
<td>REFL</td>
<td>reflexive</td>
<td></td>
</tr>
<tr>
<td>REL</td>
<td>relative</td>
<td></td>
</tr>
<tr>
<td>RMP</td>
<td>remote past</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>single argument of canonical intransitive verb</td>
<td></td>
</tr>
<tr>
<td>SBJ</td>
<td>subject</td>
<td></td>
</tr>
<tr>
<td>SBJV</td>
<td>subjunctive</td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
<td></td>
</tr>
<tr>
<td>SHIFT</td>
<td>perspective-shifting suffix</td>
<td></td>
</tr>
<tr>
<td>SIM</td>
<td>simultaneous</td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>same-subject marker</td>
<td></td>
</tr>
<tr>
<td>STD</td>
<td>standing</td>
<td></td>
</tr>
<tr>
<td>SUB</td>
<td>subordinator</td>
<td></td>
</tr>
<tr>
<td>TC</td>
<td>thematic consonant</td>
<td></td>
</tr>
<tr>
<td>TEL</td>
<td>telic</td>
<td></td>
</tr>
<tr>
<td>TH</td>
<td>thematic prefix/ thematic suffix</td>
<td></td>
</tr>
<tr>
<td>TR</td>
<td>transitive</td>
<td></td>
</tr>
<tr>
<td>VBLZ</td>
<td>verbalizer/verbalization</td>
<td></td>
</tr>
</tbody>
</table>
List of online materials

Chapter 3: Dataset
http://dx.doi.org/10.21942/uva.6834188

Chapter 4: Dataset
https://doi.org/10.21942/uva.7172012

Chapter 5: Exact glosses of the incorporating verbs found in ten or more of the sample languages in the typological survey
https://doi.org/10.21942/uva.12161724

Chapter 5: Questionnaire on verb-based restrictions on incorporation
https://doi.org/10.21942/uva.12161748

Chapter 5: Examples and lexical translations for (non-)incorporating verb meanings
https://doi.org/10.21942/uva.12161751

Chapter 5: Incorporation ability and frequency across verbs and languages
https://doi.org/10.21942/uva.12161769
1 Introduction

1.1 Background
Incorporation involves the integration of a lexical element into a word containing another lexical element, such that two lexical elements occur in a single word (Mithun 1994: 5024; Gerds 1998: 84; Haugen 2015: 414). This phenomenon exists in a large number of genealogically and geographically diverse languages and is often associated with the notion of polysynthesis (Mithun 1994: 5024, 2000: 926–927; Haspelmath and Sims 2010: 138; Velupillai 2012a: 120; Murasugi 2014: 283–284; Genee 2018: 243). The most extensively studied type of incorporation is noun incorporation, in which a noun and a verb constitute a verbal word together (Gerds 1998: 84; Iturrioz Leza 2001: 714). This type of incorporation is illustrated in example (1) from Mapudungun.¹

(1) Noun incorporation in Mapudungun
   a. Ñi chao kintu-le-y ta.chi pu waka.
      my father seek-PROG-3SG.SBJ.IND the COLL cow
      ‘My father is looking for the cows.’
   b. Ñi chao kintu-waka-le-y.
      my father seek-cow-PROG-3SG.SBJ.IND
      ‘My father is looking for the cows.’

In many incorporating languages, noun incorporation constructions may correspond to multi-word constructions in which the same noun and verb form separate words (Gerds 1998: 83–85; Massam 2017), as exemplified for Mapudungun in (1). The clause in example (1a) is a regular transitive clause. It contains, in addition to the subject noun phrase Ñi chao ‘my father’, a transitive verb with the stem kintu ‘seek’ and the direct object noun phrase ta.chi pu waka ‘the cows’. Example (1b) consists of the same lexical elements, but here the noun waka is incorporated into the verb. The position of the noun between the verbal stem kintu and the verbal suffixes -le and -y indicates that the construction is a single verbal word.

Constructions that are considered to result from incorporation are known to display large variation in their grammatical properties, both cross- and intra-linguistically (Murasugi 2014: 284; Johns 2017; Massam 2017). This variation concerns pragmatic, semantic, morphological, phonological as well as lexical characteristics of the constructions. With respect to pragmatics, there is variation

¹ Glosses in the examples are adapted to the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/resources/glossing-rules.php). The use of ‘*’ shows that an example is ungrammatical.
Incorporation: Constraints on variation

between incorporated nouns that are used to refer and incorporated nouns that function non-referentially (Massam 2001: 169–171, 174–175, 2009: 1084, 2017; Chung and Ladusaw 2003: 126–128; Farkas and De Swart 2003: 148; Murasugi 2014: 284–285; Borik and Gehrke 2015: 6). In Paraguayan Guaraní, for instance, incorporated nouns have a non-referential function: it is not possible to refer to them anaphorically (Velázquez-Castillo 1995: 677–678, 694), as demonstrated in example (2).

(2) Incorporation of a non-referentially used noun in Paraguayan Guaraní
*A-hova-hei-se pe-mitã, pero i-sy he’i
1.ACSBJ-face-wash-DES that-child but 3.INCSBJ-mother say
nda-i-ky’a-i ha.
NEG-3.INCSBJ-dirty-NEG that
‘I wanted to wash the child’s face but his mother said that it wasn’t dirty.’
(Velázquez-Castillo 1995: 694; Velazquez Castillo 1996: 144)

In Bininj Kun-Wok, by contrast, an incorporated noun can co-occur with a demonstrative that appears external to the incorporation construction, but that nevertheless relates to the incorporated noun (Evans 2003: 452). This possibility, exemplified in (3), shows that incorporated nouns can be used referentially in this language.

(3) Incorporation of a referentially used noun in Bininj Kun-Wok
Nga-murrng-bimbom na-mekke.
1>3-bone-paint.PST.PFV M-DEM
‘I painted those bones.’
(Evans 2003: 235)

The demonstrative na-mekke in (3) also illustrates that incorporated nouns may be semantically modified. Languages vary with respect to the possibility to combine an incorporation construction with a demonstrative or another modifier such as an adjective, quantifier or relative clause outside the incorporation construction, modifying the incorporated noun: some languages generally allow the presence of such modifiers, but in other languages this type of modification is restricted or not possible at all (Gerds 1998: 89–90; Muro 2009: 100; Murasugi 2014: 284). Western Frisian differs from Bininj Kun-Wok in that its incorporated nouns cannot be combined with external modifiers (Dijk 1997: 15–16). Thus, the incorporated noun *jerappel ‘potato’ cannot co-occur with an adjective, as demonstrated in (4a), or with an article or demonstrative, as can be seen in (4b).
Incorporation of a non-modifiable noun in Western Frisian

a. *Heit sit te grouwte jerappel-skilen
   father sits to huge potato-peel
   ‘Father is sitting, peeling huge potatoes.’
   (Dijk 1997: 16)

b. Heit sit te (*de/*in/*dy)jerappel-skilen
   father sits to DEF/INDF/DEM potato-peel
   ‘Father is sitting, peeling (*the/a/that/those) potatoes.’
   (Dijk 1997: 44)

Another domain of semantic variation in noun incorporation constructions concerns the status of the incorporated nouns as arguments or modifiers of the incorporating verbs (Gerdts 1998: 86–87; Mithun 2000: 917; Lehmann and Verhoeven 2005: 118; Sadock 2006: 585; Aikhenvald 2007: 19; Mursasugi 2014: 284; Borik and Gehrke 2015: 2; Haugen 2015: 414–415; Massam 2017). Incorporated nouns typically represent semantic arguments of their incorporating verbs. However, incorporated nominal modifiers are found in some languages as well. For example, Chukchi does not only show incorporated nominal arguments, such as qora ‘reindeer’ in (5), but also allows incorporated nominal modifiers (Spencer 1995: 455–459), exemplified by gətg ‘lake’ in (6b).

(5) Incorporation of a nominal argument in Chukchi
   taŋ-amən Cəkwənqəqəj ɣa-qora-nm-at-len
   INTS-alone personal.name.ABS.SG PRF-reindeer-kill-TH-3SG.S
   ‘Cəkwənqəqəj all by himself slaughtered reindeer.’
   (Dunn 1999: 222, 226)

(6) Incorporation of a nominal modifier in Chukchi
   a. gətg-eto qat-gəl wələŋən
      lake-DAT go-3SG.S raven.ABS.SG
   b. gətg-əqat-gəe wələŋən
      lake-go-3SG.S raven.ABS.SG
      ‘Raven went to the lake.’
      (Skorik 1948: 72–73, cited in Spencer 1995: 458)

Incorporation constructions also display variation in their morphosyntactic properties. One domain of morphosyntactic variation in incorporation constructions that has not received much attention in the incorporation literature so far (but see Iturrioz Leza 2001; Aikhenvald 2007; Muro 2009; Barrie and Mathieu 2016) concerns the morphosyntactic form of incorporated elements, i.e. their formal complexity. While the incorporated nouns in the examples in (1)–(6) above are simple stems,
Incorporation: Constraints on variation

consisting of a single morpheme, in some languages it is also possible to incorporate nominal inflected words, as shown for Ket in (7). Here, the incorporated noun don'-aŋ contains the inflectional suffix -aŋ marking plural number.

(7) Incorporation of a nominal inflected word in Ket
don'-aŋ-s'i-vet
1SG.SBJ-knife-PL-PRS-E-make
‘I’m making knives.’
(Drossard 2002: 235)

Phonologically, incorporated nouns are often identical to corresponding unincorporated nouns, but in some languages they take special phonological forms (Aikhenvald 2007: 13; Caballero et al. 2008: 387–388). The noun k’ab ‘hand’ from Yucatec Maya has the same phonological form in the incorporation construction in (8b) as in the construction without incorporation in (8a) (Lehmann and Verhoeven 2005: 149).

(8) Incorporation of a noun in its regular phonological form in Yucatec Maya
a. k-u lom-ik yéetel u k’ab
   IPFV-3.SBJ poke-INCOMPL with 3.POSS hand
   ‘He pokes it with his hand.’
b. k-u loom-k’ab-t-ik
   IPFV-3.SBJ poke-hand-TR-INCOMPL
   ‘He pokes it with his hand.’

Example (9b) from Sora, by contrast, shows the incorporated noun te ‘banana’, which differs phonologically from the corresponding noun found in contexts without incorporation, i.e. konte-n, exemplified in (9a).

(9) Incorporation of a suppletive noun in Sora
a. pen konte-n dyum-t-ai
   I banana-NSFX eat-NPST-1.SBJ
   ‘I am eating a banana.’
b. pen dyum-te-ti-n-ai
   I eat-banana-NPST-INTR-1.SBJ
   ‘I am eating a banana.’
   (Anderson 2017: 939)

Finally, incorporation constructions vary in the extent to which they are subject to lexical restrictions. For example, in some languages only body-part nouns can be
incorporated (Aikhenvald 2007: 20; Massam 2009: 1090), while other languages allow other incorporated nouns as well (see e.g. the examples in Spencer 1995: 449–450 [for Chukchi]; Lehmann and Verhoeven 2005: 148–149 [for Yucatec Maya]; Haude 2006: 378 [for Movima]; the examples in Anderson 2017: 937 [for Sora]).

1.2 Research aim and research topics
The aim of this thesis is to add to our understanding of the cross- and intra-linguistic variation in incorporation constructions by examining the range of variation regarding their pragmatic, semantic, morphosyntactic, phonological and lexical properties and by focusing specifically on constraints on this variation. The thesis starts out with a broad review of the pragmatic, semantic and phonological variation in incorporation and the constraints on this variation identified in previous literature, in Chapter 2. Three in-depth studies of particular domains of morphosyntactic, pragmatic-semantic and lexical variation follow in Chapter 3, 4 and 5, respectively. These domains of variation are analyzed in detail, based on data from a varied sample of languages. The variation and constraints investigated in Chapter 2 to 5 are subsequently discussed in light of several theoretical approaches to incorporation in Chapter 6, in order to examine their implications for theoretical accounts of incorporation.

In Chapter 2, several pragmatic, semantic and phonological domains of variation in noun incorporation are considered on the basis of findings from previous literature. For each of these domains, constraints are formulated that capture their variation. The constraints generally follow from implicational relationships describing the cross-linguistic distribution of the various possibilities within a particular domain. These implicational relationships are often called implicational hierarchies, especially when more than two possibilities within a domain are implicationally related. For the relevant domain, incorporating languages then only vary with respect to their cut-off point on the hierarchy. Alternatively, the constraints take the form of what may be called basic settings. A basic setting specifies a restricted set of possibilities within a certain pragmatic, semantic or phonological domain and for each incorporating language only one of these possibilities holds. The different constraints are illustrated on the basis of examples from 27 noun-incorporating languages. Moreover, it is shown how the constraints together determine the total range of pragmatic, semantic and phonological possibilities for incorporation constructions in a particular noun-incorporating language.

The morphosyntax of incorporated elements across languages is the research topic of Chapter 3. Here, the morphosyntactic possibilities for incorporated elements are examined on the basis of a systematic analysis of data from a sample of 30 incorporating languages. While it has often been claimed that incorporated elements may only be simple stems, I make a distinction between the incorporation of single lexical morphemes, derived stems, inflected words, phrases and clauses. The data
show that an implicational hierarchy can be set up ranging from the simplest forms, i.e. single lexical morphemes, to forms as complex as phrases: the morphosyntactic forms of incorporated elements vary within and across languages, but, at the same time, the variation is constrained in that the incorporation of more complex forms implies the incorporation of simpler forms. Note that in this chapter, in contrast to the other chapters, not only incorporated nouns but also incorporated verbs, adjectives, adverbs and adpositions are included.

In Chapter 4 I present an investigation of the pragmatic referentiality and semantic modifiability of incorporated nouns. These two domains have already attracted considerable attention in the incorporation literature, but this chapter’s aim is to study the variation in the referentiality and modifiability of incorporated nouns in a balanced sample of languages. I distinguish between three pragmatic-semantic types of nouns: referentially used modifiable nouns (+R/+M nouns), non-referentially used modifiable nouns (−R/+M nouns) and non-referentially used non-modifiable nouns (−R/−M nouns). Data from the 21 noun-incorporating languages included in the sample demonstrate that all three types of nouns can be incorporated and that languages may display incorporated nouns of more than one of these types. While incorporated nouns of the different types may generally occur in languages independently of each other, the incorporation of −R/+M nouns appears to be implicationally related to the incorporation of +R/+M incorporated nouns, i.e. all languages showing −R/+M incorporated nouns also allow +R/+M incorporated nouns. This interdependency indicates that the distribution of incorporated nouns of some of the pragmatic-semantic types is nevertheless restricted.

An examination of the types of verbs incorporating nouns across languages follows in Chapter 5. Constraints on the verbs that can be used in noun incorporation constructions have not yet been systematically explored: most noun incorporation research has primarily concentrated on the properties of incorporated nouns rather than on those of incorporating verbs. The chapter consists of a typological survey of incorporating verbs in descriptive sources for a sample of 50 incorporating languages and a more detailed set of case studies on incorporating verbs in corpus data from eight languages. The variation in incorporating verbs is investigated, and it is examined whether particular restrictions are imposed on incorporating verbs cross-linguistically. The findings indicate that both the morphosyntactic transitivity of a verb, which is likely to have a semantic basis, and idiosyncratic factors have an effect on the possibility for a verb to show noun incorporation. Thus, there are constraints on the set of individual verbs that allow noun incorporation, but these constraints are predictable only to a certain extent.

Importantly, variation in incorporation constructions and constraints on this variation are also relevant for the recurrent question in incorporation research how or in which part of the linguistic system incorporation constructions are formed. This question relates to the traditional distinction made in the incorporation literature
between approaches in which incorporation is considered a lexical process and approaches in which it is assumed that incorporation is syntactic in nature (Mithun 1994: 5025; Massam 2009: 1083–1086, 2017; Štekauer et al. 2012: 43–47; Murasugi 2014: 286–288; Haugen 2015: 415–421). According to studies such as Sapir (1911), Mithun (1984, 1986a), Di Sciullo and Williams (1987), Rosen (1989) and Anderson (2000, 2005), incorporation is a type of lexical word formation or, more specifically, a type of lexical compounding. By contrast, researchers including Sadock (1985, 1986, 1991), Baker (1988, 1996, 2009) and Barrie and Mathieu (2016) regard incorporation as a syntactic process, arguing that incorporated elements show properties that are characteristic of independent syntactic constituents rather than of internal parts of words. In addition, syntactic accounts of incorporation and compounding have been proposed that are based on a single-engine hypothesis, i.e. in which all word formation is assumed to be syntactic, such that incorporation constructions must be the result of a syntactic process as well (Harley 2009; Wiltschko 2009). As the explanatory potential of theoretical accounts of incorporation depends on their ability to capture the full range of variation in incorporation and the constraints on this variation, the findings of the present thesis may also affect the debate about the theoretical status of incorporation. I discuss the theoretical contribution of the four studies in Chapter 2 to 5 in Chapter 6.

1.3 Approach and data

With the exception of the study in Chapter 5, this thesis takes a Functional Discourse Grammar approach to incorporation. Functional Discourse Grammar (FDG, Hengeveld and Mackenzie 2008) is a typologically-oriented, functional linguistic framework that aims to account for the properties of linguistic utterances on the basis of their communicative functions. The framework allows for a characterization of incorporation as a grammatical process involving semantics, morphosyntax and, more indirectly, pragmatics and phonology. Based on this characterization, the FDG approach may be grouped with the syntactic accounts of incorporation discussed in the previous section, rather than with the lexical ones. FDG’s Grammatical Component distinguishes four independent but interrelated levels: an Interpersonal Level, which contains pragmatic units, a Representational Level, responsible for semantics, a Morphosyntactic Level and a Phonological Level. Incorporation can then be defined as a process in which two units that are in a semantic dependency relation at the Representational Level form a single word at the Morphosyntactic Level. In the case of noun incorporation, the units are nominal and verbal.

According to this FDG-based definition, there are no a priori requirements or restrictions with respect to other semantic or morphosyntactic characteristics of

---

2 Technical terms as applied in FDG are capitalized (see Hengeveld and Mackenzie 2008: 44).
incorporation or regarding pragmatic or phonological characteristics of incorporation. Correspondingly, on the basis of the FDG approach taken in this thesis, it is expected that there is variation within both semantic, morphosyntactic, pragmatic and phonological domains of incorporation. At the same time, FDG generally seeks to formulate constraints on linguistic variation and assumes that such constraints may often take the form of implicational hierarchies (Hengeveld and Mackenzie 2008: 31–37, in prep.). In the same way, this thesis aims to establish constraints on the semantic, morphosyntactic, pragmatic and phonological variation in incorporation constructions.

Incorporation is closely connected to compounding: both phenomena involve the combination of two lexical elements in a single morphosyntactic word. The relation between incorporation and compounding is an important but complex issue, as it can be understood in different ways. While researchers taking a lexical approach to incorporation tend to claim that incorporation is a subtype of compounding (Sapir 1911: 257; Mithun 1984: 847; Di Sciullo and Williams 1987: 63–69), researchers such as Baker (1988: 78) and Barrie and Mathieu (2016: 3–5), taking a syntactic approach to incorporation, argue that compounding and incorporation are two distinct processes, the former lexical and the latter syntactic in nature. In FDG, a distinction is made between lexical compounding and grammatical compounding (Hengeveld and Mackenzie 2016: 1150–1153). Lexical compounds are formed in the Lexicon and involve a non-productive process that combines two lexical items to create a new item with unpredictable semantics. Grammatical compounds, by contrast, are productively created in the Grammatical Component by combining semantic units at the Representational Level into a semantically compositional whole that is expressed as a single word at the Morphosyntactic Level. In this thesis, incorporation then comprises all grammatical compounding in which a semantic dependency relation exists between the two units at the Representational Level.

As the thesis focuses on cross- and intra-linguistic variation in incorporation and constraints on this variation, the four studies in Chapter 2 to 5 are based on data from a large set of incorporating languages with different genealogical and geographical backgrounds. In order to enable a suitable selection of incorporating languages for the studies, I first compiled a list of languages that were identified as incorporating languages in earlier literature. This list, which is presented in Appendix 1, is based primarily on a survey of noun-incorporating languages by Velupillai (2012b), which includes the noun-incorporating languages referred to in the well-known typological study on noun incorporation by Mithun (1984), the review articles

---

3 Because various definitions of incorporation occur in the incorporation literature, these languages do not necessarily show constructions that can be considered incorporation constructions according to the FDG-based definition of incorporation used in the present thesis. For the studies in Chapter 2, 3 and 4, which make use of the FDG approach, it was therefore specifically verified that the selected languages display incorporation as defined in this thesis.
on incorporation by Gerdts (1998) and Aikhenvald (2007) and a number of studies on individual incorporating languages. The list was extended on the basis of various theoretical works on incorporation (Sapir 1911; Sadock 1980, 1985, 1986; Baker 1988, 1996; Rosen 1989; Anderson 2000), several review articles on incorporation (Mithun 1994, 2010; Iturrioz Leza 2001; Anderson 2007; Massam 2009) and a number of articles on incorporation taking a cross-linguistic perspective (Caballero et al. 2008; Štukauer et al. 2012; Barrie and Mathieu 2016). Based on a search of the *Linguistic bibliography* (Bobyleva et al. n.d.) and the *Modern Language Association international bibliography*, I also added languages to the list that are studied in incorporation articles found in these bibliographies. Finally, I came across a number of additional incorporating languages during the research process itself, and these languages were included in the list as well.

The languages investigated in the studies in Chapter 2 to 5 were all selected from this list of incorporating languages. For Chapter 2, no systematic sampling procedure was used, but all languages illustrating the various patterns of variation were taken from this list. For the studies in Chapter 3 and 4, by contrast, diversity samples were drawn from this list in a consistent way (see Section 3.4.1 and 4.3.1 for details). The languages studied in the typological survey in Chapter 5 form a genealogically diverse set of languages from the list, and the corpus-based case studies in the same chapter relate to eight languages that can be regarded as a convenience sample of languages from the list. Table 1 presents the set of languages addressed in the four studies. Note that in this table, as in the rest of the thesis, the language names as given in Glottolog (Hammarström et al. 2017) are used, which means that in some cases the language names in the thesis differ from the names used in the literature consulted.

**Table 1.** Languages studied in the various chapters, indicated in the numbered columns, included in the thesis. The language names, language families, macro-areas and countries are based on Glottolog (Hammarström et al. 2017). “*1*” indicates that a language was initially included in the sample of Chapter 4 but had to be excluded eventually due to insufficient available data. “*2*” indicates that a language is considered a language family rather than a single language in Glottolog. The last column mentions the language experts consulted for the analysis of the languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>Language family</th>
<th>Macro-area</th>
<th>Country</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Expert consulted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alamblak</td>
<td>Sepik</td>
<td>Papunesia</td>
<td>Papua New Guinea</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Atsugewi</td>
<td>Palaihnihan</td>
<td>North America</td>
<td>United States</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Baure</td>
<td>Arawakan</td>
<td>South America</td>
<td>Bolivia, Plurinational State of</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>Swintha Danielsen</td>
</tr>
<tr>
<td>Language</td>
<td>Subgroup</td>
<td>Region</td>
<td>Country/Region</td>
<td>Inclusion</td>
<td>Variation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------------</td>
<td>--------------</td>
<td>------------------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bininj Kun-Wok</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Nicholas Evans</td>
<td></td>
</tr>
<tr>
<td>Caddo</td>
<td>Caddoan</td>
<td>North America</td>
<td>United States</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cayuga</td>
<td>Iroquoian</td>
<td>North America</td>
<td>Canada; United States</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chamalapa Zoque</td>
<td>Mixe-Zoque</td>
<td>North America</td>
<td>Mexico</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chukchi</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia</td>
<td>Russian Federation</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Michael Dunn</td>
<td></td>
</tr>
<tr>
<td>Crow</td>
<td>Siouan</td>
<td>North America</td>
<td>United States</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Canadian Inuktitut</td>
<td>Eskimo-Aleut</td>
<td>North America</td>
<td>Canada</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Ojibwa</td>
<td>Algic</td>
<td>North America</td>
<td>Canada</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ese Eja</td>
<td>Puco-Tacanan</td>
<td>South America</td>
<td>Bolivia; Plurinational state of Peru</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>Marine Vuillermet</td>
<td></td>
</tr>
<tr>
<td>Guahibo</td>
<td>Guahiboan</td>
<td>South America</td>
<td>Colombia; Venezuela; Bolivarian Republic of</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guarayu</td>
<td>Tupian</td>
<td>South America</td>
<td>Bolivia; Plurinational State of</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>Swintha Danielsen</td>
<td></td>
</tr>
<tr>
<td>Haida²</td>
<td>Haida</td>
<td>North America</td>
<td>Canada; United States</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halkomelem</td>
<td>Salishan</td>
<td>North America</td>
<td>Canada; United States</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hokkaido AINU</td>
<td>Ainu</td>
<td>Eurasia</td>
<td>Japan</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Anna Bugaeva, Tomomi Sato</td>
<td></td>
</tr>
<tr>
<td>Iraqw</td>
<td>Afro-Asiatic</td>
<td>Africa</td>
<td>Tanzania; United Republic of</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Tjeu Claessen, Maarten Mous</td>
<td></td>
</tr>
<tr>
<td>Kalaallisut</td>
<td>Eskimo-Aleut</td>
<td>Eurasia</td>
<td>Greenland</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Michael Fortescue</td>
<td></td>
</tr>
<tr>
<td>Kalamang</td>
<td>West Bomberai</td>
<td>Papunesia</td>
<td>Indonesia</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>Eline Visser</td>
<td></td>
</tr>
<tr>
<td>Karajá</td>
<td>Nuclear-Macro-Je</td>
<td>South America</td>
<td>Brazil</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ket</td>
<td>Yeniseian</td>
<td>Eurasia</td>
<td>Russian Federation</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Stefan Georg</td>
<td></td>
</tr>
<tr>
<td>Malayo</td>
<td>Chibchan</td>
<td>South America</td>
<td>Colombia</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mamaindé</td>
<td>Nambiquaran</td>
<td>South America</td>
<td>Brazil</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mandinka</td>
<td>Mande</td>
<td>Africa</td>
<td>Gambia; Guinea; Guinea-Bissau; Senegal</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mapudungun</td>
<td>Araucanian</td>
<td>South America</td>
<td>Argentina; Chile</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Fernando Zúñiga</td>
<td></td>
</tr>
<tr>
<td>Marithiel</td>
<td>Western Daly</td>
<td>Australia</td>
<td>Australia</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>Language Type</td>
<td>Region</td>
<td>Countries/Country</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Authors</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>--------</td>
<td>-------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Mohawk</td>
<td>Iroquoian</td>
<td>North America</td>
<td>Canada; United States</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Marianne Mithun</td>
<td></td>
</tr>
<tr>
<td>Movima</td>
<td>Movima (Isolate)</td>
<td>South America</td>
<td>Bolivia, Plurinational State of Peru</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Katharina Haude</td>
<td></td>
</tr>
<tr>
<td>Murui Huitoto</td>
<td>South America</td>
<td>Colombia; Peru</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nadeb</td>
<td>Nadahup</td>
<td>South America</td>
<td>Brazil</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nisga’a</td>
<td>Tsimshian</td>
<td>North America</td>
<td>Canada</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Niuean</td>
<td>Austronesian</td>
<td>Papuasia</td>
<td>Niue</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Diane Massam</td>
<td></td>
</tr>
<tr>
<td>Nivkh</td>
<td>Nivkh (Isolate)</td>
<td>Eurasia</td>
<td>Russian Federation</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast Maidu</td>
<td>North America</td>
<td>United States</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern Gumuz</td>
<td>Africa</td>
<td>Ethiopia; Sudan</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>Colleen Ahland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuu-chah-nulth</td>
<td>South America</td>
<td>United States</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palikúr</td>
<td>Arawakan</td>
<td>South America</td>
<td>Brazil, French Guiana</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Alexandra Aikhenvald</td>
<td></td>
</tr>
<tr>
<td>Panamint</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>United States</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panare</td>
<td>Cariban</td>
<td>South America</td>
<td>Venezuela, Bolivarian Republic of</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Thomas Payne</td>
<td></td>
</tr>
<tr>
<td>Paraguayan Guaraní</td>
<td>South America</td>
<td>Argentina</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plains Cree</td>
<td>Algic</td>
<td>North America</td>
<td>Canada; United States</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>Arok Wolvengrey</td>
<td></td>
</tr>
<tr>
<td>Sora</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Gregory Anderson</td>
<td></td>
</tr>
<tr>
<td>South Slavey</td>
<td>Athapaskan-Eyak-Tlingít</td>
<td>North America</td>
<td>Canada</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Keren Rice</td>
<td></td>
</tr>
<tr>
<td>Southern Tiwa</td>
<td>Kiowa-Tanoan</td>
<td>North America</td>
<td>United States</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Donald Frantz</td>
<td></td>
</tr>
<tr>
<td>Takelma</td>
<td>Takelma (Isolate)</td>
<td>North America</td>
<td>United States</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanimuca-Ketuará</td>
<td>Tarucano</td>
<td>South America</td>
<td>Colombia</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiwi</td>
<td>Tiwi (Isolate)</td>
<td>Australia</td>
<td>Australia</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ute-Southern Pauite</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>United States</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Talmy Givón</td>
<td></td>
</tr>
<tr>
<td>Warembori</td>
<td>Austronesian</td>
<td>Papuasia</td>
<td>Indonesia</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washo</td>
<td>Washo (Isolate)</td>
<td>North America</td>
<td>United States</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Frisian</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Netherlands</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Highland Chatino</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>William Foley</td>
<td></td>
</tr>
<tr>
<td>Yele</td>
<td>Yele (Isolate)</td>
<td>Papuasia</td>
<td>Papua New Guinea</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yimas</td>
<td>Lower Sepik-Ramu</td>
<td>Papuasia</td>
<td>Papua New Guinea</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>William Foley</td>
<td></td>
</tr>
</tbody>
</table>
Each of the studies that forms part of the thesis makes use of descriptive sources on incorporation in the selected languages. These sources include reference grammars and articles on incorporation and related processes in the relevant languages. In addition, experts on the incorporating languages investigated were consulted. The researchers who helped to analyze the languages of their expertise are mentioned in Table 1 above. Finally, the case studies in Chapter 5 are based primarily on corpus data and data elicited by language experts. The data from the descriptive sources, language experts and corpora form the basis of the four studies on the cross- and intra-linguistic variation in incorporation constructions and the constraints on this variation, which are presented in the next chapters.
2 Noun incorporation in Functional Discourse Grammar

2.1 Introduction

Noun incorporation concerns the situation in which a nominal unit combines with a verbal unit to form a single verbal word (Gerdts 1998: 84; Mithun 2000: 916; Aikhenvald 2007: 11; Massam 2017). An initial example from Yucatec Maya is given in (1).

(1) Noun incorporation in Yucatec Maya
   a. t-in ch’ak-ah che’ ichil in kòol
      PST-1SG.SBJ cut-COMPL tree in 1SG.POSS milpa
      ‘I chopped trees in my cornfield.’
   b. h ch’ak-che’-nah-en ichil in kòol
      PST cut-tree-COMPL-1SG.ABS in 1SG.POSS milpa
      ‘I chopped trees in my cornfield.’

Example (1a) shows a regular transitive clause in Yucatec Maya, with a verb with the stem ch’ak ‘cut’ and an object noun che’ ‘tree’. In example (1b), the noun che’ is incorporated into the verb: the noun here follows the verbal stem ch’ak but precedes the verbal inflectional suffixes.

Noun incorporation constructions show highly varied properties cross-linguistically. For instance, languages differ in whether or not their incorporated nouns can be used to refer (Massam 2009: 1084; Murasugi 2014: 284–285; Borik and Gehrke 2015: 6) and whether or not incorporation functions to background the participant designated by the incorporated noun (Mithun 1984: 859; Gerdts 1998: 86). In addition, whereas some languages restrict noun incorporation to semantic arguments, others also show incorporated modifiers (Mithun 1984: 875; Gerdts 1998: 87; Murasugi 2014: 284). Besides, in many but not all incorporating languages there are, for most noun incorporation constructions, corresponding constructions in which the noun and verb appear as separate morphosyntactic words (Mithun 1984: 847–848; Gerdts 1998: 84–85; Massam 2017), as exemplified for Yucatec Maya in (1). Furthermore, incorporated nouns may be phonologically identical to non-incorporated nouns.

---

1 This chapter is a slightly adapted version of: Olthof, Marieke & Kees Hengeveld. subm. Noun incorporation in Functional Discourse Grammar.

2 Glosses in the examples are adapted to the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/resources/glossing-rules.php). The use of “*” shows that an example is ungrammatical, whereas the use of “#” indicates that an example is semantically anomalous.
nouns in the same language or may have specialized forms (Mithun 1984: 876; Caballero et al. 2008: 387–388).

Because the various pragmatic, semantic, morphosyntactic and (morpho)phonological properties associated with incorporated nouns appear to be combined in different ways in different languages, noun incorporation is particularly interesting for discussions about interfaces in grammatical theory. An interface can be defined as a set of rules that states the possible relations between different types of grammatical representations. With its four independent but interrelated grammatical levels, i.e. the Interpersonal Level (IL), Representational Level (RL), Morphosyntactic Level (ML) and Phonological Level (PL), which contain pragmatic, semantic, morphosyntactic and phonological representations respectively, Functional Discourse Grammar (FDG, Hengeveld and Mackenzie 2008) provides a suitable framework to study interface conditions in noun incorporation (see Section 2.2).3

In this chapter, we provide an FDG analysis of the interface conditions involved in noun incorporation. Following Hengeveld and Mackenzie (in prep.), we assume that differences between interface conditions across languages can often be defined in terms of implicational hierarchies or constraints, such that for every language its cut-off point on the many hierarchies involved will predict the working of the interfaces. Correspondingly, this chapter proposes, based on earlier literature and data from a large number of incorporating languages, a set of hierarchies that determine the constraints on the possible mappings between the FDG levels in noun incorporation in different languages. In addition to these hierarchies, a number of basic settings concerning noun incorporation is provided, which state, for instance, whether a language allows incorporation at all and which alignment system is applied in incorporation.

We first introduce the FDG framework and its approach to interfaces in Section 2.2. Our FDG definition of noun incorporation follows in Section 2.3. Subsequently, we look at the relevant interfaces between pairs of levels, where the pairs are presented in a top-down manner. As incorporation is a morphosyntactic phenomenon, ML is always involved in these pairs. Thus, the relevant interfaces are the IL-ML interface, discussed in Section 2.4, the RL-ML interface, addressed in Section 2.5, and the ML-PL interface, examined in Section 2.6. We provide examples concerning the relevant pairs in each of these sections, but in Section 2.7 we exemplify for one language, Kalaallisut, how the interfaces between the three different pairs of levels together capture the possibilities for noun incorporation in this language. Here we also exemplify how the pragmatic, semantic, morphosyntactic and phonological properties of incorporated nouns may or may not match across the different levels in FDG. In Section 2.8 we then discuss our findings and draw our conclusions.

---

3 Technical terms as applied in FDG are capitalized (see Hengeveld and Mackenzie 2008: 44).
2.2 Functional Discourse Grammar

FDG is a typologically-based theory of language structure with the four-level architecture shown in Figure 1. The figure shows that FDG is the Grammatical Component of a wider theory of verbal interaction, in which it interacts with a Conceptual, Contextual and Output Component. Figure 1 also shows that FDG has a top-down organization, working down from larger to smaller units.

Within the Grammatical Component itself, there are four levels of analysis. Two of these, the Interpersonal Level and the Representational Level, are the output
of the operation of Formulation. This operation converts conceptual representations into semantic and pragmatic representations. The Morphosyntactic Level and the Phonological Level are the output of the operation of Encoding, which translates pragmatic and semantic representations into morphosyntactic and phonological ones.

Internally, every level is hierarchically organized in terms of layers relevant to that level. For instance, at the (actional) Interpersonal Level, layers such as the Discourse Act and the Referential Subact are relevant; at the (designational) Representational Level, layers such as the Propositional Content and the State-of-Affairs are needed; at the Morphosyntactic Level layers such as the Noun Phrase and the Clause are used; finally, at the Phonological Level prosodic units such as the Intonation Phrase and the Phonological Word are relevant.

Layers may be further modified by modifiers, operators and functions. Modifiers differ from operators and functions in being lexical rather than grammatical. The difference between operators and functions is that the latter are relational while the former are not. Examples of operators that will show up later in this article are identifiability and specificity operators that operate on Referential Subacts at the Interpersonal Level. Examples of modifiers are adjectives that modify Individuals and locative phrases that modify States-of-Affairs, both at the Representational Level. Finally, examples of functions are the Actor and Undergoer functions of arguments at the Representational Level and the Subject function of Noun Phrases at the Morphosyntactic Level.

In the next section we will consider how noun incorporation fits into this general architecture.

2.3 Defining noun incorporation in FDG

Before moving to the actual interfaces involved in noun incorporation, it is important to indicate how we define noun incorporation in FDG. Various definitions of noun incorporation have been proposed in the literature, differing chiefly in whether they characterize noun incorporation as a lexical or syntactic process (Massam 2009: 1077; Murasugi 2014: 284; Haugen 2015: 414; Johns 2017). In this study, we restrict the term noun incorporation to productive, semantically transparent processes in which a nominal and a verbal unit at RL together form a single verbal Word at ML. Thus, we consider noun incorporation a phenomenon that takes place at the grammatical levels rather than in the Lexicon. More specifically, we define noun incorporation constructions as cases in which a nominal and a verbal unit that are in a dependency relation of the form head-modifier or predicate-argument at RL form a single verbal Morphosyntactic Word.

---

4 Modifiers are examples of the lexemes included in the box on the left-hand side of the Grammatical Component in Figure 1 that feeds into Formulation.
Noun incorporation in Functional Discourse Grammar 17

Note that this definition entails that noun incorporation shows a certain degree of overlap with compounding. In FDG, a distinction can be made between compounds formed by combining lexemes in the Lexicon and semantically transparent compounds that are productively created in the Grammatical Component at RL (Hengeveld and Mackenzie 2016: 1150–1153). The latter type can be further divided into head-modifier, predicate-argument and conjunct-conjunct compounds. Head-modifier and predicate-argument compounds consisting of a nominal and a verbal unit that are morphosyntactically verbal equal noun incorporation as we define it here.

Noun incorporation can also be linked to the notion of polysynthesis. In the literature on noun incorporation, the phenomenon has sometimes even been considered a necessary feature of polysynthetic languages (Genee 2018: 243). Within the FDG framework, Genee (2018: 264) has identified five parameters that contribute to a language’s degree of polysynthesis and noun incorporation may play a role in each of them. Most importantly, noun incorporation leads to higher lexical density, because incorporation of a noun into a verb always results in a Morphosyntactic Word with at least two lexical Morphemes.

According to our definition, noun incorporation takes place at ML. This level distinguishes the morphosyntactic layers presented in (2).

\[(2) \quad \text{Morphosyntactic layers in FDG}\]
\[\text{Le}_n = \text{Linguistic Expression} \]
\[\text{Cl}_n = \text{Clause} \]
\[\text{Xp}_n = \text{Phrase (of type x)} \]
\[\text{Xw}_n = \text{Word (of type x)} \]
\[\text{Xm}_n = \text{Morpheme} \]

Morphemes are further divided into three types: Stems (Xs), Roots (Xr) and Affixes (Aff). Note that the difference between a Stem and a Root in FDG is that a Root cannot occur independently, i.e. without being attached to another lexical Morpheme, while a Stem can (Hengeveld and Mackenzie 2008: 404).

All morphosyntactic layers in (2), except for Le, which represents the maximal morphosyntactic unit, may be embedded into other units, leading potentially to full recursivity. This means that Morphosyntactic Words too may embed other morphosyntactic units, with incorporation as the result. Many different configurations are possible within the template of a Morphosyntactic Word. In (3), we illustrate the possibilities for noun incorporation. For reasons of space, we limit ourselves here to

---

5 Correspondingly, the hierarchies proposed in this chapter for the interface conditions on noun incorporation may also be relevant for other grammatical head-modifier and predicate-argument compounds.
configurations in which the nominal unit precedes the verbal one, even though the reverse can also be found (Caballero et al. 2008), and in which the verbal unit is a Stem (Vs₁), although verbal Roots may incorporate nouns as well.⁶

(3) Morphosyntactic Word templates with incorporated nouns

a. \((Vw₁; \{(Affₙ₁) (Nr₁) (Vs₁) (Affₙ₂)\} (Vw₁))\)

b. \((Vw₁; \{(Affₙₐ) (Ns₁) (Vs₁) (Affₙₐ)\} (Vw₁))\)

c. \((Vw₁; \{(Affₙₐ) (Nw₁) (Vs₁) (Affₙₐ)\} (Vw₁))\)

d. \((Vw₁; \{(Affₙₐ) (Np₁) (Vs₁) (Affₙₐ)\} (Vw₁))\)

The different possible configurations also illustrate the morphosyntactic characteristics that we use to verify that apparent noun incorporation constructions are single Morphosyntactic Words, which is important because our definition of noun incorporation depends on Morphosyntactic Word status. In most cases, the position of a nominal unit between a verbal Affix and a verbal Root or Stem shows that it is incorporated into the verbal Word. In a few languages that do not tend to use Affixes, verbal clitics and particles can be considered in the same way as verbal Affixes. Finally, in some languages the Root status of either the nominal or the verbal unit can be used to recognize incorporation: as Roots necessarily combine with another lexical Morpheme in a Morphosyntactic Word, the occurrence of a nominal Root next to a verbal lexical Morpheme or the occurrence of a verbal Root next to a nominal lexical Morpheme shows that the two form a single Morphosyntactic Word.

Another important aspect of noun incorporation shown in the configurations in (3) is that the incorporated unit may be a nominal Root (Nr₁), as in (3a), a nominal Stem (Ns₁), as in (3b), a nominal Word (Nw₁), as in (3c), or a Noun Phrase (Np₁), as in (3d). A terminological comment is in order now: what is generally called “noun incorporation” is not always “nominal Stem incorporation” but may also be “nominal Root incorporation”, “nominal Word incorporation” or “Noun Phrase incorporation”. In order to avoid unnecessary terminological complexities, we use the term “noun incorporation” for all four situations.⁷

The possibilities represented in (3) do not appear in languages randomly. Chapter 3 investigates the range of morphosyntactic units that may be incorporated cross-linguistically. Based on the results for noun incorporation specifically, the implicational hierarchy given in (4), in which the class of lexical Morphemes includes

---

⁶ We also include constructions with bound verbal units that are sometimes called derivational affixes as noun incorporation constructions, as long as these verbal units form a large group in the relevant language and have concrete, verb-like meanings. These considerations concern languages like Eastern Canadian Inuktitut, Kalaallisut and Nuu-chah-nulth.

⁷ The different types of incorporated units contribute in different degrees to the polysynthetic character of a language, as the inclusion of higher morphosyntactic layers within one Word may be assumed to make a language more polysynthetic than the inclusion of lower morphosyntactic layers within one Word (Genee 2018: 264).
Noun incorporation in Functional Discourse Grammar

both Roots and Stems, may be proposed to describe the possibilities for incorporated nouns.

(4) Implicational hierarchy of incorporated nouns with different morphosyntactic forms

\[ \text{lexical Morpheme} \supset \text{derived Stem} \supset \text{inflected Word} \supset \text{Phrase} \]

This hierarchy expresses that typologically the most common form of noun incorporation concerns the incorporation of lexical Morphemes, followed by grammatically derived Stems, inflected Words and Phrases. Also, the hierarchy states that if, in a particular language, a noun of a category more to the right in the hierarchy can be incorporated, then nouns of all categories to the left can be incorporated as well. Data from 30 languages presented in Chapter 3 largely confirm the hierarchy in (4). It thus seems that languages can be parametrized, in that for every language a particular cut-off point in (4) can be specified at ML. Note that this is not an interface condition, but a restriction that applies in the Morphosyntactic Encoder itself.

Interfaces between ML on the one hand and IL, RL and PL on the other hand are, however, highly relevant for noun incorporation. In noun incorporation constructions, the nominal unit at ML may map onto various units at IL, RL and PL, as will be discussed in the next sections. Some of these mappings create mismatches between levels. Most importantly, noun incorporation constructions typically involve two separate units at RL that form a single unit at ML. Usually it is one of the two arguments of a transitive verb that is incorporated into this verb (see also Section 2.5.3). This means that two units from a single Configurational Property at RL form a unit at ML, while the other argument that plays a role in the same Configurational Property is expressed as a separate unit at ML. In this way, noun incorporation entails a mismatch between RL and ML, i.e. in Encoding (see Section 2.7 for an illustration).

2.4 The IL-ML interface

There are several aspects of IL that (co-)determine whether noun incorporation is or is not allowed in a language. The relevant aspects are the following:

i. The interpersonal category of the incorporated noun: is it a Referential Subact or not?

ii. The head of the incorporated noun: is it a proper name or a common noun?

---

[8] Only one of the 30 languages does not conform to this hierarchy of the forms of incorporated nouns: in Yimas, incorporated nominal Stems and incorporated nominal Words occur, while no examples of incorporated nominal derived Stems are found in the study. Note, however, that the incorporation of adverbial derived Stems is attested.
Incorporation: Constraints on variation

iii. The pragmatic operators applying to the incorporated noun: what are its identifiability (+id/−id) and specificity (+s/−s) values?

iv. The pragmatic function of the incorporated noun: does it have a Focus function, a Background function or none of these?

We will address these aspects one by one in what follows.

2.4.1 The interpersonal category of the incorporated noun

Languages may show non-referential incorporated nouns only, limit incorporation to referential nouns or allow both referential and non-referential incorporated nouns. As mentioned in Chapter 1, in Paraguayan Guaraní incorporation is restricted to non-referential nouns. Example (5) shows that it is not possible in this language to refer anaphorically to an incorporated noun.

(5) Incorporation of a non-referential noun in Paraguayan Guaraní

\[ \text{*A-hova-hei-se pe-mità, pero i-sy he'i} \]
\[ \text{1.ACSBJ-face-wash-DES that-child but 3.INACSBJ-mother say} \]
\[ \text{nda-i-ky'a-i} \]
\[ \text{ha.} \]
\[ \text{NEG-3.INACSBJ-dirty-NEG that} \]
\[ \text{‘I wanted to wash the child’s face but his mother said that it wasn’t dirty.’} \]

(Velázquez-Castillo 1995: 694; Velazquez Castillo 1996: 144)

Paraguayan Guaraní incorporated nouns do thus not correspond to Referential Subacts at IL. Instead, they are part of the Ascriptive Subacts corresponding to the incorporating verbs (see also Smit 2005: 105).9

By contrast, in Panare “incorporation can be used when the incorporated unit refers to a highly referential and specific entity” (Payne 1995: 309). In this language, incorporation has “specific semantic effects which do not include ‘downplaying’ the identity, referentiality or identifiability of an O[bject] argument” (Payne and Payne 2013: 330). Thus, we conclude that incorporated nouns in Panare instantiate Referential Subacts. An example of noun incorporation from Panare is shown in (6).

9 Alternatively, non-referential incorporated nouns like the ones in Paraguayan Guaraní may correspond to independent Ascriptive Subacts. It does not seem possible to decide which of these possibilities is correct. Ascriptive Subacts can be recognized on the basis of the presence of a modifier or operator of approximation (Hengeveld and Mackenzie 2008: 111–112), but for most of the languages we studied we have not been able to verify whether or not non-referential incorporated nouns can combine with such a modifier or operator. Moreover, the potential unavailability of such modifiers and operators could also be due to morphosyntactic restrictions on what can be incorporated rather than on interpersonal ones. Note that incorporated nouns functioning as nominal predicates are an exception to these considerations, as these always instantiate their own Ascriptive Subacts.
Incorporation of a referential noun in Panare
Yu’pêtyaka-ñe kēj kēn.
y-pu-pêtyaka-ñe kēj kēn
3-head-split-NSPEC.TR AN.PROX AN.INVIS
‘He’s gonna split his head.’
(Payne 1995: 301; Payne and Payne 2013: 332)

In Bininj Kun-Wok, both referential and non-referential incorporated nouns can be found. Example (7) shows the incorporation of the noun *murrng* ‘bone’ into the verb *bimbom* ‘paint’. Here, *murrng* is used referentially, i.e. it corresponds to a Referential Subact, as evidenced by the demonstrative *na-mekke*, which appears as modifier of the incorporated noun external to the incorporation construction. In example (8), on the other hand, the incorporated noun *yaw* ‘baby, child’ is non-referential. It is used as a secondary predicate and correspondingly instantiates an Ascriptive Subact.

(7) Incorporation of a referential noun in Bininj Kun-Wok
Nga-murrng-bimbom na-mekke.
1>3-bone-paint.PST.PFV M-DEM
‘I painted those bones.’
(Evans 2003: 235)

(8) Incorporation of a non-referential noun in Bininj Kun-Wok
Birri-yaw-ni.
3.AU-baby/child-sit.PST.IPFV
‘They sat down like children.’
(Evans 2003: 484)

Chapter 4 studies the referential potential and modification possibilities of incorporated nouns in a sample of 21 incorporating languages, showing that 2 languages restrict incorporation to referential nouns, in eight languages only non-referential incorporated nouns occur and 11 languages show both referential and non-referential incorporated nouns. Based on these data, there does therefore not seem to be an implicational relationship between the incorporation of nouns used referentially and those used non-referentially. All possible combinations occur. We can thus formulate a basic setting regarding the pragmatic category of incorporated nouns, where languages belong to one of the three following types:
Incorporation: Constraints on variation

(9) Basic setting regarding incorporated nouns of different interpersonal categories

Incorporation of referential nouns/Incorporation of non-referential nouns/Incorporation of both referential and non-referential nouns

2.4.2 The head of the incorporated noun

In addition to this basic setting, a number of hierarchies concerning the pragmatic characteristics of incorporated nouns seem to emerge from the data. The first of these has to do with the question whether the incorporated noun is a common noun or a proper name. The incorporation of proper names is cross-linguistically rare (Mithun 1984: 864; Borik and Gehrke 2015: 5) and has even been proposed to be impossible (Mardirussian 1975: 386). It appears that the few languages that do allow the incorporation of proper names, such as Eastern Canadian Inuktitut (Johns 2009: 190–191), Kalaallisut (Sadock 1980: 314; see also example [77]), Nivkh (Mattissen 2017: 861) and Ute-Southern Paiute (Givón 2013: 322–323), additionally show the incorporation of common nouns. Thus, in Eastern Canadian Inuktitut, we find both construction (10), with the incorporated common noun *savi* ‘knife’, and construction (11), with the incorporated proper name *Miali*.

(10) Incorporation of a common noun in Eastern Canadian Inuktitut

*savi-siuq-tunga.*

knife-look.for-1SG.PART

‘I am looking for a knife.’

(Johns 2009: 187)

(11) Incorporation of a proper name in Eastern Canadian Inuktitut

*Qallupilluq Miali-tu-niaq-pa?*

Qallupilluq Mary-consume-NEARFUT-3SG.INTERR

‘Is Qallupilluq (a sea monster) going to eat Mary?’

(Johns 2009: 191)

Most other languages, however, limit incorporation to common nouns. For instance, in Mapudungun (Loncon Antileo 2017: 46), Nadèb (Weir 1990: 325), Nuu-chah-nulth (Stonham 2008: 524) and Southern Tiwa (Allen et al. 1984: 301), common nouns may be incorporated, but proper names may not. A possible explanation for the rare occurrence of incorporated proper names could be that languages generally only allow the incorporation of lexemes inserted at RL, while proper names differ from other nouns in appearing at IL (Hengeveld and Mackenzie 2008: 19). In addition, proper names are special in that they are only used for referents that are assumed to be identifiable for the addressee (Hengeveld and Mackenzie 2008: 117), while many languages limit incorporation to non-referential nouns, as discussed in the previous
subsection, or to nouns with referents that are not identifiable for the addressee, as will be discussed in the next subsection.

The hierarchy in (12) captures the data concerning the heads of incorporated nouns observed so far.

(12) Implicational hierarchy of incorporated nouns with different types of interpersonal heads

Incorporation of common nouns $\supset$ Incorporation of proper names

2.4.3 The pragmatic operators applying to the incorporated noun

With respect to pragmatic operators, we consider here the restrictions on noun incorporation that have to do with the identifiability of the referent for the addressee and the identifiability of the referent for the speaker. In languages with referential incorporated nouns, speakers may or may not assume these referential nouns to be identifiable for the addressee. In several languages, such as Chimalapa Zoque (Johnson 2000: 274) and Nuu-chah-nulth (Walde 2004: 52), the referents evoked by referential incorporated nouns are necessarily non-identifiable for the addressee. Other languages, including Kalaallisut (Sadock 1985: 399), Mapudungun (Baker et al. 2005: 174), Mohawk (Baker 1996: 288), Nivkh (Mattissen 2003: 175–176) and Sora (Anderson 2017: 941, fn. 12), do show incorporated nouns with referents that are taken to be identifiable for the addressee, which is often evidenced by the possibility to combine them with demonstratives, as in example (7) from Bininj Kun-Wok above. However, these languages allow the incorporation of nouns with referents that are not identifiable for the addressee as well. In example (13) from Mohawk, for instance, the noun other ‘basket’ is assumed not to be identifiable for the addressee in the first incorporation construction, but in the second incorporation construction it is identifiable for the addressee (Baker 1996: 288).

(13) Incorporation of a non-identifiable and an identifiable incorporated noun in Mohawk

\[
\text{Thetire iska w-ather-a-yi-tah-kwe’ nek tsi Wishe yesterday one N.SBJ-basket-Ø-lie-HAB-PST but PRT Michael i-k-ehr-e’ wa-ha-[a]ther-a-hnu-’.
\]

Ø-1SG.SBJ-think-IPFV FAC-M.SBJ-basket-Ø-buy-PNCT

‘There was a basket (here) yesterday, but I think Michael (basket-)bought it.’

(Baker 1996: 288)

From facts like these we tentatively derive the following implication:
Incorporation: Constraints on variation

(14) Implicational hierarchy of incorporated nouns with different identifiability values

\[ \text{Incorporation of } -id \text{ nouns } \Rightarrow \text{Incorporation of } +id \text{ nouns} \]

Languages may also restrict the incorporation of referential nouns to those with referents that are not identifiable for the speaker, i.e. that are non-specific. Thus, Chimalapa Zoque (Johnson 2000: 274) and Nuu-chah-nulth (Nakayama 2014: 455) do not show the incorporation of nouns that evoke specific referents. By contrast, Kalaallisut (Fortescue 1984: 251, 300), Mohawk (Baker 1988: 79, 1996: 288), Southern Tiwa (Allen et al. 1984: 297) and Washo (Lemieux 2010: 154; Bochnak and Rhomieux 2013: 271) do allow the incorporation of nouns with specific reference. These languages additionally show incorporated nouns with referents that are not identifiable for the speaker. For instance, the incorporated noun qimmi ‘dog’ in the Kalaallisut example in (15) has a non-specific incorporation, while the Kalaallisut noun piili ‘car’ in example (16) has a specific interpretation.

(15) Incorporation of a non-specific noun in Kalaallisut

\[ \text{qimmi-}qar-puq \]
\[ \text{dog-have-3SG.IND} \]

‘He has a dog/dogs/there are dogs.’

(Fortescue 1984: 300)

(16) Incorporation of a specific noun in Kalaallisut

\[ \text{sukka-su-}mik \quad \text{piili-}si-vuq \]
\[ \text{be.fast-INTR,PTCP-INS.SG} \quad \text{car-buy-3SG.IND} \]

‘He bought a (fast) car.’

(Fortescue 1984: 251)

Based on facts like these we preliminarily propose the implication in (17).

(17) Implicational hierarchy of incorporated nouns with different specificity values

\[ \text{Incorporation of } -s \text{ nouns } \Rightarrow \text{Incorporation of } +s \text{ nouns} \]

2.4.4 The pragmatic function of the incorporated noun

Finally, the possible pragmatic functions of incorporated nouns play a role in the IL-ML interface. It has been noted that in many languages, noun incorporation is a backgrounding device (Mithun 1984: 874; Gerdts 1998: 86; Massam 2017). Thus, nouns may be incorporated in order to mark them as having a Background function. Focal nouns, by contrast, are generally not found in incorporation constructions (Baker 1988: 78–79; Gronemeyer 1996: 29; Aikhenvald and Green 1998: 453; Lehmann and Verhoeven 2005: 117; DeClaire et al. 2017: 5, 7).
Mohawk is an example of a language in which incorporated nouns may have a background function but not a focus function (Mithun 1984: 869; Baker 1996: 290; DeClaire et al. 2017: 5–7). More precisely, in this language noun incorporation is obligatory unless either the noun or the verb has a focus function. Thus, example (18b), in which the incorporated noun honwa ‘boat’ has a background function, is grammatical, while example (19b), in which the incorporated noun ‘sereht ‘car’ has a focus function, is not accepted.

(18) Incorporation of a noun with background function in Mohawk

a. Önha wa’ehonwahni:non’
   onhka wa’-e-honw-a-hninon-‘
   who FAC-F.SG-boat-LK-buy-PNCT
   ‘Who bought a boat?’

b. Wá:ri wa’ehonwahni:non’
   Wari wa’-e-honw-a-hninon-‘
   Mary FAC-F.SG-boat-LK-buy-PNCT
   ‘MARY bought a boat.’
   (DeClaire et al. 2017: 4)

(19) Incorporation of a noun with focus function in Mohawk

a. Wahahonwahni:non’ ken ne Sewatis?
   wa-ha-honw-a-hninon-‘  ken ne Sewatis
   FAC-M.SG-boat-LK-buy-PNCT Q PRT John
   ‘Did John buy a boat?’

b. #Iah. Waha’serehtahni:non’
   iah wa-ha- ‘sereht-a-hninon-‘
   no FAC-M.SG-car-LK-buy-PNCT
   ‘No. He bought a car.’
   (DeClaire et al. 2017: 4)

Similarly, in Ket “incorporation [tends] to be used to background an item in discourse”, while a construction without incorporation is “used to topicalize the same item” or “expresses instead a focused, unexpected, or otherwise individuated verb-external object” (Vajda 2017: 910–911).

However, there are also languages in which both backgrounded and focal nouns can be incorporated. In the Kalaallisut example (20a), kaage ‘cake’ is part of the focal part of the message. In (20b) it is picked up again and therefore now part of the background. In (21), the incorporated noun aput ‘snow’ refers to the new topic introduced in this sentence and is therefore focal in nature, just like kaage in (20a).
Incorporation: Constraints on variation

(20) Incorporation of a noun with focus function and incorporation of a noun with background function in Kalaallisut
a. *Ipasaq kaage-liur-pugut.*
   yesterday cake-make-1PL.IND
   ‘Yesterday, we made cake.’

b. *Ullumi kaage-rnias-pugut.*
   today cake-sell-1PL.IND
   ‘Today, we are selling cake.’

(Van Geenhoven 1998: 37)

(21) Incorporation of a noun with topic and focus function in Kalaallisut
(Piututsq was unable to continue)
*Nuna-Ø aput-qar-liir-riir-pug.*
land-ABS.SG snow-have-INGR-already-3SG.IND
   ‘Snow was on the land already.’

Given that we have not encountered languages in which focal nouns can be incorporated while backgrounded ones cannot, we speculate that the hierarchy in (22) correctly describes the distribution of incorporated nouns with Background and Focus function.

(22) Implicational hierarchy of incorporated nouns with Background and Focus function

\[ \text{Incorporation of nouns with Background function} \supset \text{Incorporation of nouns with Focus function} \]

Based on the findings for noun incorporation and Background and Focus function, it could also be expected that other dimensions of information structure, such as the one dividing a discourse act into Comment versus Topic and the one distinguishing Overlap and Contrast, are subject to similar hierarchies, as suggested in (23) and (24).

(23) Implicational hierarchy of incorporated nouns with Comment and Topic function

\[ \text{Incorporation of nouns with Comment function} \supset \text{Incorporation of nouns with Topic function} \]
(24) Implicational hierarchy of incorporated nouns with Overlap and Contrast function

Incorporation of nouns with Overlap function \( \supset \) Incorporation of nouns with Contrast function

However, data concerning pragmatic functions of incorporated nouns are very limited and the definitions of topic and contrast used in different studies vary greatly. For this reason, these expectations could not be tested.

2.5 The RL-ML interface

At RL, too, there are many factors that (co-)determine whether or not noun incorporation is allowed in a language. These include the following:

i. The semantic layer of the incorporated noun: does the noun designate a Property or an Entity?
ii. The semantic function of the incorporated noun: is it an Undergoer, an Actor or something else?
iii. The type of dependent with respect to the incorporating verb: is the incorporated noun an intransitive argument, transitive argument or a modifier?
iv. Alignment system: for verbs with more than one argument, which argument can be incorporated?
v. Relationality: is the incorporated noun relational or non-relational?

2.5.1 The semantic layer of the incorporated noun

Incorporated nouns may either be Property-denoting nouns, i.e. nouns at the RL layer of the Property, here called f-nouns, or Entity-designating nouns, such as nouns at the RL layer of the Individual or other RL layers, here indicated as \( \alpha \)-nouns (Smit 2005: 102–103). These types of incorporated nouns can be differentiated based on their modification possibilities: f-nouns are non-modifiable, while \( \alpha \)-nouns can be modified.\(^\text{10}\) Languages differ in which of these types of nouns they show in incorporation constructions: they may limit incorporation to f-nouns, only allow \( \alpha \)-nouns as incorporated nouns or show both incorporated f-nouns and incorporated \( \alpha \)-nouns.

In Western Frisian all incorporated nouns are f-nouns. The examples in (25) show that it is not possible to modify an incorporated noun in this language by means

\(^{10}\) It is possible for f-nouns to combine with Property modifiers (Hengeveld and Mackenzie 2008: 230–231). However, because this type of modification is highly marginal, we consider f-nouns as non-modifiable here.
Incorporation: Constraints on variation

of plural inflection (25a), determiners (25b), adjectives (25c) or adpositional phrases (25d).\(^{11}\)

(25) Incorporation of an f-noun in Western Frisian

a. *Heit jerappel/*jerappel-s dolt de hiele dei
   father potato/potato-PL digs DEF whole day
   ‘Our father is digging potatoes all day long.’
   (Dijk 1997: 15)

b. *De buorljju sieten bâten te *de/*dy/*sokke wyn-drinken
   DEF neighbours sat outdoors to DEF/DEM/such wine-drink
   (Dijk 1997: 16)

c. *Heit sit te grouwe jerappel-skilen
   father sits to huge potato-peel
   ‘Father is sitting, peeling huge potatoes.’
   (Dijk 1997: 16)

d. *Heit sit te jerappel mei in soad spruten skilen
   father sits to potato with INDF lot sprouts peel
   ‘Father is sitting, peeling potatoes with a lot of sprouts.’
   (Dijk 1997: 16)

In contrast to the Western Frisian incorporated nouns, incorporated nouns in Niuean are always α-nouns. In this language three types of noun incorporation can be recognized, which are called “general”, “existential” and “instrumental” (Seiter 1980, cited in Massam 2001: 167). Incorporated nouns in each of these types are α-nouns, as they may be modified by relative clauses external to the incorporation constructions and/or constitute the head of full incorporated noun phrases (Massam 2001: 169, fn. 18, 175, 178). An example of a Niuean incorporated noun modified by a relative clause is shown in (26).\(^{12}\)

(26) Incorporation of an α-noun in Niuean

Ne fai fale a Sione ne tā e au.
PST have house ABS Sione PST build ABS I
‘Sione has a house that I built.’
(Massam 2001: 175)

---

\(^{11}\) In Western Frisian, the morphosyntactic word status of noun incorporation constructions can be identified on the basis of the verbal infinitive marker te, which usually directly precedes the verbal word but precedes the noun in a noun incorporation construction, as in (25b–d).

\(^{12}\) In Niuean, verbal enclitics follow incorporated nouns (Seiter 1980: 69), thus showing that the incorporated noun and the incorporating verb form a single morphosyntactic word.
Finally, in Bininj Kun-Wok both incorporated f-nouns and incorporated α-nouns are found. Incorporated body-part nouns and incorporated generic nouns, which function semantically as arguments of incorporating verbs, may be modified by adjectives, possessive pronouns, demonstratives, numerals and relative clauses (Evans 2003: 452), as exemplified in (7) above. By contrast, incorporated nouns functioning as secondary predicates, shown in (8), are not modifiable (Evans 2018: p.c.) and can thus be considered f-nouns.

These facts from Western Frisian, Niuean and Bininj Kun-Wok illustrate that an implicational relationship cannot be established between the incorporation possibilities of the two semantic types of nouns (see also Chapter 4). Languages therefore need a basic setting for this parameter, as given in (27).

(27) Basic setting regarding incorporated nouns at different semantic layers

*Incorporation of f-nouns/Incorporation of α-nouns/Incorporation of both f-nouns and α-nouns*

Within the class of α-nouns the ones designating Individuals exhibit in many languages a distinction between those designating animate Entities and those designating inanimate Entities. In these languages, inanimate nouns may be the only type of nouns that can be incorporated (Mithun 1984: 863; Borik and Gehrke 2015: 5) or may “incorporate more readily than animate nouns” (Gerdt 1998: 85; see also Lehmann and Verhoeven 2005: 115; Sadock 2006: 585). This asymmetry between animate and inanimate nouns may be related to the different functions of animate and inanimate nouns in discourse, as animate nouns are typically more central in discourse than inanimate ones, while incorporation often functions to background nouns (Mithun 1984: 863; Gerdt 1998: 85–86).

In Southern Tiwa, incorporation is obligatory for inanimate direct objects, inanimate subjects of intransitive verbs, animate non-human direct objects (unless they are singular and co-occur with an external modifier, in which case incorporation is optional) and plural human direct objects (unless they co-occur with an external modifier, in which case incorporation is optional) (Allen et al. 1984: 293, 295, 296, 299–300). By contrast, human singular direct objects are only optionally incorporated (unless when the subject is third person, in which case the incorporation is obligatory) (Allen et al. 1984: 294) and animate subjects are never incorporated (Allen et al. 1984: 298). Animacy thus influences the possibility or obligation to use an incorporation construction in Southern Tiwa, and the language prefers inanimate incorporated nouns.

Animacy is also relevant for incorporation in Bininj Kun-Wok. This language makes use of three types of productive, semantically transparent noun incorporation: body-part noun incorporation, generic noun incorporation and secondary predicate incorporation (Evans 2003: 325). Although incorporated secondary predicate nouns
Incorporation: Constraints on variation

may be animate and body-part nouns could be considered animate nouns, generic noun incorporation uses a closed set of around 60 nouns (Evans 2003: 332–333), which are almost all inanimate (Evans 2003: 390). This set includes only three human nouns, *daluk* ‘woman’, *bininj* ‘man’ and *beyard* ‘child’ and one other animate noun *bod* ‘bee’ (Evans 2003: 473).

Finally, there are also languages in which animacy does not play a role in incorporation. For instance, in Nuu-chah-nulth, both human entities, other animate entities and inanimate entities can be found in noun incorporation constructions (Stonham 2008: 512).

These facts lead us to tentatively postulate the following implicational hierarchy:

\[(28) \text{Implicational hierarchy of incorporated nouns with different animacy values} \]

\[\text{Incorporation of inanimate nouns} \supset \text{Incorporation of non-human animate nouns} \supset \text{Incorporation of human animate nouns}\]

### 2.5.2 The semantic function of the incorporated noun

There seems to be a general preference for the incorporation of nouns in Undergoer function: languages that allow the incorporation of nouns with other semantic functions always allow the incorporation of Undergoers as well (Mithun 1984: 875; Lehmann and Verhoeven 2005: 118). In addition, it has often been argued in the literature that nouns functioning as Actors cannot be incorporated or are at least very unlikely to be incorporated (Mithun 1984: 863; Gerdts 1998: 87; Massam 2009: 1089; Johns 2017). Nevertheless, a few languages have been shown to allow such incorporation. Based on examples from these languages, we speculate that the incorporation of Actors is not impossible but rather appears at the lowest position of the hierarchy regarding the semantic functions of incorporated nouns.

In Palikúr, incorporation is restricted to Undergoer arguments (Aikhenvald and Green 1998: 451). Example (29) shows the incorporation of the Undergoer argument of a transitive verb, while in example (30) the Undergoer argument of an intransitive verb is incorporated. Note that in the latter example the possessor of the Undergoer argument appears as the subject of the verb. Such constructions with possessors occurring as clausal arguments are also known as external possessor constructions (Payne and Barshi 1999: 3, 6).

---

13 Undergoer arguments are alternatively called patients, objects of transitive verbs or subjects of stative verbs in the sources used here.

14 Actor arguments are alternatively called agents, subjects of active intransitives, subjects of transitives or agentive subjects in the sources used here.
(29) Incorporation of an Undergoer argument into a transitive verb in Palikûr

\[ \text{kuri ig hakis-ota-ne han akiw} \]

now 3.M rub-eye-CONT.NF thus again

‘He continued rubbing his eyes again.’

(Aikhenvald and Green 1998: 452)

(30) Incorporation of an Undergoer argument into an intransitive verb in Palikûr

\[ \text{eg barew-kug} \]

3.F clean-foot

‘She is clean-footed.’ (i.e. ‘Her feet are clean.’)

(Aikhenvald and Green 1998: 452)

In Mapudungun, Undergoers, such as pullku ‘wine’ in (31), and Locative modifiers of intransitive verbs, such as kawellu ‘horse’ in (32), can be incorporated, while nouns functioning as Actor arguments and other modifiers cannot occur as incorporated nouns (Baker et al. 2005: 171; Zúñiga 2017: 703–705).

(31) Incorporation of an Undergoer argument in Mapudungun

\[ \text{Juan ngilla-pullku-la-y.} \quad \text{ñi che ngilla-fi-n.} \]

Juan buy-wine-NEG-3SG.SBJ.IND I buy-1SG.SBJ.IND

‘Juan didn’t buy the wine. I bought it.’

(Baker et al. 2005: 146)

(32) Incorporation of a Locative modifier into an intransitive verb in Mapudungun

\[ \text{püra-kawellu-ascend-horse} \]

‘mount a horse’

(Zúñiga 2017: 705)

Yucatec Maya allows the incorporation of Undergoer arguments, Instrument modifiers and Locative modifiers (Lehmann and Verhoeven 2005: 149). Che’ ‘tree’ in example (1b) above, k’ab ‘hand’ in example (33) and pach ‘back’ in example (34) illustrate the incorporation of nouns with these semantic functions.

(33) Incorporation of an Instrument modifier in Yucatec Maya

\[ \text{in lom-k’ab-t-ik-ech} \]

3.SBJ poke-hand/finger-TR-INCOMPL-2SG.ABS

‘I poke you with my finger.’

(Sullivan 1984: 151; Lehmann and Verhoeven 2005: 161)

\[ \text{15 Locative modifiers are alternatively called location or ground in the sources used here.} \]
Incorporation: Constraints on variation

(34) Incorporation of a Locative modifier in Yucatec Maya

táan in kuch-pach-t-ik in nal
PROG 1SG.SBJ load-back-TR-INCOMPL 1SG.POSS corn
‘I am carrying my corn on my back (multiple trips).’
(Bricker et al. 1998, cited in Lehmann and Verhoeven 2005: 166)

In Movima incorporated Undergoer arguments (35), Instrument modifiers (36) and Locative modifiers (37)–(38) are found as well (Haude 2006: 368, 383, 384).

(35) Incorporation of an Undergoer argument in Movima

ij wul-a-saniya (ni-kis saniya)
2.INTR sow-DR-melon OBL-ART.PL.AB melon
‘You sow melon.’
(Haude 2006: 368)

(36) Incorporation of an Instrument modifier in Movima

jayna nis-na=is is bari=is di’ jayna
DISC wipe.clean-DR=PL.AB ART.PL foot=PL.AB REL DISC
pay’-but-el n-is bereya:-buñ
smear-BR.mud-APPL OBL-ART.PL tar-BR.mud
‘Then they wiped clean their feet (of the macaws), which were smeared with tar.’
(Haude 2006: 385)

(37) Incorporation of a Locative modifier in Movima

am-a-sil-a=is os lume’ n-os si-l-kwa
enter-DR.BR.hole-LV=PL.AB ART.N.PST agouti OBL-ART.N.PST BR.hole-ABSS
‘They (the dogs) made the agouti go into the hole.’
(Haude 2006: 384)

(38) Incorporation of a Locative modifier in Movima

kas isko-ni-wa rey ja’ ena’ kamay-chorada-net
NEG 3PL.AB-VBLZ-NMLZ again just DUR.STD yell-street-APPL
‘Those were not just yelling in the street.’
(Haude 2006: 384)

South Slavey shows incorporated nouns with various semantic functions. Example (39) shows the incorporation of the Undergoer argument too ‘night’. Example (40) demonstrates that in this language Locative modifiers can be incorporated, whereas (41) exemplifies Instrument modifier incorporation. In addition, South Slavey Actor arguments can be incorporated, as in example (42).
Incorporation of an Undergoer argument in South Slavey

(39) **too-go-d-i-tl’e**
night-area-QU-QU-be.dark
‘It (night) is dark.’
(Rice 1989: 655, cited in Rice 2008: 386)

Incorporation of a Locative modifier in South Slavey

(40) **k’e-ke-e-h-dzoh**
around-foot-ASP-1SG.SBJ-slice
‘I skated, slid on feet.’

Incorporation of an Instrument modifier in South Slavey

(41) **tse na-xee-ye’a**
wood back-pack-3.DOBJ-handle.default.object
‘S/he is packing wood back.’ (i.e. ‘S/he is handling wood by means of pack.’)

Incorporation of an Actor argument in South Slavey

(42) **be-se-we-h-see**
sleep-1SG.DOBJ-QU-CAUS-kill.SG.OBJ
‘I am sleepy.’ (i.e. ‘Sleep overcomes me.’)
(Rice 1989: 663, cited in Rice 2008: 387)

Finally, Sora shows the incorporation of Undergoer arguments of transitive verbs (Anderson 2017: 937), such as **dʒaʔ ‘snake’** in (43). Locative modifiers (44) and Instrument modifiers (45) can be incorporated too. In addition, Actor arguments of transitive verbs can be incorporated in this language (Anderson 2017: 945–946), as shown in (46).

Incorporation of an Undergoer argument in Sora

(43) **pen nam-dʒaʔ-[t]i-n-aj**
catch-snake-NPST-INTR-1.SBJ
‘I am catching a snake.’
(Anderson 2017: 939)

Incorporation of a Locative modifier in Sora

(44) **lem-dʒeŋ-te-ben-dʒi**
bow-foot-NPST-2PL.OBJ-3PL.SBJ
‘They bow to your feet.’
(Anderson 2017: 937)
Incorporation: Constraints on variation

(45) Incorporation of an Instrument modifier in Sora
\[ \text{pen a}^{-3}\text{dyi}^{-m}^{-dji}=\text{a}^{-3}\text{lo}^{-g}\quad \text{aba}^{-3}\text{s}^{-i}^{-t}^{-ai} \]
\[ \text{I 3-foot-NSFX-PL=OBJ wash-hand-NPST-1.SBJ} \]
‘I am washing their feet by hand.’
(Anderson 2017: 937)

(46) Incorporation of an Actor argument in Sora
\[ \text{pem}^{-3}\text{bud}^{-t}^{-am} \]
\[ \text{seize-bear-NPST-2.OBJ} \]
‘The bear will seize you.’
(Anderson 2017: 946)

Table 1 summarizes the possible semantic functions of incorporated nouns in these different languages.

| Table 1. Semantic functions of incorporated nouns in six languages. |
|-----------------------|-----------------|-----------------|-----------------|
| Language | Undergoer | Other semantic functions (Locative, Instrument) | Actor |
| Palikūr | + | − | − |
| Mapudungun | + | + | − |
| Yucatec Maya | + | + | − |
| Movima | + | + | − |
| South Slavey | + | + | + |
| Sora | + | + | + |

Table 1 reflects the hierarchy given in (47).

(47) Implicational hierarchy of incorporated nouns with different semantic functions
\[ \text{Incorporation of Undergoer} \supset \text{Incorporation of other semantic functions} \supset \text{Incorporation of Actor} \]

2.5.3 Type of dependent with respect to the incorporating verb
Incorporated nouns and incorporating verbs are in a dependency relation of the form head-modifier or predicate-argument. Typically, the incorporated noun is either a modifier or an argument of the incorporating verb (Mithun 2000: 917; Haugen 2015: 414–415). In addition, incorporated nouns may function as nominal predicates in constructions in which the incorporating verbs function as semi-copula (Hengeveld 1992: 34–39), as in the Ket example in (i).

(i) Incorporation of a noun functioning as a nominal predicate in Ket
\[ \text{tab}^{-3}\text{-an}^{-t}^{-o}^{-n}^{-aq} \]
\[ \text{dog.PL-3PL=AN.SBJ-TC-PST-become} \]
‘They turned into dogs.’
(Vajda 2017: 918)
Noun incorporation in Functional Discourse Grammar

occurs in languages that also show the incorporation of nominal arguments (Mithun 1984: 875; Aikhenvald 2007: 19). More specifically, it seems that all incorporating languages allow the incorporation of transitive (Undergoer) arguments, that languages may additionally incorporate intransitive (Undergoer) arguments and that languages that show both incorporated transitive and intransitive arguments may optionally also allow incorporated modifiers (Mithun 1984: 875; Haspelmath 2018: 318, fn. 9).17

In Kalamang, incorporation appears to be restricted to transitive arguments (Visser 2019: p.c.). An example of an incorporation construction in Kalamang is shown in (48).18

(48) Incorporation of a transitive argument in Kalamang

ma mua’waruo
ma muap-paruo
3SG food-make
‘She is cooking.’
(Visser et al. 2019)

In Nadëb, arguments can be incorporated into transitive and intransitive verbs, as shown in example (49) and (50) respectively.19

(49) Incorporation of a transitive argument in Nadëb

ta=ti ɨɨ i-ii
3SG=food ASP-fish
‘He is fishing his (i.e. someone else’s) food.’
(Weir 1990: 331)

(50) Incorporation of an intransitive argument in Nadëb

ɨ̃ɨ h=t ʉg da-tés
1SG=toothTH-hurt
‘I have toothache.’ (lit. ‘I tooth-hurt.’)
(Weir 1990: 323)

In such cases, the incorporating verb may be considered an operator of the nominal predicate, which then functions as the head of the verbal operator.

17 Some languages also show the incorporation of arguments into ditransitive verbs. This type of incorporation is addressed in the discussion of morphosyntactic alignment in Section 2.5.4.

18 In the isolating language Kalamang, the absence of the accusative marker on a noun that directly precedes a verb shows that it is incorporated (Visser 2019: p.c.).

19 In Nadëb, incorporated nouns, which precede the stem of the incorporating verb, follow the pronouns that appear as verbal proclitics, such as ta= in (49) (Weir 1990: 331). An incorporated noun and an incorporating verb thus form a single morphosyntactic word together. However, the noun and verb remain independent phonological words (see Section 2.6.2).
Incorporation: Constraints on variation

These types of noun incorporation are the only possible types in Nadëb (Weir 1990: 325), which entails that modifier incorporation is not found in this language.

In Hokkaido Ainu, incorporation is also limited to nouns functioning as transitive or intransitive arguments. In this language, four types of incorporation are recognized: transitive Undergoer incorporation, intransitive argument incorporation in which the argument is a natural phenomenon noun, intransitive argument incorporation in which the argument is a body-part noun in its possessive form and transitive Actor incorporation in which the incorporated Actor is a (super)natural phenomenon or insect noun (Bugaeva 2017: 897).

By contrast, in Chukchi both incorporated transitive arguments, incorporated intransitive arguments and incorporated modifiers are found. Firstly, in example (51), the incorporated noun wala ‘knife’ functions as the Undergoer argument of the transitive incorporating verb mna ‘sharpen’.

(51) Incorporation of a transitive argument in Chukchi

*Mə-wala-mna-rkən
1PL.S.INT-knife-sharpen-1PL.S
‘Let us sharpen the knives.’

Secondly, the incorporated noun ətlə ‘mother’ in example (52) is the intransitive Undergoer argument of the intransitive incorporating verb wʔə ‘die’.

(52) Incorporation of an intransitive argument in Chukchi

atlag-ən ətlə-wʔə-gəe
father-ABS.SG mother-die-3SG.S
‘Father’s mother died (on him).’

Thirdly, in example (53), the incorporated noun amnə ‘fish’ is a modifier that in a clause without incorporation would take so-called instrumental case-marking (Skorik 1948: 72, cited in Spencer 1995: 457).

(53) Incorporation of a modifier (adjunct) in Chukchi

[…] amnə-tke-rkən
fish-smell-3SG.S
‘[…] (it) smells of fish.’
(Skorik 1948: 72, cited in Spencer 1995: 457)

Like Chukchi, Ket shows the incorporation of arguments that normally function as transitive Undergoers, exemplified in (54), the incorporation of arguments
of intransitive verbs, as in example (55), as well as the incorporation of modifiers, shown in example (56).

(54) Incorporation of a transitive argument in Ket  
    \text{da=nan-si-bed}  
    3.F.SBJ=bread-PRS-make  
    ‘She is making bread.’  
    (Vajda 2017: 912)

(55) Incorporation of an intransitive argument in Ket  
    \text{ul-a-ta}  
    rain-PRS-falls  
    ‘It rains.’  
    (Vajda 2017: 921)

(56) Incorporation of a modifier (adjunct) in Ket  
    \text{assano ke’d tib d=sal-a-t-a-kit}  
    hunting person dog 3.SBJ=tobacco-3SG.M.OBJ-TC-PRS-rub  
    ‘The hunter “tobaccoed” the dog (to rid it of fleas).’  
    (Vajda 2017: 916)

Finally, Western Frisian also shows incorporated arguments and modifiers (Dijk 1997: 94, 136, 162). Note, however, that the incorporation of intransitive incorporated arguments is limited to a few isolated cases in sentences with expletive subjects, such as the one in example (57) (Dijk 1997: 162).

(57) Incorporation of an intransitive argument in Western Frisian  
    \text{It begjint te sne-wiskjen}  
    It begins to snow-fly  
    ‘The snow begins to fly.’  
    (Dijk 1997: 162)

Table 2 shows the types of dependents found in incorporation constructions in the different languages.
Incorporation: Constraints on variation

Table 2. Types of incorporated dependents in five languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>Transitive argument</th>
<th>Intransitive argument</th>
<th>Modifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalamang</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nadëb</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Hokkaido Ainu</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Chukchi</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ket</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Western Frisian</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

The implicational hierarchy that may be derived from Table 2 is given in (58).

(58) Implicational hierarchy of incorporated nouns functioning as semantic arguments and modifiers

\[
\text{Incorporation of transitive arguments} \supset \text{Incorporation of intransitive arguments} \supset \text{Incorporation of modifiers (adjuncts)}
\]

Interestingly, in cases in which the only argument of an intransitive verb is incorporated, the incorporation construction, i.e. a single Morphosyntactic Word at ML, may correspond to a complete Configurational Property at RL. In such cases, noun incorporation creates a transparent match between a single unit at RL and a single unit at ML. In this respect, there is a contrast between the incorporation of arguments into intransitive verbs on the one hand and the incorporation of arguments into transitive verbs and the incorporation of modifiers on the other hand. Note finally that a language’s ability to incorporate full Configurational Properties also adds to this language’s degree of polysynthesis, in that a relatively high unit at RL corresponds to a single Word at ML (see Genee 2018: 257–260).

2.5.4 Morphosyntactic alignment

In FDG, the selection of arguments with specific semantic functions in certain privileged syntactic positions is handled by the interface between RL and ML as well. For instance, at the Clause layer, the choice of arguments with certain semantic functions to fulfill the role of Subject and Object is handled by this interface. A language shows an accusative or ergative alignment if there is neutralization between the argument of an intransitive verb and the Actor or Undergoer argument of a transitive verb. Furthermore, on the basis of neutralization between the Undergoer argument of a transitive verb and the Undergoer or Locative argument of a ditransitive verb, languages can be characterized as either indirective or secundative.

At the Morphosyntactic Word layer, similarly the choice of arguments that can be incorporated is an issue of alignment. The alignment system of a language for noun incorporation may simply depend on interpersonal or representational characteristics of the arguments, but may also be of the morphosyntactic type. The following examples demonstrate that the different morphosyntactic alignment systems
distinguished for the Clause layer are found at the Morphosyntactic Word layer, i.e. in noun incorporation, as well.

In Bininj Kun-Wok the only argument of an intransitive (59)–(60) and the Undergoer argument of a transitive verb (61) can be incorporated, while Actor arguments of transitive verbs cannot (Evans 2003: 455, 468–471).

(59) Incorporation of the only argument of an intransitive verb in Bininj Kun-Wok
    Ga-warnde-djabdi.
    3-rock-stand.up.straight.NPST
    ‘There is a rock standing up straight.’
    (Evans 2003: 451)

(60) Incorporation of the only argument of an intransitive verb in Bininj Kun-Wok
    Ga-yau-dolga-n.
    3-baby/child-get.up-NPST
    ‘The baby (kangaroo) gets out of its pouch.’
    (Evans 2003: 468)

(61) Incorporation of the Undergoer argument of a transitive verb in Bininj Kun-Wok
    Al-ekge al-gohbanj ba-gurlah-bimbu-ni.
    F-DEM II-old.person 3>3.PST-skin-paint-PST.IPfv
    ‘That old lady used to paint buffalo hides.’
    (Evans 2003: 451)

Note that in the case of intransitive verbs, both Actors and Undergoers can be incorporated (Evans 2003: 468), which shows that the alignment system for incorporation in Bininj Kun-Wok cannot be explained solely on the basis of semantic functions. Thus, the alignment system for noun incorporation in this language is not representational in nature. Rather, the language has a morphosyntactic alignment system of the ergative type (see also Hengeveld and Mackenzie 2008: 408).

Kalamang, on the other hand, has an accusative system, as it allows the incorporation of transitive Undergoers, as exemplified in (48) in Section 2.5.3, but not of intransitive arguments (Visser 2019: p.c.).

With respect to ditransitive verbs, noun incorporation is typically limited to Undergoer arguments, such that most languages have an indirective alignment system (Malchukov et al. 2010: 42). Thus, Southern Tiwa incorporates the Undergoer arguments of both transitive (62) and ditransitive (63) verbs (Allen et al. 1984: 293, 303; see also Hengeveld and Mackenzie 2008: 408–409).
Incorporation: Constraints on variation

(62) Incorporation of the Undergoer argument of a transitive verb in Southern Tiwa
$Ti$-$seuan$-$mũ$-$ban.
1SG$>$SG-man$-$see$-$PST
‘I saw the/a man.’
(Allen et al. 1984: 294)

(63) Incorporation of the Undergoer argument of a ditransitive verb in Southern Tiwa
$Ti$-$’u’u$-$wia$-$ban $ɨ$-$’ay.$
1SG$>$SG-baby$-$give$-$PST 2SG-ALL
‘I gave the baby to you.’
(Allen et al. 1984: 303)

Nivkh, on the other hand, has secundative alignment in noun incorporation, as
the Locative of a ditransitive verb (64) can be incorporated, just like the Undergoer of
a transitive verb, as shown in (65) (Mattissen 2003: 137, 140; see also Hengeveld and

(64) Incorporation of the Locative argument of a ditransitive verb in Nivkh
$objetd$ɨ $k’$e $atak$-$asqam$-$d
bay$.$watcher $net$ grandfather$-$take.away$-$IND
‘The bay watcher took the net away from grandfather.’
(Otaina 1978: 34, cited in Mattissen 2003: 142)

(65) Incorporation of the Undergoer argument of a transitive verb in Nivkh
$atak $k’$e$-$seu$-$d
grandfather $net$-dry$-$IND
‘Grandfather dried the net.’
(Otaina 1978: 34, cited in Mattissen 2003: 137)

Finally, in Hokkaido Ainu, ditransitive verbs may sometimes incorporate both
their Undergoer argument and their Locative argument at the same time (Bugaeva
2017: 899), as in (66).

(66) Incorporation of the Undergoer argument and the Locative argument of a
ditransitive verb in Ainu
$cep$-$ya$-$o$-$kuta$=$an
fish$-$shore$-$APPL$-$throw$=$INDF.$
‘I threw the fish (he caught) onto the shore.’
Based on examples like (66), we conclude that languages may also have a neutral alignment system for noun incorporation.

The alignment system of a language in its incorporation strategies is not predictable from other properties and therefore has to be stipulated as a basic property of the language, as in (67).

(67) Basic settings regarding alignment systems in the context of noun incorporation

**Accusative/Ergative/Neutral**

**Indirective/Secundative/Neutral**

### 2.5.5 Relationality

In many languages, relational nouns or, more specifically, body-part nouns are either the only type of nouns that can be incorporated or the type of nouns that is incorporated most frequently or easily (Mithun 1986b: 383; Aikhenvald 2007: 20; Massam 2009: 1090). Moreover, in some languages incorporation is limited to constructions in which a body-part noun or another relational noun is incorporated and its (alienable) possessor is expressed as an argument of the incorporating verb, i.e. as an external possessor.

In Palikūr, incorporation is limited to body-part nouns (Aikhenvald and Green 1998: 451; Aikhenvald 2007: 20). These nouns are obligatory possessed, and when they are incorporated, their possessor is generally expressed as the direct object, in the case of a transitive incorporating verb, or as the subject, in the case of an intransitive incorporating verb (Aikhenvald and Green 1998: 451–452). An example of an incorporated body-part noun with its possessor expressed as direct object is shown in (68), in which the noun *ot ‘eye’ is incorporated and its 3rd person singular possessor is expressed as the verbal suffix -gi. The incorporation of a body-part noun and the expression of its possessor as subject was exemplified in (30) above.

(68) Incorporation of a relational noun in Palikūr

\[
\text{ig-kis } hapis patuk-ot-bet-h-e-gi
\]


‘They shot his eyes out.’ (lit. ‘They eye-shot-him.’)

(Aikhenvald and Green 1998: 452)

In contrast to Palikūr, Yucatec Maya does not restrict incorporation to body-part nouns or relational nouns. This language shows both the incorporation of body-part nouns, such as *k’ab ‘hand’ and *pach ‘back’ in example (33) and (34) above, and the incorporation of non-body-part nouns, such as *che ‘tree’ in example (1b).

Based on these facts we tentatively suggest the hierarchy given in (69).
2.6 The ML-PL interface

PL receives its input from ML. It is here that it is determined how the incorporated noun is realized segmentally and prosodically. We therefore consider here the following issues:

i. Type of head: is it suppletive or non-suppletive?
ii. The phonological layer of the incorporated noun: is it a separate Phonological Word (Pw) or is it part of the verbal Pw?

In the area of the interface between ML and PL no hierarchies have been proposed that would capture the cross-linguistic constraints on incorporation. Rather, it seems that languages use two basic settings in the interaction between these two levels.

2.6.1 The head of the incorporated noun

In some languages, (some) nouns take suppletive or phonologically alternate forms when they are incorporated, while in other languages incorporated nouns have the same form as unincorporated ones (Mithun 1984: 876; Aikhenvald 2007: 13; Caballero et al. 2008: 387–388). In Sora incorporated nouns have special forms, called “combining forms”, which are monosyllabic or mono-moraic counterparts of the “full forms” that are used in contexts without incorporation (Anderson 2007: 175). The full forms typically show some similarity to the combined forms in that the full forms often appear to be derived from the corresponding combining forms by either reduplication, prefixation, suffixation or compounding (Anderson 2007: 175). For instance, the noun meaning ‘banana’ has the full form kante and the combining form -te, as shown in (70).

(70) Incorporation of a suppletive noun in Sora
   a. jen kante-n  dʒum-i-ai
      I banana-NSFX eat-NPST-1.SBJ
      ‘I am eating a banana.’
   b. jen dʒum-te-li-n-ai
      I eat-banana-NPST-INTR-1.SBJ
      ‘I am eating a banana.’
      (Anderson 2017: 939)

Incorporated body-part nouns in Palikûr either have the same form as unincorporated body-part nouns or alternate forms that are clearly related to the
Unincorporated body-part nouns (Aikhenvald and Green 1998: 451). The set of body-part nouns that can be incorporated is presented in Table 3, in which both the independent and incorporated forms are included.

**Table 3.** Forms of unincorporated body-part nouns, body-part nouns incorporated into stative verbs and body-part nouns incorporated into transitive verbs in Palikúr (Aikhenvald and Green 1998: 451).

<table>
<thead>
<tr>
<th>Form of unincorporated noun</th>
<th>Form of noun incorporated into a stative verb</th>
<th>Form of noun incorporated into a transitive verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>duk 'chest'</td>
<td>-duk</td>
<td>-duka</td>
</tr>
<tr>
<td>kugku 'foot'</td>
<td>-kug</td>
<td>-kuga</td>
</tr>
<tr>
<td>wak 'hand'</td>
<td>-ok</td>
<td>-oka</td>
</tr>
<tr>
<td>tew 'head'</td>
<td>-tiw</td>
<td>-tew</td>
</tr>
<tr>
<td>arvak 'eye'</td>
<td>-ot</td>
<td>-(h)ott(a)</td>
</tr>
<tr>
<td>biy 'mouth'</td>
<td>-bi</td>
<td>-biya</td>
</tr>
<tr>
<td>tip 'top (lid)'</td>
<td>-tip</td>
<td>-tipa</td>
</tr>
</tbody>
</table>

Finally, in Mapudungun incorporated and unincorporated nouns have the same form, as shown by the nouns wün 'snout' and waka 'cow' in example (71) and (72) respectively.

(71) Incorporation of a non-suppletive noun in Mapudungun

Puff pi nga ñi wün ngürü,
Paff say.3SG.SBJ 3.POSS snout fox
wichaf-wün-tu-y [...].
become.big-snout-RE-3SG.SBJ.IND
‘The fox said “paff!” with his snout, (and) his snout became big again […]’

(72) Incorporation of a non-suppletive noun in Mapudungun

a. Ñi choa kintu-le-y ta.chi pu waka.
   my father seek-PROG-3SG.SBJ.IND the COLL cow
   ‘My father is looking for the cows.’

b. Ñi choa kintu-waka-le-y.
   my father seek-cow-PROG-3SG.SBJ.IND
   ‘My father is looking for the cows.’

Whether or not an incorporating language displays suppletive forms cannot be predicted from other properties of the language. The grammar therefore needs a basic setting as in (73).
Basic setting regarding incorporated nouns with different phonological heads

(73) *Incorporated nouns have suppletive forms*/Incorporated nouns have non-suppletive forms

2.6.2 The phonological layer of the incorporated noun

Incorporation constructions may or may not form single Phonological Words (Mithun 1984: 849; Aikhenvald 2007: 14–15; Caballero et al. 2008: 385–386). In some languages there is clear evidence for the status of incorporation structures as Phonological Words. For instance, in Chukchi the vowel harmony rules that operate in phonological words are also at work in incorporation constructions (Mithun 1984: 875; Spencer 1995: 445), as shown in example (74), part of which repeats example (51).

(74) Incorporation of a noun that becomes part of the verbal phonological word in Chukchi

a. *Wala-t ma-mne-rkənet*
   knife-ABS.PL 1PL.A..INT-sharpen-3PL.P
   ‘Let us sharpen the knives.’

b. *Me-wala-mna-rkən*
   1PL.S..INT-knife-sharpen-1PL.S

The recessive vowel e in the verbal stem *mne* ‘sharpen’ changes into the dominant vowel /a/ under influence of the dominant a vowels in the incorporated noun in (74b).

In Cayuga, an incorporated noun and its incorporating verb also form a single phonological word. In this language, phonological words have stress on their fourth syllable, and this pattern also holds for incorporation constructions (Mithun 1994, cited in Aikhenvald 2007: 14).

In other languages, however, incorporation constructions do not form single Phonological Words, even though they constitute Morphosyntactic Words. In Yimas, for instance, incorporated nouns and their incorporating verbs may both carry stress like independent phonological words (Foley 1991: 84). Thus, in example (75), both the incorporated deverbal noun /wacakm/ and the verb including the stem /ti/ carry phonological word stress.

(75) Incorporation of a noun that remains a separate phonological word in Yimas

*mamam p-na-waca-k-m-ti-n*

sore.VILSG VII.SG.S-DEF-small-IRR-VIL.SG-become-PRS

‘The sore is getting smaller.’

(Foley 1991: 83)
In addition, the form of the class and number agreement marker on the incorporated noun /wacakm/ shows that this incorporated noun is an independent phonological word. The marker takes the form /m/, which is the allomorph that is used word-finally, rather than the form /mp/, which is the allomorph that normally occurs in word-medial position (Foley 1991: 84).

In Nadêb, the position of verbal clitics shows that an incorporated noun and an incorporating verb form a single morphosyntactic word (Weir 1990: 330–331). Nevertheless, just like in Yimas, the noun and the verb remain independent phonological words in terms of stress placement (Weir 1990: 323, 330–331). In example (49) above, for instance, the verbal proclitic ta appears in front of the incorporated noun, thus showing that the noun ti ‘food’ is part of the verb with the stem ti ‘fish’ morphosyntactically. At the same time, nouns and verbs in incorporation constructions are stressed independently and can therefore be considered independent phonological words.

Whether or not incorporated nouns in a particular language form separate Phonological Words cannot be predicted from other properties of the language. It therefore has to be specified as a basic setting, as given in (76), in the grammar.

(76) Basic setting regarding incorporated nouns at different phonological layers

Incorporated nouns as separate Pw / Incorporated nouns as part of the verbal Pw

2.7 A worked example

2.7.1 Introduction

One complete set of interface conditions for noun incorporation can be exemplified for Kalaallisut on the basis of the constructions in example (77)–(81). In order to show how these interface conditions are dealt with in FDG, we provide the underlying representations at the four levels of analysis in FDG for these examples, which will serve as a point of reference for the ensuing discussion. Our representations at the Phonological Level are tentative and based on Arnhold (2014). Arnhold (2014: 221) assumes that for Kalaallisut the mora, the phonological word and the intonation phrase are the relevant prosodic units. We will only consider the latter two. She furthermore argues that generally the phonological word coincides with the morphological and syntactic word, something we will assume below as well. Intonation contours are not indicated, as these are not relevant to our concerns here.
Incorporation: Constraints on variation

(77) Noun incorporation in Kalaallisut

\textit{pinnir-su-nik} \quad \textit{pani-qar-puq}

be.beautiful-inTR.PTCP-INS.PL \quad daughter-have-3SG.IND

‘He has beautiful daughters.’

(Kristoffersen 1992: 154)

IL: \quad (A_{i}: [(F_{i}: \text{DECL.} (F_{i})) (P_{i})_{3} (P_{i})_{3} (C_{i}: [(T_{i}) (R_{i}) (-\text{id} +s R_{i}: (T_{i}) (R_{i}))_{\text{INPFOC}}] (C_{i}))] (A_{i}))

RL: \quad (p_{i}: (e_{i}: (f_{i}: ([f_{i}: qar (f_{i})] (x_{i})_{3} (m x_{i}: (f_{i}: \text{panik} (f_{i})] (x_{i})_{3}: (f_{i}: \text{pinnir} (f_{i})) (x_{i})_{3} (f_{i})): (f_{i})) (e_{i})) (ep_{i})) (p_{i}))

ML: \quad ([C_{i}: [(Np_{i}: [Nw_{i}: [Ns_{i}: \text{pinnir} (Ns_{i})] (Vp_{i}: [Vw_{i}: [Nw_{i}: [Ns_{i}: \text{panik} (Ns_{i})] (Vr_{i}: qar (Vr_{i})] (Aff_{i}: \text{voq} (Aff_{i}))] (Vw_{i}))] (Vp_{i})) (C_{i}))]

PL: \quad (IP_{i}: [(PW_{i}: /\text{pinnir}\text{-su-nik}/ (PW_{i})) (PW_{i}: /\text{pani}\text{-qar-puq}/ (PW_{i}))]) (IP_{i}))

(78) Noun incorporation in Kalaallisut

\textit{Nuuk-liar-puq}

Godthaab-go.to-3SG.IND

‘He went to Godthaab.’

(Sadock 1980: 314)

IL: \quad (A_{i}: [(F_{i}: \text{DECL.} (F_{i})) (P_{i})_{3} (P_{i})_{3} (C_{i}: [(T_{i}) (R_{i}) (+\text{id} +s R_{i}: \text{Nuuk} (R_{i}))] (C_{i}))] (A_{i}))

RL: \quad (p_{i}: (e_{i}: (f_{i}: ([f_{i}: liar (f_{i})] (x_{i})_{3} (m x_{i}: [f_{i}: \text{panik} (f_{i})] (x_{i})_{3}: (f_{i}: \text{pinnir} (f_{i})) (x_{i})_{3} (f_{i})): (f_{i})) (e_{i})) (ep_{i})) (p_{i}))

ML: \quad ([C_{i}: [(Vp_{i}: [Vw_{i}: [Nw_{i}: \text{Nuuk} (Nw_{i})] (Vr_{i}: liar (Vr_{i})] (Aff_{i}: \text{voq} (Aff_{i}))] (Vw_{i}))] (Vp_{i})) (C_{i}))

PL: \quad (IP_{i}: [(PW_{i}: /\text{nuuk}\text{-liar-puq}/ (PW_{i})) (PW_{i})]) (IP_{i}))

(79) Noun incorporation in Kalaallisut

\textit{(*utuqqar-mik)} \quad \textit{palasi-rpalup-puq} \quad \textit{(*utuqqaq-Θ)}

old.one-INS.SG \quad priest-be.like-3SG.IND \quad old.one-ABS.SG

‘He is like an old priest.’

(Kristoffersen 1992: 154)

IL: \quad (A_{i}: [(F_{i}: \text{DECL.} (F_{i})) (P_{i})_{3} (P_{i})_{3} (C_{i}: [(T_{i}) (R_{i})] (C_{i}))] (A_{i}))

RL: \quad (p_{i}: (e_{i}: (sim [f_{i}: [f_{i}: \text{palasi} (f_{i})] (x_{i})_{3} (f_{i})): (f_{i})) (e_{i})) (ep_{i})) (p_{i}))

ML: \quad ([C_{i}: [(Vp_{i}: [Vw_{i}: [Ns_{i}: \text{palasi} (Ns_{i})] (Vr_{i}: (r)palug (Vr_{i})] (Aff_{i}: \text{voq} (Aff_{i}))] (Vw_{i}))] (Vp_{i})) (C_{i}))

PL: \quad (IP_{i}: [(PW_{i}: /\text{palasi}\text{-rpalup-puq}/ (PW_{i})) (IP_{i})])
2.7.2 The IL-ML interface

Starting with the IL-ML interface, we observe that in Kalaallisut both referential nouns, such as panik ‘daughter’ ((Ri) in [77]) and Nuuk ‘Godthaab’ ((Ri) in [78]), and non-referential nouns, like the predicatively used noun palasi ‘priest’ ((Ti) in [79]), can be incorporated. Incorporated referential nouns are usually common nouns, but referential proper names are also found in incorporation constructions in Kalaallisut, as shown by example (78), where Nuuk ‘Godthaab’ is the direct head of (Ri). The language also shows several possibilities with respect to the pragmatic operators of referential incorporated nouns. Firstly, the noun panik ‘daughter’ in (77) “cannot be understood as definite” (Kristoffersen 1992: 156), hence the operator −id on (Ri), whereas the noun Nuuk ‘Godthaab’ in (78), being a proper name, has a referent that is presented as identifiable for the addressee (Sadock 1980: 314), hence the operator +id R on (Ri).
Incorporation: Constraints on variation

+id on \((R_i)\). Secondly, although *panik* ‘daughter’ in \((77)\) “can refer to [a] specific [entity]” (Kristoffersen 1992: 156), incorporated nouns in Kalaallisut can also refer to non-specific entities (Fortescue 1984: 300; see also example [15] above). In addition, while non-referential incorporated nouns like *palasi* ‘priest’ in \((79)\) lack a pragmatic function, referential incorporated nouns in Kalaallisut, as illustrated in Section 2.4.4, may either be focal or backgrounded.

The basic settings and position on the hierarchies for the IL-ML interface for noun incorporation in Kalaallisut are summarized in \((82)\)–\((86)\).

\((82)\) Basic setting regarding incorporated nouns of different interpersonal categories

\[ \text{Incorporation of referential nouns/Incorporation of non-referential nouns/Incorporation of both referential and non-referential nouns} \]

\((83)\) Implicational hierarchy of incorporated nouns with different types of interpersonal heads

\[ \text{Incorporation of common nouns } \supseteq \text{Incorporation of proper names} \]

\((84)\) Implicational hierarchy of incorporated nouns with different identifiability values

\[ \text{Incorporation of } -id \text{ nouns } \supseteq \text{Incorporation of } +id \text{ nouns} \]

\((85)\) Implicational hierarchy of incorporated nouns with different specificity values

\[ \text{Incorporation of } -s \text{ nouns } \supseteq \text{Incorporation of } +s \text{ nouns} \]

\((86)\) Implicational hierarchy of incorporated nouns with Background and Focus function

\[ \text{Incorporation of nouns with Background function } \supseteq \text{Incorporation of nouns with Focus function} \]

2.7.3 The RL-ML interface

The examples in \((77)\)–\((81)\) also illustrate the RL-ML interface conditions for noun incorporation in Kalaallisut. Firstly, the examples include both the incorporated non-modifiable f-noun *palasi* ‘priest’ ((l) in \([79]\)) and the incorporated a-noun *panik* ‘daughter’ that is modified by *pinnir* ‘beautiful’ ((x) in \([77]\)). Secondly, incorporated a-nouns can both designate animate entities, as with \((x)\) in example \((77)\) and \((x)\) in \((81)\), and inanimate entities, as exemplified with \((l)\) in \((78)\) and \((x)\) in \((80)\). Note that animate incorporated entities may both be human and non-human: the verb *-qar* ‘have’ incorporates the human noun *panik* ‘daughter’ in example \((77)\), while the verb *si* ‘get’ in \((81)\) incorporates the non-human noun *aalisagaaq* ‘fish’. Thirdly, the
constructions in (77), (80) and (81) show an incorporated noun with the semantic function of Undergoer ((x_j)U), while Nuuk ‘Godthaab’ in (78) is an incorporated noun that has the semantic function of Locative (l_i). Crucially, Sadock (2003: 31, 46) notes that an incorporated noun in Kalaallisut always corresponds to a verb’s “semantic object”, with the exception of predicatively used incorporated nouns like palasi ‘priest’ in (79). From this we infer that the incorporation of nouns with the semantic function of Actor is not possible. In addition, this information indicates that the incorporation of intransitive arguments and modifiers is excluded. Correspondingly, the morphosyntactic alignment system for noun incorporation in Kalaallisut is accusative, as arguments of intransitive verbs and Actor arguments of transitive verbs contrast with Undergoer arguments of transitive verbs in not being able to be incorporated. Kalaallisut also predominantly shows neutralization between Undergoer arguments of transitive and ditransitive verbs: in the same way as transitive verbs, ditransitive verbs tend to incorporate their Undergoer arguments, as illustrated in example (80). The language thus shows a primarily accusative-indirective morphosyntactic alignment system in noun incorporation. Interestingly, at the clausal layer Kalaallisut generally uses an ergative-secundative system for case-marking (Fortescue 1984: 80, 82; Malchukov 2013: 283), i.e. in Kalaallisut the morphosyntactic alignment system for the Clause and Word layer differ. Finally, the examples show that both relational nouns, such as panik ‘daughter’ (x_j) in (77), and non-relational nouns, such as Nuuk (l_i) in (78), can be incorporated in Kalaallisut.

The basic settings and position on the hierarchies for the RL-ML interface for noun incorporation in Kalaallisut can thus be presented as in (87)–(92).

(87) Basic setting regarding incorporated nouns at different semantic layers
Incorporation of f-nouns/Incorporation of α-nouns/Incorporation of both f-nouns and α-nouns

(88) Implicational hierarchy of incorporated nouns with different animacy values
Incorporation of inanimate nouns ⇒ Incorporation of non-human animate nouns ⇒ Incorporation of human animate nouns

(89) Implicational hierarchy of incorporated nouns with different semantic functions
Incorporation of Undergoer ⇒ Incorporation of other semantic functions ⇒ Incorporation of Actor
Incorporation: Constraints on variation

(90) Implicational hierarchy of incorporated nouns functioning as semantic arguments and modifiers

\[ \text{Incorporation of transitive arguments} \supseteq \text{Incorporation of intransitive arguments} \supseteq \text{Incorporation of modifiers (adjuncts)} \]

(91) Basic settings regarding alignment systems in the context of noun incorporation

- Accusative/Ergative/Neutral
- Indirective/Secundative/Neutral

(92) Implicational hierarchy of relational and non-relational incorporated nouns

\[ \text{Incorporation of relational nouns} \supseteq \text{Incorporation of non-relational nouns} \]

2.7.4 The ML-PL interface

The examples in (77)–(81) also provide information about the two basic settings for Kalaallisut that are relevant for the ML-PL interface. Incorporated nouns in Kalaallisut do not take suppletive forms: they simply correspond to the stems of independently used nouns (Sadock 1985: 399). Finally, an incorporated noun and its incorporating verb form a single phonological word in Kalaallisut, which can be shown on the basis of several morphophonological processes (Sadock 2003: 12–19). For instance, the incorporated noun panik ‘daughter’ in (77) loses its final consonant /k/ under influence of the word-internal following /q/ of the verbal Root -qar ‘have’, while the incorporating verb -(r)paluq ‘be like’ in (79) takes the form /rpaluq/ rather than /paluq/ because it is attached to a nominal stem that ends in a vowel, i.e. palasi ‘priest’.

The basic settings for the ML-PL interface for noun incorporation in Kalaallisut are shown in (93) and (94).

(93) Basic setting regarding incorporated nouns with different phonological heads

\( \text{(Some) incorporated nouns have suppletive forms} / \text{Incorporated nouns have non-suppletive forms} \)

(94) Basic setting regarding incorporated nouns at different phonological layers

\( \text{Incorporated nouns as separate Pw} / \text{Incorporated nouns as part of the verbal Pw} \)

---

20 More rarely, incorporated nouns may correspond to independent inflected nouns (Sadock 1980: 315), but such incorporated nouns are not found in the examples in (77)–(81).
2.7.5 Mappings

When comparing the representations at the various levels in (77)–(81) it is remarkable that, although some constructions show one-to-one mappings between IL, RL, ML and PL, in other constructions mismatches can be found. Interestingly, the construction in (81) shows one-to-one mappings between IL and RL on the one hand and between ML and PL on the other, but mismatches occur between IL/RL on the one hand and ML/PL on the other. This is shown in (95) and in Figure 2 below.

(95) Noun incorporation in Kalaallisut

Esther fresh-INS.SG fish-get-3SG.IND
‘Esther got (a) fresh fish.’
(Van Geenhoven 1998: 18)

That is, in example (95) the mismatches are purely a matter of Encoding. In this example a single Referential Subact (RJ) at IL maps onto a single Individual (xJ) at RL. The two Ascriptive Subacts (TJ) and (TK) that make up (RJ), map onto one Property, (fJ) and (fK), each. So there is a straightforward mapping from IL to RL. In the step from IL/RL to ML things are radically different. The Property (fJ) and the head of its Undergoer argument (fJ) form a single Verbal word (VwJ) at ML. The modifier of the Undergoer argument (fK) forms a single Noun phrase (NpK), and the Actor argument (xI) constitutes another Noun phrase (NpI). The elements that make up the Verbal word at ML thus do not make up any unit at RL. The mapping from ML to PL is then straightforward again, as Morphosyntactic Words at ML correspond to Phonological Words at PL.

2.8 Discussion and conclusions

This chapter has shown that the constraints on noun incorporation require a multi-level analysis, such as that provided by FDG. The four levels of analysis in this model provide the means to capture the pragmatic, semantic, morphological and phonological properties of incorporated nouns, while the operations connecting these levels provide the means to define the constraints that govern the possible mappings between levels in incorporation in a given language. By defining these constraints as a combination of implicational hierarchies and basic settings, the cross-linguistic variation in the field of noun incorporation can be described in a systematic way.

Furthermore, in studying how combinations of properties from all levels of analysis play a role in the system of noun incorporation in a single language, Kalaallisut, in Section 2.7, we have demonstrated that the basic distinction in FDG between Formulation and Encoding, i.e. between IL/RL on the one hand and ML/PL on the other, is neatly reflected in the mismatches that incorporation can bring along.
Figure 2. FDG representations for example (95).
3 Formal variation in incorporation: A typological study and a unified approach

3.1 Introduction

This chapter 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 (Gerdt 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).²

(1)  Incorporation of a nominal stem into a verb in Chukchi³
a. ʔətt-e piri-nin-Ø melota-lyrn
dog-ERG catch-3SG>3SG-PST hare-ABS.SG
‘The dog caught the hare.’
b. ʔətt-ə-an milute-piri-yi-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 ʔətt-e ‘dog’, a verb with the stem piri ‘catch’ and a direct object noun in the absolutive case melota-lyrn ‘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

---

³The vowel differences between melota in (1a) and milute in (1b) are due to a vowel harmony rule (Kurebito 2012: 188, n. 3).

²Glosses in the examples are adapted to the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/resources/glossing-rules.php).

¹This chapter is a slightly adapted version of: Olthof, Marieke. 2020. Formal variation in incorporation: A typological study and a unified approach. Linguistics 58(1). 131–205.
incorporation process detransitivizes the verbal stem, such that the subject noun ʔəttə-n appears in the absolutive case in (1b). In addition, milute in (1b) is interpreted as an indefinite noun, in contrast to the free-standing noun melota-ɣən in (1a). Note also that while example (1b) shows a case of noun incorporation, cross-linguistically 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. neneno ‘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

\[
\text{Wisi ibi-musa-tuwi-ban.} \\
\text{two 1.PL>1.PL-cat-buy-PST} \\
\text{‘They bought two cats.’} \\
(\text{Allen et al. 1984: 297})
\]

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

\[
\text{t-ə-manyə-ə-yak-wən-Ø neneno-Ø} \\
\text{1SG.A-E-hand-E-wipe-3SG.P-PST child-ABS.SG} \\
\text{‘I wiped a child’s hands.’} \\
(\text{Kurebito 2012: 182})
\]

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 cross-linguistic variation in incorporation structures, proposed definitions of incorporation diverge greatly. One characteristic that is nevertheless argued to be shared by many incorporation structures and often even considered one of the distinctive properties of incorporation is that the incorporated 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 single morpheme (Baker 1988: 71–72, 2003: 306; Evans 1996: 65; Gerdt 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 I use the term root exclusively for bound lexical morphemes (see Section 2.3 and 3.2.3), 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 derivational 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 complex 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

\[ \text{?on=can-kuyʔ-ciʔ-šuk-wə} \quad \text{?i} \quad \emptyset \quad \text{can-šuk-wə} \]

1.ERG=sit-INS-give-2/3PL-COMPL and 3.ABS sit-2/3PL-COMPL.

‘I gave them seats and they sat down.’

(Johnson 2000: 276)

In addition, some languages allow the incorporation of inflected words (Iturrio 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).
Incorporation: Constraints on variation

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

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

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 \text{mansun} ‘ox’ and the modifier \text{ngilla-n} ‘bought’, which is again modified by the adverb \text{we} ‘newly’, can as a whole be incorporated into the verb with the stem \text{adkintu} ‘watch’ (Zúñiga 2017: 705), as demonstrated in (6b).

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

(a) \text{Adkintu-yaw-i we ngilla-n mansun}.

\[ \text{watch-PERAMB-IND newly-buy-PTCP ox} \]

(b) \text{Adkintu-we-ngilla-n-mansun-kiyaw-i}.

\[ \text{watch-newly-buy-PTCP-ox-PERAMB-IND} \]

‘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 cross-linguistic 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 3.2.1). The general focus in theoretical research on incorporation has been on the incorporation of stems (Mithun 1984, 1986a; Baker 1988, 1996, 2003, 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 incorporation elements.

\[ ^{\text{4}} \text{The change in the form of the perambulative suffix is due to allomorphy: when the suffix follows a vowel it takes the form -\text{yaw}, as in (6a), while in all other cases it is -\text{kiyaw}, as in (6b) (Zúñiga 2000: 50).} \]
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 chapter therefore investigates the range of variation in the forms of incorporated elements and their cross-linguistic 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 cross-linguistic 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, geographically 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 chapter is as follows. Section 3.2 discusses the theoretical background for the research, addressing previous theoretical work on incorporation and outlining the FDG approach to incorporation taken in the study. Section 3.3 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 3.4, Section 3.5 presents the results. Finally, Section 3.6 provides the
conclusions about the formal variation of incorporated elements, the cross-linguistic distribution of the different elements and the suitability of a unified treatment of the different forms.

3.2 Theoretical background

3.2.1 Previous accounts of incorporation

As indicated above, the aim of this chapter 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 subsection.

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 (Massam 2009: 1083–1086; Štekauer et al. 2012: 43–47; Haugen 2015: 415–421).

In the morphological or lexical approach, incorporation is considered a word-formation process similar or identical to compounding (Caballero et al. 2008: 390; Štekauer et al. 2012: 46; Muro 2009: 24). Important studies supporting this morphological account of incorporation are an early paper by Sapir (1911) and the research by Mithun (1984, 1986a), 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 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 (2016) analysis 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

---

5 Baker (1988: Ch. 4 and 5) also addresses verb and adverb incorporation, in which verbs and adpositions undergo head movement.
6 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.
7 Wiltschko (2009) also proposes a Distributed Morphology-based account of incorporation, but limits incorporation to bare roots and nominalized roots.
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-referral 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; Baker 1988: 144–145; Rosen 1989: 297, 302–304). As my aim is to provide a broad approach to incorporation that enables a unified treatment of all forms and types of incorporation, a more flexible approach than the ones discussed in this subsection is needed. FDG appears to provide this flexibility. Its approach to incorporation is discussed in the next subsection.

### 3.2.2 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

---

8 Technical terms as applied in FDG are capitalized (see Hengeveld and Mackenzie 2008: 44).
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) Incorporation of an adverb (modifier) into a verb (head) in Hokkaido Ainu

A-

*toyko-kikkik*.  

INDF.A

thoroughly-beat  

‘I beat (him) up thoroughly.’  

(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 *sal* ‘tobacco’ is incorporated into the verbal predicate *kit* ‘rub’ of the predication frame ‘the hunter rubbed the dog’.

(8) Incorporation of a nominal adjunct (modifier) into a verb (head) in Ket

*assano ke’d tib d=sal-a-t-a-kit*  

hunting person dog 3.SBJ=tobacco-3SG.M.OBJ-TC-PRS-rub  

‘The hunter “tobaccoed” the dog (to rid it of fleas).’  

(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 *shut* ‘shirt’ is incorporated into its verbal predicate *pe* ‘make’, and *shut* and *pe* together form the semantic head of the predication.
Incorporation: Constraints on variation

(9) Incorporation of a nominal argument (dependent) into a predicate (nucleus) in Southern Tiwa

\[ Ti-shut-pe-ban. \]

1SG>1SG-shirt-make-PST

‘I made the/a shirt.’

(Allen et al. 1984: 293)

At RL, incorporation constructions may thus either involve 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 (Xmn), other words (Xwn), phrases (Xpn) and clauses (Cln), such that the maximal template for words is the one shown in (10) (Hengeveld and Mackenzie 2008: 400).

(10) Maximally elaborated morphosyntactic word template

\[ (Xw1): [(Xmn0) (Xwn0) (Xpn0) (Cln0)] (Xw1) \]

(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 (Xmn0), (Xwn0), (Xpn0) and (Cln0) 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.9

---

9 Definitions of lexical morphemes, words, phrases and/or clauses used in FDG follow in Section 3.3.1.
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 cross-linguistically 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 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 cross-linguistic 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.

3.2.3 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: 849), “loose incorporation” (Miner 1986: 252) and “pseudo-incorporation” (Massam 2001, 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).

(11) (Pseudo-)incorporation of a nominal stem into a verb in Niuean

\[
\text{Takafaga } ika^=\tūmau^=nī \ a \ ia. \\
\text{hunt fish=always=EMPH ABS he} \\
\text{‘He’s always fishing.’}
\]

(Seiter 1980: 69)

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 free-standing 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 \(kənte\) 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

\[
\text{ɲen jum-te-ti-n-ai} \\
\text{I eat-banana-NPST-INTR-1} \\
\text{‘I am eating a banana.’}
\]

(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 \(b\)-\(a\)\(shṭā\) ‘my eye(s)’ can be incorporated, as shown in (13), which includes the nominal stem \(išṭa\) ‘eye’ that
shows a vowel alternation (í > á) when it combines with the first person singular possessive prefix b-. (Graczyk 2007: 54–55).

(13) Incorporation of a nominal inflected word with stem alternation into a verb in Crow

\[
\begin{align*}
baapée-sh & \quad bim<ma>hp-dk \\
\text{day-DET} & \quad \text{swim}<1.ACSBJ>-SS \\
hawa-ss-b-ikusk-adh-aache & \\
\text{around-GOAL.-1.ACSBJ-come.out-DISTR-APPROX} & \\
\text{a} & \quad b-\text{ashtá-wis-aat-bee-m} & \quad \text{bittáchi-k} & \\
\text{until 1.POSS-eye(s)-open-APPROX.-1.ACSBJ.MIR-DS} & \quad 1.PRO.\text{alone-DECL} & \\
\end{align*}
\]

‘Today I went swimming, I was coming out here and there until I opened my eyes and to my surprise I was alone.’

(Graczyk 2007: 211)

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. 10 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.

---

10 Clauses do not correspond to a part of speech so the division between nouns, verbs, adjectives, adverbs and adpositions is not relevant for them.
Incorporation: Constraints on variation

With respect to incorporated elements, many studies address only or primarily nouns, i.e. they concentrate on noun incorporation (Gerðts 1998; Mithun 2000; Iturrioiz Leza 2001; Massam 2009). Nevertheless, there are also several studies that discuss constructions with incorporated elements of other parts of speech, such as incorporation of adpositions into verbs (Baker 1988: 229; Gerðts 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; Gerðts 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, 2000: 922; Gerðts 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, 1998, 1999: 48–56; Gerðts 1998: 94–98; Gerðts 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, 1999: 54, 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 (Gerðts 1998: 94; Kinkade 1998: 266; Mithun 1999: 54; Bischoff 2011: 1), are structures such as the one in (15) from Halkomelem.

(14) Denominal verb construction in Chukchi
\[ t\text{-}i\text{-}r\text{-}t\text{-}a\text{-}t\text{-}w\text{-}a\text{-}r\text{k}\text{a}n \]
\[ 1\text{SG}\text{-}S\text{-}skin\text{-}coat\text{-}E\text{-}take\text{-}off\text{-}E\text{-}PRS \]
\[ \text{‘I am taking off my skin coat.’} \]
(Kurebito 2001: 73)
Formal variation in incorporation

(15) Lexical affix construction in Halkomelem

\[ niʔ šk̓ʷ-əyəɫ \] \( \theta \) Mary.

AUX bathe-baby DET Mary

‘Mary bathed the/a baby.’

(Gerdts 2003: 347)

In (14), the element \(-tw\) ‘to take off’ has a verbal meaning but is a bound morpheme that obligatorily combines with an incorporated element. In (15), \(-əyəɫ\) ‘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 1986a, 1997: 364, 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, 1999: 50–54, 2009: 11–12), their discourse function of backgrounding information that is already known (Mithun 1997: 364–365, 1999: 51–54, 2009: 12; Mathieu 2013: 117–118) and syntactic characteristics such as the possibility to strand modifiers of the incorporated element (Haugen 2007: 150, 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, 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, 1998: 63, 65, 1999: 48–49; Gerdts 1998: 94; Kurebito 2001: 65; Stonham 2008: 514) and very numerous (Czaykowska-Higgins et al. 1996: 29; Mithun 1997: 264, 1999: 54; Gerdts 1998: 94, 2003: 346; Kurebito 2001; 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).

\[\text{Doubling is not possible in all languages with denominal verb constructions (Mathieu 2013: 127–128 [for Eastern Ojibwa]), but for some types of incorporation and in some languages with incorporation doubling is not possible either (i.e. for Compound NI [Rosen 1989] or “lexical compounding” incorporation [Mithun 1984]) (Gerdts 1998: 95–96).}\]
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 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

\[
\begin{align*}
\text{kati-ssamaar-put} \\
get.married-plan-3PL.IND \\
\end{align*}
\]

‘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). 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 truckdriver, in which truck is an argument of drive; and conjunct-conjunct compounds, such as singer-composer (Hengeveld and Mackenzie 2016: 1151). 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.

---

12 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 forthc.).

13 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 (Scalis 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.
3.3 Hypotheses

3.3.1 The possible forms of incorporated elements

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

(17) Maximally elaborated morphosyntactic word template

\( (Xw_1): [(Xm_n) (Xw_n) (Xp_n) (Cl_n)] (Xw_1) \)  

(Hengeveld and Mackenzie 2008: 400)

As this template specifies that words, including incorporation structures, may cross-linguistically include morphemes \((Xm_n)\), words \((Xw_n)\), phrases \((Xp_n)\) and clauses \((Cl_n)\), 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 3.2.3, 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, 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 3.2.3).

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).

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.

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 3.2.3, lexically derived stems are considered single morphemes and not incorporation structures at ML (Hengeveld and Mackenzie 2008: 229–230, 2016: 1150).
3.3.2 The cross-linguistic distribution of the various forms of incorporated elements

While FDG predicts that incorporated elements may cross-linguistically 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 3.2.2). 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.

Cross-linguistic 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; Velupillai 2012a: 35; Moravcsik 2013: 54). 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
Formal variation in incorporation 73
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 cross-linguistic 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

\[\text{lexical morpheme} \supset \text{derived stem} \supset \text{inflected word} \supset \text{phrase} \supset \text{clause}\]

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.

3.4 Method

3.4.1 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, 1985, 1986; Baker 1988, 1996; Rosen 1989; Anderson 2000), a number of overview articles on incorporation (Mithun 1994, 2010; Iturrioz Leza 2001; Anderson 2007; Massam 2009) and a few cross-linguistic 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, geographical 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 ten 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

---

15 The list in Appendix 1 includes 11 additional languages that were identified as incorporating languages after the sample of the study presented here was drawn. These languages, marked with “**”, were not taken into account in the sampling procedure of the present study.
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.

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, Gunwinygu, Mayali]</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Chumalapa Zoque [San Miguel Chumalapa 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>Kalallisut [Eskimo, 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>Marrthiel [Marrthiuyel]</td>
<td>Western Daly</td>
<td>Australia</td>
<td>Australia</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 of</td>
</tr>
<tr>
<td>Nadeb</td>
<td>Nadahup</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Nuean [Niue]</td>
<td>Austronesian</td>
<td>Papunesia</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, Nuuchahuluth]</td>
<td>Wakashan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Palikur [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 Republic of</td>
</tr>
<tr>
<td>Paraguayan Guarani [Guaraní, Paraguayan Guarani]</td>
<td>Tupian</td>
<td>South America</td>
<td>Argentina; Paraguay</td>
</tr>
<tr>
<td>Sora</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
</tr>
</tbody>
</table>
### Incorporation: Constraints on variation

| South Slavey [Slave, Slavey] | Athapaskan-Eyak-Tlingit | North America | Canada |
| South Tiwa [Tiwa] | Kiowa-Tanoan | North America | United States |
| Ute-Southern Paiute [Chemehuevi, Southern Paiute, Ute] | Uto-Aztecan | North America | United States |
| Washo | Washo (Isolate) | North America | United States |
| Western Frisian [Frisian, Frisian] | Indo-European | Eurasia | Netherlands |
| Western Highland Chatino [Yaitpepec Chatino] | Otomanguean | North America | Mexico |
| Yimas | Lower Sepik-Ramu | Papuasia | Papua New Guinea |
| Yucatec Maya [Maya Yucateco, Yucatecan Mayan] | Mayan | North America | Belize; Guatemala; Mexico |

#### 3.4.2 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 3.3.1. 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.

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 3.3.1, 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
Formal variation in incorporation  77

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

\[
\begin{align*}
\text{juulli-p} & \quad \text{kingurn-a-gut} & \quad \text{Nuum-mukar-puq} \\
\text{July-RC SG} & \quad \text{after-3SG POSS PROS} & \quad \text{Nuuk-go.IO-3SG IND}
\end{align*}
\]

‘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,
synchronously -mut(t) and -kar are recognized as a single morpheme -mukar, i.e. it is
not possible to use -kar without -mut(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
Incorporation: Constraints on variation

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

\[
\begin{align*}
\text{illu-mi-} & \text{-niip-puq} \\
\text{house-REFL.POSS-be.in-3SG.IND} \\
\text{‘He is in his (own) house.’} \\
\text{(Fortescue 1984: 300–301)}
\end{align*}
\]

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

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

\[
\begin{align*}
\text{Anta-} & \text{-p} \quad \text{Aani-} & \text{-lu} \quad \text{irnir-an-niip-puq} \\
\text{Anta-RC.SG} & \text{Aani-RC.SG=} \text{and} \text{ son-3PL.POSS-be.at-3SG.IND} \\
\text{‘She is at the house of Anta and Aani’s son.’} \\
\text{(Fortescue 1984: 135)}
\end{align*}
\]

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.

3.5 Results

The hypotheses concerning the variation in the forms of incorporated elements and the cross-linguistic 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.

---

16 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).
Table 2. The occurrence of the different forms of incorporated elements in the sample languages. “+” means that the form can be incorporated in the language, “−” shows that the form cannot be incorporated in the language and “#” indicates that the form is irrelevant for the parts of speech that can be incorporated in the language.

<table>
<thead>
<tr>
<th>Language</th>
<th>Lexical morpheme</th>
<th>Derived stem</th>
<th>Inflected word</th>
<th>Phrase</th>
<th>Clause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crow</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Hokkaido Ainu</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Ket</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Mapudungun</td>
<td>+</td>
<td>+</td>
<td>#</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Nadeb</td>
<td>+</td>
<td>+</td>
<td>#</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Niuan</td>
<td>+</td>
<td>+</td>
<td>#</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Northern Guuguimidjaal</td>
<td>+</td>
<td>+</td>
<td>#</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Bininj Kun-Wok</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Halkomelem</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Kalaallisut</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Nuu-chah-nulth</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>Sora</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>South Slavey</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Ute-Southern Paiute</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Yimas</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Chimalapa Zoque</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Chukchi</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Eastern Ojibwa</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Iraqw</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Mohawk</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Movima</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Southern Tiwa</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Western Frisian</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Marthiel</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Palikuir</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Panare</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Paraguayan Guaraní</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Washo</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Western Highland Chatino</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Yucatec Maya</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

3.5.1 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
Incorporation: Constraints on variation

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
\[nam\mathrm{ma} \ ə\mathrm{y}=w\mathrm{ə}h-a-\mathrm{cak}-\mathrm{šuk}-wə\]
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
\[Le’ \ t’\mathrm{eel}-o’ \ […] \ ken \ uy-u(’u)b \ a-t’aan-e’\]
DEF hen-DDEIX SUB.FUT 3.SBJ-hear 2.SBJ-speak-LNDEIX
k-u=chen-t’aan
HAB.3.SBJ=just-speak
'The rooster […] when he hears you speaking, he just speaks.'

In (22) *wəhə* ‘good’ is a modifier of the head *cak* ‘do’, while in (23) *chen* ‘just’ modifies *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 -*mo* ‘bird’ is a nominal bound morpheme (Haude 2006: 220), functioning as an argument of the verb *yok* ‘catch’.

(24) Incorporation of a nominal bound root into a verb in Movima
\[asko \ yok-ə-mo-na=\mathrm{is} \ \mathrm{ney}=s \ \mathrm{kara’} \ di’ \ \mathrm{sere}:\mathrm{re}\]
3SG.N.AB catch-DR.BE,bird-DR=PL.AB here=DEF 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.

17 The incorporated element *wəhə* ‘good’ is an adjective (Johnson 2000: 68, 278), although it functions as an adverb here, modifying the verb (Johnson 2000: 278).
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

\[
\text{baa-luúsh-chiili} \\
\text{INDF-eat-look for} \\
\text{‘look for food’}
\]

(Graczyk 2007: 281)

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

\[
d=\text{kahwe-ʔuk-tuk-wə} \\
1\text{ABS=}\text{coffee-drink-finish-COMPL} \\
\text{‘I finished drinking coffee.’}
\]

(Johnson 2000: 221)

The incorporated element in the Crow example in (25) consists of the verb \text{luúsh} ‘to eat’ and a nominalizing indefinite prefix \text{baa-} (Graczyk 2007: 48, 50) and is an argument of its host, \text{chiili} ‘look for’ (Graczyk 2007: 297). The Chimalapa Zoque verbal compound \text{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 \text{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 \text{mampi} ‘again’, consisting of the adjective \text{ma} ‘other’ and the adverbializing suffix \text{-mpi} (Foley 1991: 343), is incorporated in (27), in which it modifies the predicate \text{wampujkra} ‘angry’.

(27) Incorporation of a adverbial derived stem into a verb in Yimas

\[
\text{na-n-ma-mpi-ira-wampujkra-ntut} \\
3\text{SG.P-3SG.A-other-ADVZ-ALL-angry-RMP} \\
\text{‘He was angry with her again.’}
\]

(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 don’-aŋ ‘knives’ incorporated into its predicate, the verb vet ‘make’.

(28) Incorporation of a nominal inflected word in Ket
d-don’-aŋ-s’-i-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 nilir-sima ‘had been cold’, part of the dependent imiq nilir-sima ‘the water had been cold’, is incorporated into the nucleus nirar ‘say that’.

(29) Incorporation of a verbal inflected word in Kalaallisut
imiq nilir-sima-nirar-paa
water be.cold-PFV-say.that-3SG>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 wą ‘to’ with the first person singular affix se- (Rice 1989: 269), modifying tlah ‘go’, in (30).

(30) Incorporation of an adpositional inflected word into a verb in South Slavey
se-wą-e-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).
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.’

The incorporated phrase in the Niuean example in (32) contains a noun, mena ‘thing’, and a subjunctive relative clause, ke nonofo=ai ‘to settle there’ (Massam 2001: 161). The Crow incorporated element in (33) is considered a phrase because the

(33) Incorporation of an adposition phrase into a verb in Crow

ash-bacheeitchi-m Apsáalooke kuxshi-kadż-ak
lodge-chief-SIM Crows help-AUG-SS
ak=baa’ilápixaahkuua=ss-dee-sh hééleé-la-k
REL=Washington=GOAL-go-DET among-be.at-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, mena ‘thing’, and a conjunctive relative clause, ke nonofo=ai ‘to settle there’ (Massam 2001: 161). The Crow incorporated element in (33) is considered a phrase because the

---

18 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: Constraints on variation

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 baailiki·p·xisaahkuua ‘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-tay·palwanta·pojga ‘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.19

\begin{enumerate}
\item 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
\begin{verbatim}
tor-tay-palwanta-pojga-pela-rkan
\end{verbatim}
1SG.S-new-good-metal-spear-leave-PRS
\begin{quote}
‘I am leaving a good, new, metal spear.’
\end{quote}
\end{enumerate}

As it is not possible to identify the incorporated element tor-tay-palwanta-pojga ‘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 3.4.1). 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 3.2.2 and 3.3.1 is somewhat too broad with respect to the forms of incorporated elements.

\begin{enumerate}
\item 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 absolute case-marking, even though it functions as a direct object. The absence of absolute case-marking does thus not exclude the possibility that the incorporated element is really a phrase.
\end{enumerate}
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-externally, such that examples of clause incorporation may simply not have been included in the literature consulted. 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 3.4.2, 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-īluu-h
2.ACSBJ-stand-IMP
dii-līi-wah-kūnnaa-wuu-o-k”  huu-k
2.PRO-2.OBJ-1.ACSBJ-fetch-1.ACSBJ.PL.come-PL-DECL  say.PL-DECL

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

(Graczyk 2007: 313)

In this example the incorporated element *dii-līi-wah-kūnnaa* ‘we fetch you’ contains a verb with a first person subject prefix *wah-* and a second person object prefix *līi-* 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 3.4.1). 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.

(37)  \(ngune\) ‘eat’ and \(ganj\) ‘meat’ as independent words in a clause in Bininj Kun-Wok

\[Barri-ngune-ngun-ganj.\]

3.AU>PST-eat-PST.PFV  IV-meat

‘They ate the meat.’

(Evans 2003: 330)

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.
(38) Incorporation of a verbal derived stem, in the form of a grammatical compound, into a verb/incorporation of a clause into a verb in Chukchi

\[ \textit{ənko mat-mac-qora-gərke-plako-mə} \]

\[ \text{"then 1PLS-almost-deer-hunt-finish-1PLS"} \]

‘Then we almost finished hunting reindeer.’


Example (38) can be analyzed in various ways. First of all, because the incorporated adverb \textit{mac} ‘almost’ semantically modifies the host of the incorporation construction, \textit{plako} ‘finish’, rather than the incorporated verb \textit{gərke} ‘hunt’ (Spencer 1995: 461), it seems likely that this construction involves two incorporated elements, i.e. the adverb \textit{mac} ‘almost’ and the combination of the verbal predicate \textit{gərke} ‘hunt’ and its nominal argument \textit{qora} ‘deer’. Indeed, Spencer (1995: 459) discusses this example as a case of “multiple incorporation”. The adverb \textit{mac} ‘almost’ thus seems to be incorporated as a single adverbial morpheme.\(^{20}\) However, because \textit{gərke} ‘hunt’ and \textit{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 \textit{qora} is incorporated into \textit{gərke} or a clause consisting of a nominal word \textit{qora} and a verbal word \textit{gərke}, 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 case of example (36) from Bininj Kun-Wok, there is no way to determine whether the incorporated element \textit{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

\(^{20}\) Note that Dunn (1999: 141) actually describes \textit{mac} as an approximative verbal prefix rather than as a lexical adverb. According to that analysis, the combination of \textit{qora} ‘deer’ and \textit{gərke} ‘hunt’ is the only incorporated element in example (38).
an alternation when used as an incorporated stem, as most nouns in Sora do for example (see Section 3.2.3), 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 nonof=ai* ‘a place to settle’ is a noun phrase consisting of a head noun *mena* ‘thing’ and a relative clause *ke nonof=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.

### 3.5.2 The cross-linguistic distribution of the various forms of incorporated elements

The second hypothesis presented in Section 3.3 was that the various forms that incorporated elements may take constitute the implicational hierarchy presented in (18), repeated here as (39).

\[(39) \quad \text{Hypothesized implicational hierarchy of incorporated elements} \]

\[
\text{lexical morpheme} \supset \text{derived stem} \supset \text{inflected word} \supset \text{phrase} \supset \text{clause}
\]

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 3.5.1, 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.

3.6 Conclusion

The aim of the present study has been to survey the forms that incorporated elements may take cross-linguistically 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 cross-linguistically. 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 which 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.
4 Referentiality and modifiability of incorporated nouns: Cross- and intra-linguistic variation¹

4.1 Introduction

Noun incorporation can be broadly defined as the combination of a noun and a verb such that they together form a new, complex predicate (Gerdts 1998: 84; Mithun 2000: 916; Massam 2009: 1078, 2017; see also Sapir 1911: 257). To illustrate the phenomenon, an example from Bininj Kun-Wok is presented in (1).²

(1)  Noun incorporation in Bininj Kun-Wok
      3.AU>PST-eat-PST.PFV IV-meat
        ‘They ate the meat.’
  b. Barri-ganj-ngune-ng.
      3.AU>PST-meat-eat-PST.PFV
        ‘They ate the meat.’
      (Evans 2003: 330)

While example (1a) shows a regular transitive clause with a verb with the stem ngune ‘eat’ and a direct object noun gun-ganj ‘meat’, in example (1b) the stem of the direct object noun, ganj, is incorporated into the verb barri-ngune-ng. The position of the nominal stem ganj in (1b) between the verbal prefix barri- and the verbal stem ngune overtly shows its status as an incorporated noun. Incorporated nouns are arguments, as in (1b), or modifiers of the verbs in which they are included (Mithun 2000: 917; Haugen 2015: 414–415; Massam 2017). Incorporating languages are genetically diverse and are especially numerous in North and South America, northern Australia, Austronesia and Siberia (Mithun 2000: 926–927; Velupillai 2012a: 120).

The present chapter investigates whether incorporated nouns like ganj in (1b) are used to refer in the same way as unincorporated nouns can be and whether they can be modified by, for instance, adjectives, demonstratives and relative clauses, just like most unincorporated nouns. These questions are persistent issues in the literature on noun incorporation (Mithun 1994: 5025–5026; Farkas and De Swart 2003: 17, 2004: 46; Baker 2009: 152–153; Massam 2009: 1084, 1086, 2017; Murasugi 2014: 284; Borik and Gehrke 2015: 5–6; Barrie and Mathieu 2016: 36–37). While some

¹ This chapter is a slightly adapted version of: Olthof, Marieke. forthcoming. Referentiality and modifiability of incorporated nouns: Cross- and intra-linguistic variation. Sprachtypologie und Universalienforschung.

² Glosses in the examples are adapted to the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/resources/glossing-rules.php). The use of "*" shows that an example is ungrammatical, whereas the use of "#" indicates that an example is semantically anomalous.
researchers argue that incorporated nouns have a referential function, others claim that they are not used to refer. Similarly, some hold that incorporated nouns are modifiable, whereas other researchers maintain that they are non-modifiable.

An important reason why researchers disagree about the referential potential and modification possibilities of incorporated nouns is that they understand the notion of referentiality in different ways (Mattissen 2003: 173; Massam 2017) and put opposing interpretations on apparent modifiers of incorporated nouns. In addition, studies focusing on different languages and sometimes even studies concentrating on one and the same language draw different conclusions about the referentiality and modifiability of incorporated nouns (see Mithun 1984: 866–872 and Baker 1988: Ch. 4, 1996: Ch. 7, who both discuss Mohawk and Bininj Kun-Wok). Thus, there appears to be cross- and intra-linguistic variation regarding the referential potential and modification possibilities of incorporated nouns (Sadock 1991: 86–88; Massam 2001: 169–171, 174–175, 2009: 1084, 2017; Chung and Ladusaw 2003: 126–128; Farkas and De Swart 2003: 148; Murasugi 2009: 100, 129; Murasugi 2014: 284–285; Borik and Gehrke 2015: 6).

The aim of the present study is therefore to systematically explore the range of cross- and intra-linguistic variation with respect to the referentiality and modifiability of incorporated nouns based on a fixed set of criteria for the identification of referentially and non-referentially used nouns and a consistent approach to elements that appear to modify incorporated nouns. Examining noun incorporation constructions in a sample of 21 incorporating languages, the study attempts to tease apart the conflicting ideas about the referential potential and modification possibilities of incorporated nouns presented in the literature. The findings are compared to the predictions and assumptions about the referentiality and modifiability of incorporated nouns made by various theoretical approaches to noun incorporation in order to assess the theoretical implications of the attested cross- and intra-linguistic variation.

The criteria for referentiality and the approach to modifiers used in the study are taken from Functional Discourse Grammar (FDG, Hengeveld and Mackenzie 2008). The FDG framework is suitable for this study because it assumes that referentially used nouns can be distinguished from non-referentially used ones based on their ability to function as antecedents in anaphoric reference (Hengeveld and Mackenzie 2008: 114), and anaphoric reference is also used as a test for referentiality in much of the research on noun incorporation (Sadock 1980: 311, 1991: 86–88; Mithun 1984: 866–867, 871; Baker 1988: 78–79, 1996: 287–291; Van Geenhoven 1998: 47–49; Massam 2001: 169–171, 174–175; Chung and Ladusaw 2003: 121–124; Farkas and De Swart 2003: 148; Wojdak 2005: 55; Barrie and Mathieu 2016: 3). In addition, FDG separates the semantic contribution of modifiers from their morphosyntactic expression, such that the modifiability of incorporated nouns can be addressed from a semantic perspective, i.e. independently of the possible morphosyntax of modifiers. Because incorporated nouns do not take the regular
morphosyntactic position of nouns, it is possible that their modifiers, if they can have any, also have special morphosyntactic characteristics. The FDG approach leads to a three-way typology of common nouns in which referentially used modifiable nouns, non-referentially used modifiable nouns and non-referentially used non-modifiable nouns are recognized (Smit 2005: 102–103). This typology is applied to incorporated nouns in the sample languages in order to investigate the cross- and intra-linguistic possibilities regarding the referentiality and modifiability of incorporated nouns.

The outline of the chapter is as follows. Section 4.2 introduces the notions of referentiality and modifiability as defined in FDG and the pragmatic-semantic typology of nouns that follows from these. Section 4.3 describes the method, focusing on the sampling procedure, the definition of noun incorporation employed in the study and the data analysis. In Section 4.4, the results of the study are presented, i.e. this section shows the range of variation with respect to the referentiality and modifiability of incorporated nouns that is found. Section 4.5 then discusses the results and their implications for theoretical approaches to noun incorporation. Finally, Section 4.6 provides the conclusions about the cross- and intra-linguistic variation regarding the referential potential and modification possibilities of incorporated nouns.

4.2 Referentiality and modifiability in FDG

4.2.1 Referentiality

The term referentiality, or reference, has a long history in the linguistic literature and is used in highly different ways (Chen 2009: 1657). A primary distinction can be made between a semantic and a pragmatic notion of referentiality (Abbott 2017: 240; see also Keizer 2015: 83). Very generally, a linguistic expression is semantically referential if its semantics inherently point at an entity in the world (Chen 2009: 1658; Abbott 2010: 3, 2017: 240). Pragmatic referentiality, by contrast, concerns the way a speaker uses a linguistic expression in context: a linguistic expression is pragmatically referential if a speaker uses it, in a particular discourse, to point at an entity (Chen 2009: 1659; Abbott 2010: 2, 2017: 240).

In FDG, the term referentiality is used for the pragmatic notion that pertains to the way in which nominal expressions are used in context (Hengeveld and Mackenzie 2008: 107, referring to Dik 1978: 55, 128; Hannay and Hengeveld 2009: 105, referring to Lyons 1977: 177; Keizer 2015: 83). The framework makes a pragmatic distinction between nouns that are used by a speaker to evoke an entity as a referent and nouns that a speaker employs to ascribe a property or entity (Hengeveld and Mackenzie 2008: 108–109, 113, 192–193; Keizer 2015: 83, 90–91). Because the pragmatic referentiality of a noun is determined by the way a speaker uses it, one and the same noun can both have a referential and a non-referential usage, in different contexts (Hannay and Hengeveld 2009: 105). Thus, whereas girl in example (2) is a
referentially used noun, as it, together with the article the and the adjective intelligent, is used to evoke an entity as a referent, the noun girl in example (3) is a non-referentially used noun, as the noun phrase an intelligent girl is here only used to ascribe the entity ‘an intelligent girl’ to Hannah.

(2) Referentially used noun
The intelligent girl passed the exam.
(Hengeveld 2008: 46)

(3) Non-referentially used noun
Hannah is an intelligent girl.

Referentially used nouns typically represent arguments, as in (2), or adjuncts, while non-referentially used nouns often function as predicates, as in (3). Nevertheless, there is no direct relation between the pragmatic referentiality of a noun and its status as argument or adjunct, on the one hand, or predicate, on the other hand. Although referentially used nouns never predicate by themselves, they may be part of larger, relational predicative expressions.³ Thus, in (4), the noun Shakespeare has a referential function but is at the same time part of the predicate by Shakespeare (Hengeveld and Mackenzie 2008: 190). Moreover, non-referentially used nouns may also represent arguments or adjuncts, as exemplified by piano and bike in (5) and (6) respectively.⁴

(4) Referentially used noun functioning as part of a predicative expression
This play is by Shakespeare.
(Hengeveld and Mackenzie 2008: 190)

³ Referentially used nouns are also found in identificational constructions, exemplified in (i).

(i) Identificational construction
My teacher is Peter.
(Hengeveld and Mackenzie 2008: 193)

As identificational constructions simply equate two entities with each other, neither of the nouns predicates something of the other (Hengeveld and Mackenzie 2008: 205; Keizer 2015: 137–138). Thus, referentially used nouns in identificational constructions are neither arguments, adjuncts nor predicates.

⁴ Evidence for the referential status of the noun Shakespeare as used in example (4) and for the non-referential status of the nouns piano and bike as used in example (5) and (6) comes from their possibilities with respect to anaphoric reference, which are used as a test for referentiality in FDG (see below). Note also that constructions like (5), in which a noun without marking for definiteness, number and/or case appears adjacent to a verb, are sometimes considered to involve “pseudo-incorporation” or “semantic incorporation” (Massam 2001: 157; Stvan 2009: 314; Borik and Gehrke 2015: 10). The construction in (6) may also be analyzed as a complex predicate.
Referentiality and modifiability of incorporated nouns

Non-referentially used noun functioning as an argument
Phil is playing piano for the choir.
(Le Bruyn et al. 2017)

Non-referentially used noun functioning as an adjunct
I went to Amsterdam by bike.
(Keizer 2015: 91)

Importantly, in FDG not only nouns that are used to refer to specific entities but also nouns that are used to evoke non-specific entities are considered to be pragmatically referential (Hengeveld and Mackenzie 2008: 113, 122; Hannay and Hengeveld 2009: 112; Keizer 2015: 95–96). Thus, both the noun cottage in (7), which is used to refer to a particular cottage that the speaker can identify, and the noun cottage in (8), which is used to evoke “any entity that fits the description” (Rijkhoff 2002: 235), are used referentially.\(^\text{5}\)

(7) Referentially used noun evoking a specific entity
We saw a lovely cottage yesterday.
(Keizer 2015: 99)

(8) Referentially used noun evoking a non-specific entity
We are looking for a cottage, preferably in the Lake District.
(Keizer 2015: 96)

In this respect, the FDG notion of referentiality differs from some other pragmatic notions of referentiality in which only nouns used to refer to specific entities are considered to function referentially (Payne 1997: 264; Lyons 1999: 165; Chen 2009: 1659). Payne (1997: 264) argues, for instance, that a noun is only used pragmatically referentially “if it exists as a bounded, individuated entity in the message world”. When a noun is used non-specifically, it is not clear if a corresponding entity really exists.

The reason why FDG nevertheless regards both nouns like cottage as used in (7) and nouns like cottage as used in (8) as referentially used nouns is that they are both available for anaphoric reference (Hengeveld and Mackenzie 2008: 114). Thus, not only the specific entities evoked by the nouns girl in (9a) and cottage in (9b) but also the non-specific entity referred to by the noun cottage in (10) can function as the antecedent for an anaphoric pronoun.

---

\(^5\) In English, specificity is not marked on nouns. The specific vs. non-specific distinction between example (7) and (8) can therefore only be understood from the context.
Incorporation: Constraints on variation

(9) Anaphoric reference to specific entities
   a. The intelligent girl, passed the exam even though she, hadn’t spent much time studying.  
      (adapted from Hengeveld 2008: 46)
   b. We saw a lovely cottage, yesterday. We are thinking of buying it.  
      (Keizer 2015: 91)

(10) Anaphoric reference to a non-specific entity
    We are looking for a cottage, preferably in the Lake District. It, should be available from next summer.  
    (adapted from Keizer 2015: 96)

By contrast, non-referentially used nouns do not evoke entities that can function as antecedents for regular anaphoric reference, whether they are used as predicates (11a–b), arguments (11c) or adjuncts (11d) (Hengeveld and Mackenzie 2008: 114; Keizer 2015: 91). Correspondingly, anaphoric reference is either inappropriate, as in (11c–d), or a special type of anaphor is required that does not refer back to a referentially evoked entity but only to an entity or property that is ascribed, as in (11a–b) (Hengeveld and Mackenzie 2008: 121, 193; see also Doron 1988: 284; Chen 2009: 1663; De Swart and Zwarts 2009: 289).6

(11) Anaphoric reference to nouns that are used non-referentially
   a. Hannah is an intelligent girl, That’s what she is.  
      (based on Hengeveld and Mackenzie 2008: 193)
   b. Hannah is an intelligent girl, and so, is Lucy.  
      (based on Hengeveld and Mackenzie 2008: 193)
   c. Phil is playing piano for the choir. He complains it is out of tune.  
      (Le Bruyn et al. 2017)
   d. #I went to Amsterdam by bike, but it, broke down.  
      (Keizer 2015: 91)

---
6The anaphors that and so are also regularly used to refer back to adjectival and verbal predicates, as in (ii) and (iii).

(ii) Anaphoric reference to an adjectival predicate
    Hannah is intelligent, and so, are you.  
    (based on Hengeveld 1992: 53)

(iii) Anaphoric reference to a verbal predicate
    John went swimming, That’s what he did.  
    (Hengeveld and Mackenzie 2008: 193)

This function further supports the claim that these anaphors refer back to ascribed nominal properties or entities rather than to entities that are evoked as referents in example (11a–b) (Hengeveld and Mackenzie 2008: 193).
Similarly, referentially used nouns can be co-referential with other referentially used nouns (Rijkhoff 2008: 798; see also Dik 1997: 130), while this is not possible for nouns that are used non-referentially. In example (12), the noun *cat* in the second clause is clearly used to refer to the same cat as the noun *cat* in the first clause, as emphasized by the demonstrative modifier *that*. By contrast, in example (13), the noun *girl* in the first and second clause cannot be co-referential, because *girl* in the first clause is not used to refer.

(12) Co-referential relation between referentially used nouns

_Yesterday in the park I saw a black cat. Today I saw that cat, again._

(Dik 1997: 130)

(13) Co-referential relation between a non-referentially used noun and a referentially used noun

_Hannah is an intelligent girl. #That girl, knows everything._

FDG thus uses anaphoric reference and co-reference as tests for referentiality. In addition, referentially used nouns can be recognized based on their ability to combine with elements that mark an entity as identifiable or non-identifiable for the addressee.7 A speaker may use a noun referentially in order to re-identify an entity that is already identifiable for the addressee or to introduce an entity into the discourse that, at that moment, is not identifiable for the speaker. Definite articles, demonstrative modifiers and interrogative modifiers indicate that a speaker assumes that the entity he or she refers to by means of a particular noun is identifiable for the addressee (Hengeveld and Mackenzie 2008: 122; Rijkhoff 2008: 797, 807–808; Keizer 2015: 95; see also Dik 1997: 180; Payne 1997: 102, 263; Lyons 1999: 18). Interrogative pronouns, deictic personal pronouns and proper names heading a noun phrase also present the evoked entity as identifiable for the addressee (Hengeveld and Mackenzie 2008: 117–119, 122–123; Keizer 2015: 93–94; see also Payne 1997: 39, 263; Lyons 1999: 21).8 Indefinite articles and indefinite pronouns, by contrast, can signal that the entity

---

7 Languages differ in the number and types of elements that they have available to mark the identifiability of an entity for the addressee. Nevertheless, because virtually all languages make use of demonstratives (Payne 1997: 102; Lyons 1999: 48), all languages do have the possibility to mark this type of identifiability.

8 In some contexts, proper names can be used non-referentially (Hengeveld and Mackenzie 2008: 117), as exemplified by *Houdini* in (iv).

(iv) Non-referentially used proper name

_My sister Houdini’d her way out of the locked closet._

(C Clark and Clark 1979: 784, cited in Hengeveld and Mackenzie 2008: 117)

However, because this use of proper name is highly restricted, it is not taken into account in this study.
referred to is assumed to be non-identifiable for the addressee (Hengeveld and Mackenzie 2008: 122–123; Keizer 2015: 46, 91).9

Finally, referentially used nouns can be identified based on their ability to form a noun phrase with a possessive noun or pronoun that is used to refer to a possessor entity that is assumed to be identifiable for the addressee.10 Crucially, based on the possessive relation between the possessor entity that is presented as being identifiable for the addressee and the possessed entity, the addressee is assumed to be able to identify the possessed entity as well (Rijkhoff 2008: 808–809; see also Payne 1997: 263–264; Lyons 1999: 23–24). The assumed identifiability of the possessor entity is often indicated in the same way as for other entities that are expected to be identifiable for the addressee: the possessor entity may be expressed by a noun combined with a definite article, demonstrative modifier or interrogative modifier or by a pronoun or proper name (Rijkhoff 2008: 808–809; see also Payne 1997: 263–264; Lyons 1999: 24). An example of a possessive noun that is used to refer to an entity that is identifiable for the addressee is included in (14).

(14) Referentially used possessed noun
I met Leila’s fiancé yesterday.
(Hengeveld and Mackenzie 2008: 119)

In this example, the possessor entity is referred to by means of the proper name Leila, which shows that it is assumed to be identifiable for the addressee. By expressing a possessive relation between Leila and the possessed noun fiancé, the speaker presents the entity evoked by fiancé as identifiable for the addressee as well. The speaker thus uses this noun referentially.

4.2.2 Modifiability
In FDG, the issue of modifiability is closely related to the semantic distinction between what can be called property-denoting nouns and entity-designating nouns. A

---

9 Note, however, that indefinite articles can also combine with non-referentially used nouns.
10 This criterion can only be used for nouns expressing possessed entities that function as arguments and adjuncts. When nouns expressing possessed entities are used predicatively, they may have a non-referential function, despite the presence of a possessor that is presented as identifiable for the addressee (Doron 1988: 285–286; Haspelmath 1999: 232, fn. 5; Lyons 1999: 25). In addition, inalienably possessed entities combined with a possessor that is marked as identifiable for the addressee, like brother in example (v), are not necessarily identifiable for the addressee (Lyons 1999: 25–26).

(v) Non-referentially used noun
I'm going to stay with my brother for a few days.
(Lyons 1999: 26)

If the speaker producing example (v) has more than one brother, the addressee may not be able to identify which brother the speaker refers to. However, in cases like (v), the possessed entities are specific, which shows that the nouns are nevertheless used referentially.
property-denoting noun is a noun that heads a noun phrase that only denotes a property, which has no independent existence and can only be evaluated in terms of its applicability (Hengeveld and Mackenzie 2008: 131; Keizer 2015: 105). By contrast, an entity-designating noun functions as the head of a noun phrase that designates an entity to which the property that the noun expresses applies and which may be evaluated in terms of its existence (Hengeveld and Mackenzie 2008: 131, 215; Keizer 2015: 105). Thus, the noun president in example (15) is a property-denoting noun, as it is not used to designate an entity but only denotes a property that happens to apply to that man, whereas president in example (16) is used to designate a particular entity by expressing that the property ‘president’ applies to it.

(15) Property-denoting noun
That man is president.
(Stowell 1991: 53)

(16) Entity-designating noun
The president waved to the crowd.
(Hengeveld and Mackenzie 2008: 197)

As exemplified for president in (15) and (16), nouns can in principle be used both as property-denoting and as entity-designating nouns, in different contexts. Note also that property-denoting nouns may not only be nominal predicates, as in (15), but may also function as arguments or adjuncts, as shown by town and prison in (17) and (18) respectively. At the same time, entity-designating nouns can be used predicatively, as exemplified by doctors in (19), or represent adjuncts, like knife in (20), just as well as they can be arguments, as in (16) above.

(17) Property-denoting noun used as an argument
They – whoever they is – think I’ve left town and I want to keep it that way.

(18) Property-denoting noun used as an adjunct
Two are currently in foster care – one girl because her father is in prison for murdering her mother; another girl spent last year in foster care.

---

11 The entity-designating nouns are further divided in FDG into individuals, states-of-affairs and propositional contents, which correspond to Lyons’ (1977: 442–447) first-, second- and third-order entities respectively (Hengeveld and Mackenzie 2008: 131; Keizer 2015: 105; see also Dik 1997: 136–137).

12 The predicative use of property-denoting nouns happens to be highly restricted in English, but is used somewhat more freely in, for instance, the other Germanic languages and the Romance languages (Stowell 1991: 50; Hengeveld 1992: 132; De Swart et al. 2005: 447–448, 2007: 219; Roy 2013: 37).
Incorporation: Constraints on variation

(19) Entity-designating noun used as a predicate

Tom and Ron are fine doctors.
(Smit 2005: 103)

(20) Entity-designating noun used as an adjunct

John cut the meat with a knife.
(Hengeveld and Mackenzie 2008: 208)

Crucially, property-denoting nouns contrast with entity-designating nouns in terms of modifiability. Property-denoting nouns can generally not be modified (Hengeveld and Mackenzie 2008: 230–231; see also Stowell 1991: 50–51; De Swart et al. 2005: 452), as shown in example (21).\(^\text{13}\)

(21) Modification of property-denoting nouns

a. Roosevelt was (#good) president.
   (Stowell 1991: 51)

b. They think I’ve left (#busy) town.
   (Stvan 2009: 329)

c. Her father is in (#crowded) prison for murdering her mother.
   (Stvan 2009: 330)

By contrast, entity-designating nouns can combine with both grammatical and lexical modifiers.\(^\text{14}\) Grammatical modifiers may take the form of grammatical number, possessive and diminutive marking, grammatical quantifiers and demonstratives (Smit 2005: 103; Hengeveld and Mackenzie 2008: 245–247; Rijkhoff 2008: 795; Keizer 2015: 158–159). Lexical modifiers include adjectives, restrictive relative clauses, participial clauses, possessive modifiers, adpositional phrases and lexical numerals (Smit 2005: 103; Hengeveld and Mackenzie 2008: 241–245; Rijkhoff 2008: 794–797; Keizer 2015: 156–158). Note that the expression of these modifiers as bound morphemes forming a single word with the noun they modify, as independent words within the noun phrase of the modified noun or as separate appositive noun

\(^{13}\) Property-denoting nouns only allow a very restricted type of modification (Hengeveld and Mackenzie 2008: 230; Rijkhoff 2008: 790–791, 793–794; see also De Swart et al. 2005: 452–453; Broekhuis and Den Dikken 2012: 1097–1098). This type of modification specifically pertains to the property that is ascribed rather than to a designated entity. For instance, the modifier provisional in example (vi) only modifies the property ‘president’ and does not modify an individual entity.

\(^{14}\) Proper names are exceptional in that they designate entities but are, in most languages, nevertheless highly restricted in terms of modification (Payne 1997: 39; Hengeveld and Mackenzie 2008: 237).
phrases is a morphosyntactic issue that is not relevant for the semantic modifiability of nouns (Hengeveld and Mackenzie 2008: 115, 297–298).

With respect to grammatical possessive marking and lexical possessive modifiers, a difference has to be made between alienable possessors and inalienable possessors (Hengeveld and Mackenzie 2008: 243, 306–307; see also Rijkhoff 2002: 87; Nikolaeva and Spencer 2012: 207; Von Prince 2016: 71). The alienable possessor student in example (22) is an optional addition to the entity-designating noun teacher and can as such be called a possessive modifier.

(22) Alienably possessed entity
The student’s teacher

By contrast, the possessive relationship between the inalienably possessed noun brother and its possessor king in (23) is inherent to the property of being a brother: the king’s brother is only a ‘brother’ because of the specific family relationship with the king.

(23) Inalienably possessed entity
The king’s brother

The possessor king is therefore not an optional modifier of the entity-designating noun brother but rather an obligatory argument of the property brother that is here used to designate the entity. Correspondingly, property-denoting nouns may express inalienable but not alienable possession, as shown in (24) and (25), in which the nouns brother and teacher are used as non-modifiable property-denoting nouns.

(24) Property-denoting noun expressing an inalienably possessed entity
He is brother of the king.
(Hengeveld and Mackenzie 2008: 239)

(25) Property-denoting noun expressing an alienably possessed entity
#Mary is teacher of the student.

Because inalienable possessors are considered arguments of properties rather than modifiers of entities in FDG, in this study they are not taken as evidence for the modifiability of nouns.

4.2.3 Pragmatic-semantic typology of common nouns in FDG
Based on the FDG approach to referentiality and modifiability, a three-way typology of pragmatic-semantic usages of common nouns with different combinations of
Incorporation: Constraints on variation

referentiality and modifiability characteristics can be proposed (Smit 2005: 102–103; see also Genee 2018: 258–259). This typology is presented in Table 1.

Table 1. Typology of common nouns based on referentiality and modifiability.

<table>
<thead>
<tr>
<th>Modifiable (entity-designating)</th>
<th>Referential</th>
<th>Non-referential</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+R/+M, examples in [26])</td>
<td>+</td>
<td>(+R/+M, examples in [27])</td>
</tr>
<tr>
<td>Non-modifiable (property-denoting)</td>
<td>−</td>
<td>+ (−R/−M, examples in [28])</td>
</tr>
</tbody>
</table>

Prototypically, nouns are pragmatically referential and semantically modifiable, i.e. entity-designating (Hengeveld 2008: 46). Such nouns, abbreviated here as +R/+M nouns, are exemplified in (26).

(26) Referentially used modifiable nouns (+R/+M nouns)
   a. *We saw a lovely cottage, yesterday. We are thinking of buying it.*
      (Keizer 2015: 91)
   b. *The tall president waved to the crowd.*
      (adapted from Hengeveld and Mackenzie 2008: 197)

The possibility to refer back to the noun *cottage* as used in (26a) by means of the anaphor *it* shows that this noun is used referentially. In addition, the presence of the adjectival modifier *lovely* indicates the modifiability of this noun *cottage*. The noun *president* in (26b) is also a referentially used modifiable noun, as demonstrated by the definite article *the* and the adjectival modifier *tall*.

A second pragmatic-semantic possibility for nouns is to be used as non-referential but modifiable entity-designating nouns, i.e. −R/+M nouns, as demonstrated in example (27).

(27) Non-referentially used modifiable nouns (−R/+M nouns)
   a. *Hannah is an intelligent girl, and so is Lucy.*
      (based on Hengeveld and Mackenzie 2008: 193)
   b. *Tom and Ron are fine doctors, That’s what they are.*
      (adapted from Smit 2005: 103)

It should be emphasized that the typology of nouns presented in this subsection holds only for common nouns. Proper names are referentially used entity-designating nouns but are, in most languages, non-modifiable, as mentioned in fn. 14 above. Correspondingly, they neither really fit the +R/+M type nor the unattested −R/−M type. As proper names can generally not be incorporated (Mithun 1984: 864; Massam 2009: 1090; Borik and Gehrke 2015: 5), focusing on common nouns seems sufficient for the present study. The exceptional cases of proper name incorporation, found in the present study’s sample languages Kalaallisut (Sadock 1980: 314) and Ute-Southern Paiute (Givón 2013: 322–323), are thus not considered here.
The nouns girl and doctors in example (27) are used non-referentially: they can only be referred back to by using special anaphors such as so and that, which relate to ascribed properties and entities rather than to entities that are evoked as referents. The adjectival modifiers intelligent and fine indicate that girl and doctors are modifiable. Moreover, the noun doctors includes plural marking, which also shows the modifiability of that noun.

Thirdly, nouns may be non-referentially used non-modifiable nouns. For these property-denoting nouns the abbreviation −R/−M can be used. Two examples are shown in (28).

(28) Non-referentially used non-modifiable nouns (−R/−M nouns)
   a. Roosevelt was (#that good) president.
      (adapted from Stowell 1991: 51)
   b. Her father is in (#crowded) prison, for murdering her mother. #He has been in it, for six years.
      (Stvan 2009: 326, 330)

In example (28a), the noun president cannot be combined with the demonstrative modifier that without changing its pragmatic referentiality: without the demonstrative, president is used to ascribe the property ‘president’ to Roosevelt, i.e. it is used non-referentially, whereas the addition of the demonstrative gives the noun president a referential function and makes it co-referential with Roosevelt. The use of the adjectival modifier good is not possible either, i.e. president in (28a) is also non-modifiable. In (28b), the noun prison cannot be referred back to by the anaphoric pronoun it and cannot be modified by the adjectival modifier crowded, which shows that it is also a non-referentially used non-modifiable noun.

The typology presented in Table 1 and the examples in (26)–(28) show that there is no one-to-one relation between the pragmatic referentiality and semantic modifiability of nouns. On the one hand, referentially used nouns are always modifiable because non-modifiable, property-denoting nouns necessarily function non-referentially (Smit 2005: 102): they do not represent entities that can be referred to. On the other hand, non-referentially used nouns may either be modifiable or non-modifiable.

4.3 Method

4.3.1 Sampling procedure
The variation in the referentiality and modifiability of incorporated nouns is examined in this chapter on the basis of a sample of 21 languages. As the study focuses on noun incorporation, the sampling procedure was designed to include incorporating

Referentiality and modifiability of incorporated nouns
languages only. On the basis of previous research a list of incorporating languages was set up, from which the sample could be drawn. The sources used to compile this list were Velupillai’s (2012b) survey of languages with noun incorporation, which is based primarily on Mithun (1984), Gerdz (1998) and Aikhenvald (2007); several theoretical studies on incorporation (Sapir 1911; Sadock 1980, 1985, 1986; Baker 1988, 1996; Rosen 1989; Anderson 2000); overview articles on incorporation including Mithun (1994, 2010), Iturrio Leza (2001), Anderson (2007) and Massam (2009); and the typological incorporation studies by Caballero et al. (2008), Štekauer et al. (2012) and Barrie and Mathieu (2016). In addition, languages were added to the list on the basis of a search of the Linguistic bibliography (Bobyleva et al. n.d.) and the Modern Language Association international bibliography. This procedure resulted in a list of 248 languages that were described as showing incorporation. This list is included in Appendix 1.\(^\text{1}\)

Subsequently, a 30-language variety sample was drawn from this list. Variety samples are aimed to include as much of the existing linguistic variation regarding a particular linguistic phenomenon as possible in order to enable a cross-linguistic exploration of that phenomenon (Rijkhoff and Bakker 1998: 265; Croft 2003: 21; Bakker 2011: 104; Velupillai 2012a: 50). A variety sample generally includes genealogically, geographically and typologically diverse languages, as it is assumed that this way of sampling yields the highest chance of capturing all existing variation when little is known about the variation regarding the phenomenon under consideration (Croft 2003: 21; Hengeveld 2006: 46–47; Velupillai 2012a: 50; Moravcsik 2013: 18). Correspondingly, the sampling procedure in the current research also took into account the genealogical background, geographical distribution and typological properties of incorporating languages.

For both the genealogical and the geographical diversity the language classification from Glottolog (Hammarström et al. 2017) was used. The 248 languages in the list of incorporating languages belong to 79 different language families, including ten isolates. The sample languages were selected in such a way that the sample contains languages from 30 different families and thus shows genealogical variation. Regarding the geographical diversity, the spread of the languages over the six macro-areas identified in Glottolog (Africa, Australia, Eurasia, North America, Papunesia and South America) was considered. Importantly, incorporating languages are not distributed evenly over the different macro-areas. As the geographical distribution of the sample languages was aimed to reflect the spread of the 79 incorporating language families over the world, the six macro-areas are not all represented by an equal number of languages. The calculation of the proportion of

\(^1\) The list in Appendix 1 includes 11 additional languages that were identified as incorporating languages after the sample of the study presented here was drawn. These languages, marked with “\(^1\)”, were not taken into account in the sampling procedure of the present study.
Referentiality and modifiability of incorporated nouns

language families in the list of incorporating languages in each macro-area and the corresponding number of languages from each area in the sample is presented in Appendix 2.

The typological feature taken into consideration concerns the variation in the forms of incorporated nouns. Although some assume that only nominal stems can be incorporated (Baker 1988: 71–72; Gerdts 1998: 85; Mithun 2000: 917), currently incorporated derived or compounded nouns, nominal inflected words and noun phrases also receive attention in cross-linguistic studies (Iturrioz Leza 2001; Aikhenvald 2007; Muro 2009; Barrie and Mathieu 2016; see also Chapter 3). As more complex nominal forms seem to be more rarely involved in incorporation (Smit 2005: 94; Aikhenvald 2007: 12–13), four languages known to allow highly complex incorporated nominal elements, i.e. Bininj Kun-Wok (Hengeveld and Mackenzie 2008: 415), Chukchi (Hengeveld and Mackenzie 2008: 415–416), Crow (Barrie and Mathieu 2016: 33–34) and Eastern Ojibwa (Barrie and Mathieu 2016: 17–18), were deliberately included to make the sample typologically varied.

In addition to the genealogical background, geographical distribution and typological properties of the languages, the definition of noun incorporation used in the present study, to be introduced in the next subsection, was important in the sampling procedure. Authors use highly different definitions of noun incorporation (Massam 2009: 1077; Murasugi 2014: 284; Haugen 2015: 414; Johns 2017), such that the list of 248 incorporating languages includes languages with greatly varying incorporation-like constructions. For the sample, however, only languages were selected that show constructions that can be considered to involve noun incorporation according to the present study’s definition.

Finally, the amount of available data was taken into account in the sampling procedure. 30 languages were selected for which a reasonable set of data sources, including reference grammars and articles on incorporation, could be used. Additionally, experts on several of the relevant languages were consulted during the data collection. However, the process of data collection showed that for nine of the 30 sample languages the available data on the referentiality and modifiability of incorporated nouns was ultimately insufficient for the data analysis. Consequently, the results presented in this chapter only concern the remaining 21 languages. Table 2 shows the sample and additionally indicates for which languages sufficient data were available to include them in the study. The data sources consulted for these 21 languages are listed in Appendix 4.

Although three of the four languages that were included because of their formally complex incorporated elements, i.e. Chukchi, Crow and Eastern Ojibwa, had to be excluded from the final sample due to insufficient data, the sample languages still show varied forms of incorporated nouns, including simple stems, derived or compounded stems, inflected words and morphosyntactic phrases (see Chapter 3).
Table 2. 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>
<th>Sufficient data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bininj Kun-Wok [Bininj Gun-wok, Mayali]</td>
<td>Gunwinyguan</td>
<td>Australia</td>
<td>Australia</td>
<td>Yes</td>
</tr>
<tr>
<td>Chimalapa Zoque</td>
<td>Mixe-Zoque</td>
<td>North America</td>
<td>Mexico</td>
<td>No</td>
</tr>
<tr>
<td>Chukchi</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia</td>
<td>Russian Federation</td>
<td>No</td>
</tr>
<tr>
<td>Crow</td>
<td>Siouan</td>
<td>North America</td>
<td>United States</td>
<td>No</td>
</tr>
<tr>
<td>Eastern Ojibwa</td>
<td>Algic</td>
<td>North America</td>
<td>Canada</td>
<td>No</td>
</tr>
<tr>
<td>Halkomelem</td>
<td>Salishan</td>
<td>North America</td>
<td>Canada; United States</td>
<td>No</td>
</tr>
<tr>
<td>Hokkaido Ainu [Ainu, Southern Hokkaido-Ainu]</td>
<td>Ainu</td>
<td>Eurasia</td>
<td>Japan</td>
<td>Yes</td>
</tr>
<tr>
<td>Iraqw</td>
<td>Afro-Asiatic</td>
<td>Africa</td>
<td>Tanzania, United Republic of</td>
<td>Yes</td>
</tr>
<tr>
<td>Kalaallisut [Eskimo, Greenlandic, West Greenlandic]</td>
<td>Eskimo-Aleut</td>
<td>Eurasia</td>
<td>Greenland</td>
<td>Yes</td>
</tr>
<tr>
<td>Ket</td>
<td>Yeniseian</td>
<td>Eurasia</td>
<td>Russian Federation</td>
<td>Yes</td>
</tr>
<tr>
<td>Mapudungun [Mapuche]</td>
<td>Araucanian</td>
<td>South America</td>
<td>Argentina; Chile</td>
<td>Yes</td>
</tr>
<tr>
<td>Marithiel</td>
<td>Western Daly</td>
<td>Australia</td>
<td>Australia</td>
<td>No</td>
</tr>
<tr>
<td>Mohawk</td>
<td>Iroquoian</td>
<td>North America</td>
<td>Canada; United States</td>
<td>Yes</td>
</tr>
<tr>
<td>Movima</td>
<td>Movima (Isolate)</td>
<td>South America</td>
<td>Bolivia, Plurinational State of</td>
<td>Yes</td>
</tr>
<tr>
<td>Nadeb</td>
<td>Nadahup</td>
<td>South America</td>
<td>Brazil</td>
<td>Yes</td>
</tr>
<tr>
<td>Nieuw</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Nue</td>
<td>Yes</td>
</tr>
<tr>
<td>Northern Gumuz</td>
<td>Gumuz</td>
<td>Africa</td>
<td>Ethiopia; Sudan</td>
<td>Yes</td>
</tr>
<tr>
<td>Palikūr [Palikur]</td>
<td>Arawakan</td>
<td>South America</td>
<td>Brazil; French Guiana</td>
<td>Yes</td>
</tr>
<tr>
<td>Panare</td>
<td>Cariban</td>
<td>South America</td>
<td>Venezuela, Bolivarian Republic of</td>
<td>Yes</td>
</tr>
<tr>
<td>Paraguayan Guaraní [Guarani]</td>
<td>Tupian</td>
<td>South America</td>
<td>Argentina; Paraguay</td>
<td>Yes</td>
</tr>
<tr>
<td>Sora</td>
<td>Austroasiatic</td>
<td>Eurasia</td>
<td>India</td>
<td>Yes</td>
</tr>
<tr>
<td>South Slavey</td>
<td>Athapaskan-Eyak-Tingit</td>
<td>North America</td>
<td>Canada</td>
<td>No</td>
</tr>
<tr>
<td>Southern Tiwa</td>
<td>Kiowa-Tanoan</td>
<td>North America</td>
<td>United States</td>
<td>No</td>
</tr>
<tr>
<td>Ute-Southern Paiute [Ute]</td>
<td>Uto-Aztecan</td>
<td>North America</td>
<td>United States</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Referentiality and modifiability of incorporated nouns

<table>
<thead>
<tr>
<th>Washo</th>
<th>Washo (Isolate)</th>
<th>North America</th>
<th>United States</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western Frisian [Frisian]</td>
<td>Indo-European</td>
<td>Eurasia</td>
<td>Netherlands</td>
<td>Yes</td>
</tr>
<tr>
<td>Western Highland Chatino</td>
<td>Otomangean</td>
<td>North America</td>
<td>Mexico</td>
<td>No</td>
</tr>
<tr>
<td>Yimas</td>
<td>Lower Sepik-Ramu</td>
<td>Papuasia</td>
<td>Papua New Guinea</td>
<td>Yes</td>
</tr>
<tr>
<td>Yucatec Maya [Maya Yucateco, Yucatec Mayan]</td>
<td>Mayan</td>
<td>North America</td>
<td>Belize; Guatemala; Mexico</td>
<td>Yes</td>
</tr>
</tbody>
</table>

4.3.2 Definition of noun incorporation

As studies on noun incorporation do not always target the same set of constructions, a precise definition of the phenomenon of noun incorporation must be given that can be used to select the relevant constructions from the different sample languages in a systematic way. The present study defines noun incorporation on the basis of both semantic and morphosyntactic characteristics. Semantically, incorporated nouns and the verbs in which they are incorporated are in a dependency relation of the form head-modifier or predicate-argument (see also Mithun 2000: 917; Haugen 2015: 414–415). A head-modifier relation between an incorporated noun and an incorporating verb is shown in example (29) from Bininj Kun-Wok, in which yaw ‘baby, child’ modifies ni ‘sit’.

(29) Noun incorporation construction showing a head-modifier relation between noun and verb in Bininj Kun-Wok

*Birri-yaw-ni.  
3.AU-baby/child-sit.PST.JPFV
‘They sat down like children.’
(Evans 2003: 484)

A predicate-argument relation is exemplified in (30) from Mapudungun. Here, *waka* ‘cow’ functions semantically as an argument of the predicate *kintu* ‘seek’.

(30) Noun incorporation construction showing a predicate-argument relation between noun and verb in Mapudungun

*Ñí chao kintu-waka-le-y.*  
my father seek-cow-PROG-3SG.SBJ.IND
‘My father is looking for the cows.’

Morphosyntactically, an incorporated noun forms a single word with its incorporating verb (see also Caballero et al. 2008: 385). There is no general agreement about the criteria for morphosyntactic or grammatical words (Dixon and Aikhenvald 2002: 18–25; Haspelmath 2011: 38–59, 2018: 313–314, 317), but the present study uses three
types of evidence for the morphosyntactic word status of incorporation constructions. Firstly, verbal inflectional affixes and verbal pro- and enclitics may show that a noun is incorporated, as a noun appearing between a verbal affix or clitic and the stem of a verb must be incorporated into this verb (see also Caballero et al. 2008: 385). Thus, in example (31) from Yucatec Maya the position of the noun che’ ‘tree’ relative to that of the verbal stem ch’ak ‘cut’ and the verbal suffixes -nah and -en indicates that it is part of the morphosyntactic verbal word, i.e. is incorporated.

(31) Incorporated noun followed by verbal suffixes in Yucatec Maya

\[h \text{ ch’ak-che’-nah-en ichil in kòol} \]

\[\text{PST cut-tree-COMPL-1SG.ABS in 1SG.POSS milpa} \]

‘I chopped trees in my cornfield.’


Similarly, the Nadëb verbal inflectional proclitic ta= demonstrates that, in example (32), tü ‘food’ is morphosyntactically incorporated into the verb with the stem tii ‘fish’ (Weir 1990: 331).

(32) Incorporated noun preceded by a verbal proclitic in Nadëb

\[ta= \text{tü i-tii} \]

\[3SG=food ASP-fish \]

‘He is fishing his (i.e. someone else’s) food.’

(Weir 1990: 331)

Secondly, some languages make use of morphosyntactically bound verbs that obligatorily attach to a noun that functions as their argument or modifier. When a noun is directly preceded or followed by such a verb, it is here considered to be incorporated. For instance, many Kalaallisut verbs, including -si ‘receive’, are bound (Fortescue 1980, 1984: 320–324), such that allagar ‘letter’ in example (33) must be incorporated.

(33) Noun incorporated into a bound verb in Kalaallisut

\[Fari-mit allagar-si-vuq \]

\[Fari-ABL.SG letter-receive-3SG.IND \]

‘He got a letter from Fari.’

(Fortescue 1984: 214)

\[17 \text{ This criterion for morphosyntactic wordhood of incorporation constructions was used to verify that a particular language makes use of incorporation. In individual constructions, the affixes or clitics used to identify morphosyntactic incorporation may be lacking, for instance because the paradigm of the relevant inflectional features includes cells with zero-marking or the relevant inflectional features are not expressed in all contexts.} \]
Some authors claim that verbs like -si should be analyzed as derivational affixes rather than as verbs and, correspondingly, call constructions with these affixes denominal verbs (Sapir 1911: 254; Mithun 1986a: 32; Gerdts 1998: 97–98; Kurebito 2001; Stonham 2008: 513–514). However, the present study recognizes both unbound and bound verbal morphemes (Hengeveld and Mackenzie 2008: 404; see also Delahunty and Garvey 2010: 132), such that bound elements with a highly lexical, verbal meaning, like -si in (33), are considered incorporating verbs. In this respect, the present study also follows Sadock (1980, 1985), Caballero et al. (2008: 409), Barrie and Mathieu (2016: 10), Johns (2017) and several others (see Massam 2017).

Thirdly, in several languages, incorporated nouns have a special form that is only used in the context of noun incorporation (Mithun 1984: 875–876; Caballero et al. 2008: 387–388). This form can then show that a noun is morphosyntactically incorporated into the verb. Thus, in Halkomelem the noun qeq ‘baby’ takes the form -əyeɫ when it is incorporated, which means that -əyeɫ in example (34) is an incorporated noun.

(34) Incorporated noun with a special morphosyntactic form in Halkomelem

\[ niʔ ʃkʷ-əyeɫ \text{ to } Mary. \]
\[ AUX \text{ bathe-baby DET Mary } \]
\[ ‘Mary bathed the/a baby.’ \]

Elements like -əyeɫ in (34) are sometimes argued to be derivational affixes rather than suppletive forms of independent nouns (Sapir 1911: 251–252; Gerdts 1998: 94–97; Bischoff 2011: 15). However, because these affixes have lexical meanings and are large in number, they are nevertheless counted as incorporated nouns in the present study, in the same way as in Aikhenvald (2007: 13), Caballero et al. (2008: 387–388) and other studies (see Massam 2017).

The definition of noun incorporation employed in this study does not specify any phonological requirements. It is therefore assumed here that a combination of incorporated noun and incorporating verb does not have to form a single phonological word (see also Mithun 1984: 849–845; Aikhenvald 2007: 14; Massam 2017). Thus, the Nadëb example in (32) above is regarded as an incorporation construction because the noun tʉ ‘food’ and the verb tɨɨ ‘fish’ form a predicate-argument relation and constitute a single morphosyntactic word, even though the noun and verb remain independent phonological words as evidenced by stress placement (Weir 1990: 323). Constructions like (32) have also been called “juxtaposition” (Mithun 1984: 849), “loose incorporation” (Miner 1986: 252) and “pseudo-incorporation” (Massam 2009: 1087), but are included within the domain of noun incorporation in the present study (see also Aikhenvald 2007: 14–15; Caballero et al. 2008: 385–386; Massam 2017).
Finally, incorporation is here considered to be a grammatical process that is distinct from lexical compounding (Hengeveld and Mackenzie 2016: 1150–1153; see also Baker 1988: 78–80, 1996: 307–308; Barrie and Mathieu 2016: 4–5, 23). Incorporation is as such restricted to productive processes that yield verbs with semantically predictable meanings, whereas lexical compounding has limited productivity and creates verbs with possibly non-transparent semantics. In addition, the nominal components of lexical compounds necessarily have a non-referential function and cannot take modification, in contrast to incorporated nouns, of which the potential pragmatic referentiality and semantically modifiability is examined in this chapter. Constructions that classify as lexical compounds are thus excluded from the study.

4.3.3 Data analysis

In this study, the referential potential and modification possibilities of incorporated nouns in the sample languages were examined on the basis of the available examples of incorporation constructions in these languages and descriptive information about their characteristics. For each language it was investigated to which noun types its incorporated nouns belong: +R/+M nouns, –R/+M nouns and/or –R−M nouns. As the present study does not only focus on cross-linguistic variation but also on intra-linguistic variation, for each language it was necessary to find either positive or negative evidence for the existence of incorporated nouns of each of the three noun types.

Evidence for the occurrence of referentially used incorporated nouns in a particular language consisted of at least one grammatical example of an incorporated noun showing at least one of the characteristics of referentially used nouns described in Section 4.2.1. By contrast, a language was considered to show incorporated nouns with a non-referential function if at least one ungrammatical example of an incorporated noun showing at least one of these characteristics was found. The relevant characteristics include availability for anaphoric reference, ability to appear in a co-referential relation with another noun, ability to combine with an element marking the entity evoked by the noun as identifiable for the addressee and ability to combine with a referential possessor. The possibility vs. impossibility for incorporated nouns to show one or more of these characteristics could also be verified on the basis of descriptive information given by the language expert. In addition, if

---

18 Availability for anaphoric reference is no indisputable proof of the referential status of a noun, as non-referentially used nouns can sometimes function as antecedents of special anaphors that refer back to ascribed properties or entities. Moreover, the use of anaphors can be based on bridging, in which case a regular anaphor is not related to an explicit linguistic antecedent but is interpreted on the basis of the context and/or on the speaker’s prior or general world knowledge (Clark 1975; Hengeveld and Mackenzie 2008: 120; see also Mithun 1984: 871). In the data analysis, anaphoric reference is therefore only considered as evidence for the referential status of an incorporated noun when the anaphor used is a regular anaphor that unambiguously points at the entity evoked by the incorporated noun.
Referentiality and modifiability of incorporated nouns

Descriptive information was available stating that incorporated nouns in the language could or could not have a specific or definite interpretation, this information was also used to determine the pragmatic referentiality of incorporated nouns in the language. Regarding modifiability, the presentation of at least one grammatical example of an incorporated noun combined with a grammatical or lexical modifier was used as evidence for the occurrence of modifiable incorporated nouns, while an ungrammatical example of an incorporated noun with such a modifier was taken to show that a language makes use of non-modifiable incorporated nouns. The grammaticality or ungrammaticality of the use of such modifiers could also be determined based on descriptive information.

It should be emphasized that in this study, modifiability is considered a semantic issue, such that the morphosyntactic characteristics of apparent modifiers of incorporated nouns are not relevant for the issue of modifiability. Modifiers usually appear either morphologically attached to the noun they modify or syntactically adjacent to it in a noun phrase. Correspondingly, some languages allow modifiers to incorporate together with the noun they modify and to appear next to it in the incorporation construction, as shown by the adjectival modifier kurii ‘black’ in example (35) from Mapudungun, the plural infix in example (36) from Nuu-chah-nulth and the possessive suffix -mi in example (37) from Kalaallisut.

(35) Incorporation of a noun and its lexical modifier in Mapudungun
kurii-wentru-feye-l
black-man-believe-APPL.
‘believe someone to be a black man’
(Salas 1992: 197; translation from Spanish and glosses based on Smeets 2008: 521, 573; Zúñiga 2017: 709)

(36) Incorporation of a noun and its grammatical modifier in Nuu-chah-nulth
ia:<PL>ha-na k-’aƛ’-’at-qu:
child<PL>-having-TEL-SHIFT-2SG.COND
‘when you have children’
(Nakayama 2001: 64, 2014: 454)

The classification of incorporated nouns in a particular language as indefinite or non-specific was not considered sufficient evidence for referentiality, even though referentially used nouns can be indefinite and/or non-specific, because some authors may not distinguish between indefinite or non-specific referentially used nouns on the one hand and non-referentially used nouns on the other hand in the same way as FDG.

Note that a few languages, such as Chukchi (Spencer 1995: 477) and Ute-Southern Paiute (Givón 2011: 194–196, 199–200) are known to productively incorporate lexical modifiers into nouns. As these adjectives modify the entities designated by the nouns, the incorporation of a noun with such an incorporated adjective still counts as an example of the incorporation of a modifiable noun.
Incorporation: Constraints on variation

(37) Incorporation of a noun with its grammatical modifier in Kalaallisut

\[
\text{illu-mi-nilp-puq} \\
\text{house-REFL.POSS-be.in-3SG.IND}
\]

‘He is in his (own) house.’

(Fortescue 1984: 300–301)

Such incorporated modifiers are clear evidence for the modifiability of the incorporated noun.

In many other languages, by contrast, incorporation appears to be restricted to nominal stems, such that modifiers may not be incorporated. However, in some of these languages, lexical modifiers that appear to relate to the incorporated noun can occur external to the incorporation construction. Thus, in example (38) from Bininj Kun-Wok the demonstrative \textit{na-mekke} looks like a modifier of the incorporated noun \textit{murrng} ‘bone’, and in example (39) from Mohawk the relative clause \textit{nehneh a-ak-ahninuʔ} ‘that she would buy’ seems to modify the incorporated noun \textit{nuhs} ‘house’.

(38) Noun incorporation construction with an external demonstrative in Bininj Kun-Wok

\[
N\text{ga-murrng-bimbom na-mekke.} \\
1\text{-3-bone-paint.PST.PFV M-DEM}
\]

‘I painted those bones.’

(Evans 2003: 235)

(39) Noun incorporation construction with an external relative clause in Mohawk

\[
K\text{a-nuhs-rakv nehneh a-ak-ahninuʔ} \\
3.N-house-white that INDF-3.F-buy
\]

‘The house that she would buy is white.’


Importantly, there is disagreement in the literature on noun incorporation about the interpretation of such external modifiers. Whereas some claim that these modifiers indeed modify the incorporated nouns (Sadock 1980: 307–310, 1991: 91–99; Baker 1988: 92–105, 1996: 308; Van Geenhoven 1998: 17–22; Barrie and Mathieu 2016: 4), others maintain that such modifiers constitute separate noun phrases without a nominal head that are completely independent of the presence of an incorporated noun (Mithun 1984: 865–866, 870; Di Sciullo and Williams 1987: 65–66; Rosen 1989: 298; Barrie 2010: 293–294). As the present study separates the morphosyntactic position and expression of modifiers from the issue of modifiability, which concerns semantics, it is considered irrelevant whether a modifier is incorporated together with the noun it modifies or appears external to the incorporation construction: both types
of modifiers are regarded as evidence for the modifiability of incorporated nouns.\textsuperscript{21} Interestingly, however, only for two sample languages the inclusion of external modifiers turned out to affect the results (see Section 4.4).

One special type of external modifier is exemplified by the second person pronoun õm from Nadëb in (40b).

(40) Noun incorporation construction with an external possessor in Nadëb
a. a mooh ũh=hi-jaxit
   2SG.POSS hand 1SG=TH.ASP-wash
   ‘I wash your hands.’

b. õm ũh=moox ū=h=jaxit
   2SG 1SG=hand TH.ASP-wash
   ‘I wash your hands.’ (lit. ‘I hand-wash you.’)
   (Weir 1990: 324)

The unincorporated noun moox ‘hand’ in (40a) is combined with the possessive pronoun a. By contrast, in example (40b), in which moox is incorporated into the verb, this possessive pronoun is replaced by the regular pronoun õm, which is morphosyntactically the absolutive argument of the incorporating verb (Weir 1990: 323). Nouns and pronouns designating apparent possessors that are expressed as arguments of main verbs, like õm in (40b), are known as external possessors (Payne and Barshi 1999: 3; Herslund and Baron 2001: 14–15; Aikhenvald 2013: 36).

External possessors that appear in the context of noun incorporation can be analyzed in different ways. Some researchers claim that, despite their special morphosyntactic expression, these external possessors constitute semantic units together with the nouns that designate the entities that they are assumed to possess (Gerdt 2003: 352–355; Van de Velde 2013: 172–173; see also Allen et al. 1984: 306–307; Baker 1988: 96–105, who argue that the external possessors and possessed nouns form a single unit underlyingly). Such an analysis entails that an external possessor can be considered to show that an incorporated noun is modifiable. Alternatively, however, external possessors like õm in (40b) may simply be analyzed as arguments of incorporating verbs that are independent of the incorporated nouns involved (Mithun 1984: 856, 859), in which case external possessors do not form evidence for the modifiability of incorporated nouns.

Importantly, external possession is typically limited to inalienable possessive relations (Herslund and Baron 2001: 15; Aikhenvald 2013: 36). As nouns designating inalienable possessors are possessive arguments of properties rather than possessive modifiers of entities in FDG (see Section 4.2.2), inalienable external possessors used

\textsuperscript{21}In the case of demonstrative modifiers and referentially used alienable possessive modifiers, the modifiers are additionally taken as evidence for the referential status of incorporated nouns.
Incorporation: Constraints on variation

in the context of noun incorporation do not provide evidence for the modifiability of incorporated nouns, irrespective of the choice between the two possible analyses of external possessors just described. Because the data from the sample languages indeed only show inalienable external possessors, for the question of modifiability the present study can simply leave external possessors aside. However, whether or not external possessors can present evidence for the referentiality of incorporated nouns depends on the pragmatic analysis of these possessors. Because external possessors are often argued to express a special affectedness or empathy on the side of the speaker (Van de Velde 2013: 167, referring to O’Connor 2007), they may be considered to differ from regular attributive possessive nouns and pronouns in terms of their pragmatics. Regular attributive possessive nouns and pronouns form a single pragmatic unit with their possessed entities. Correspondingly, they have a shared referential status, in that a noun corresponding to an entity possessed by a referentially used attributive possessive noun or pronoun necessarily also has a referential function (see Section 4.2.1). By contrast, it seems suitable to analyze external possessors as independent pragmatic units, as they can be specified separately for pragmatic affectedness or empathy (Van de Velde 2013: 172–173). Their possible referential pragmatic status is then also independent of the referentiality of the incorporated possessed nouns, and external possessors are therefore not taken to provide evidence for the referentiality of incorporated nouns either, even if these possessors are used referentially.

4.4 Results

The present study investigates the range of cross- and intra-linguistic variation that can be found with respect to the referential potential and modification possibilities of incorporated nouns based on the FDG notions of referentiality and modifiability. More specifically, it is examined whether languages show +R/+M incorporated nouns, −R/+M incorporated nouns and/or −R/−M incorporated nouns. Table 3 presents the results of the investigation of the pragmatic referentiality and semantic modifiability of incorporated nouns, showing to which pragmatic-semantic noun types incorporated nouns in the sample languages belong.22

Table 3 demonstrates that each of the three pragmatic-semantic types of nouns is found in incorporation constructions in the sample languages and that there is both cross- and intra-linguistic variation regarding the referential potential and modification possibilities of incorporated nouns: both +R/+M incorporated nouns, −R/+M incorporated nouns and −R/−M incorporated nouns are found in a subset of the sample languages, and several sample languages show incorporated nouns of more than one pragmatic-semantic type.

22 The full set of data on which Table 3 is based can be found on http://dx.doi.org/10.21942/uva.7172012.
Table 3. The occurrence of incorporated nouns of the different pragmatic-semantic types in the sample languages. The pragmatic-semantic noun types distinguished are referentially used modifiable nouns (+R/+M), non-referentially used modifiable nouns (−R/+M) and non-referentially used non-modifiable nouns (−R/−M). “+” means that incorporated nouns in the language can be of the relevant type, while “−” shows that incorporated nouns of this type do not occur in the language.

<table>
<thead>
<tr>
<th>Language</th>
<th>+R/+M</th>
<th>−R/+M</th>
<th>−R/−M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bininj Kun-Wok</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Hokkaido Ainu</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Iraqw</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Kalaallisut</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ket</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Mapudungun</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Mohawk</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Movima</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Nadeb</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Niuecan</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Northern Gumuz</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Nuu-chah-nulth</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Palikur</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panare</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Paraguayan Guaraní</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Sora</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Ute-Southern Paiute</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Washo</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Western Frisian</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Yimas</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Yucatec Maya</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
</tbody>
</table>

Examples of incorporated nouns of each of the three types are shown in example (41)–(43). Example (41) illustrates the incorporation of a +R/+M incorporated noun in a construction from Hokkaido Ainu.

(41) Incorporation of a +R/+M noun in Hokkaido Ainu

\[
\text{siknu=an} \ wa \ \text{yay-cise-ko-hosipi=an}
\]

be.alive=INDF.S and REFL-house-to.APPL-return=INDF.S

'(Thanks to the goddess) I came back to life and returned to my own house.’

(Okuda 1993, cited in Bugaeva 2010: 789)

In this example, the verb hosipi ‘return’ combines with the applicative marker ko-, such that it takes a goal object (Bugaeva 2010: 774). This object, yay-cise ‘my house’, is here incorporated. Importantly, yay-cise includes referential reflexive possessive marking (Bugaeva 2010: 792), which shows both the referentiality and modifiability of the noun.

The incorporation construction from Ket in example (42), by contrast, contains a −R/+M noun.
Incorporation: Constraints on variation

(42) Incorporation of a −R/+M noun in Ket
\[tab-aj-t-o-n-aq\]
dog.PL-3PL.AN.SBJ-TC-PST-PST-become
‘They turned into dogs.’
(Vajda 2017: 918)

The incorporated noun tab ‘dogs’ includes grammatical plural marking, which means that it is modifiable. At the same time, it functions as a non-referential, predicatively used noun.

Finally, example (43) presents a construction with a −R/−M incorporated noun from Western Frisian.

(43) Incorporation of a −R/−M noun in Western Frisian
a. *Heit sit te (*de/*in/*dy) jerappel-skilen*
father sits to DEF/INDF/DEM potato-peel
‘Father is sitting, peeling (*the/a/that/those) potatoes.’
(Dijk 1997: 44)
b. *Heit sit te (*grouwe) jerappel (*mei in soad spruten) skilen*
father sits to huge potato with INDF lot sprouts peel
‘Father is sitting, peeling (*huge) potatoes (*with a lot of sprouts).’
(Dijk 1997: 16)

In (43), the noun jerappel ‘potato’ is incorporated into the verb skilen ‘peel’. Example (43a) demonstrates that the incorporated noun jerappel cannot be combined with an element marking definiteness, while example (43b) shows the impossibility to combine the incorporated noun jerappel with an incorporated adjectival modifier grouwe ‘huge’ or an incorporated adpositional phrase mei in soad spruten ‘with a lot of sprouts’. Thus, the incorporated noun jerappel is used non-referentially and cannot be modified.

The sample languages vary systematically as to which pragmatic-semantic types of incorporated nouns they show. Based on this variation, five groups of languages can be identified, as shown in Table 4.
Table 4. Language groups identified on the basis of the possible pragmatic-semantic noun types of incorporated nouns in the sample languages. The pragmatic-semantic noun types distinguished are referentially used modifiable nouns (+R/+M), non-referentially used modifiable nouns (−R/+M) and non-referentially used non-modifiable nouns (−R/−M). “+” means that incorporated nouns in the language can be of the relevant type, while “−” shows that incorporated nouns of this type do not occur in the language.

<table>
<thead>
<tr>
<th>Group</th>
<th>Language</th>
<th>+R/+M</th>
<th>−R/+M</th>
<th>−R/−M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hokkaido Ainu</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Kalaallisut</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Ket</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Mapudungun</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>Niuean</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Nuu-chah-nulth</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Washo</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>3</td>
<td>Nadob</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Panare</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>4</td>
<td>Binj Kun-Wok</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Mohawk</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Sora</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Ute-Southern Paiute</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>Iraqw</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Movima</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Northern Gumuz</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Palikûr</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Paraguayan Guaraní</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Western Frisian</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Yimas</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Yucatec Maya</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 4 indicates that a distinction can be made between languages that allow nouns of all three pragmatic-semantic types as incorporated nouns (group 1), languages limiting noun incorporation to modifiable nouns, i.e. +R/+M nouns and −R/+M nouns (group 2), languages that only show +R/+M incorporated nouns (group 3), languages with both +R/+M and −R/−M incorporated nouns (group 4) and languages in which noun incorporation is restricted to −R/−M nouns (group 5).

This grouping of the sample languages reflects both the variation between languages and the variation within languages with respect to the referentiality and modifiability of incorporated nouns. On the one hand, each pragmatic-semantic noun type occurs only in incorporation constructions in languages in some of the groups. For instance, −R/+M incorporated nouns are only found in group 1 and 2 languages, while the incorporation of −R/−M nouns is limited to the languages in group 1, 4 and 5. Thus, the groups indicate that there is cross-linguistic variation regarding the referential potential and modification possibilities of incorporated nouns. On the other hand, languages in three of the groups in Table 4, i.e. group 1, 2 and 4, show incorporated nouns of more than one pragmatic-semantic type. For instance,
languages in group 2 show both +R/+M incorporated nouns and −R/+M incorporated nouns. Languages in some of the groups thus show intra-linguistic variation regarding the referentiality and modifiability of incorporated nouns as well.

Note finally that the results presented in Table 3 and 4 only marginally depend on the analysis of modifiers that appear external to incorporation constructions. As discussed in Section 4.3.3, the morphosyntactic position of such external modifiers does not affect their semantic status as modifiers in FDG, but in some other studies these external modifiers are not considered to modify incorporated nouns. However, although external modifiers are found in seven of the 21 sample languages, their analysis as modifiers of incorporated nouns is only decisive in the identification of −R/+M nouns in Kalaallisut and Washo. For Kalaallisut, for instance, examples like (44) are the only type of evidence for the possibility to incorporate −R/+M nouns.

(44) Incorporation of a −R/+M noun in Kalaallisut

\[ \text{savaatili-} \text{ngur-} \text{putin=} \text{nguuq} \quad \text{pikkuris-} \text{suq} \]

\[ \text{sheep-} \text{herder-become-2SG.IND=QUOT} \quad \text{be.capable-INTR.PTCP} \]

‘They say you’ve become a capable sheep-herder.’

(Fortescue 1984: 71)

In (44), the noun incorporation construction is combined with an external participle pikkuris-suq ‘capable’, which can be considered a modifier of the incorporated noun savaatili ‘sheep-herder’. For all other languages as well as for the +R/+M incorporated nouns in Kalaallisut and Washo, other forms of positive evidence were sufficient to verify the occurrence of +R/+M and −R/+M incorporated nouns. External modifiers thus only play a minimal role in the analysis of referentiality and modifiability of incorporated nouns in the present study.

4.5 Discussion

The results presented in the previous section have important implications for theoretical accounts of noun incorporation. First of all, the data indicate that −R/−M incorporated nouns, found in 16 of the 21 sample languages, and +R/+M incorporated nouns, occurring in 13 of the 21 sample languages, are both quite frequent cross-linguistically. For this reason, a comprehensive theoretical account of noun incorporation should be able to capture both the incorporation of −R/−M nouns and the incorporation of +R/+M nouns.

Most existing theoretical approaches to incorporation, however, appear to concentrate on incorporated nouns corresponding to one of these types only. Traditionally, a distinction is made between lexical approaches and syntactic approaches to noun incorporation (Massam 2009: 1083–1086, 2017; Murasugi 2014: 286–288). Researchers taking a lexical approach argue that noun incorporation is a
type of word formation or, more specifically, a type of lexical compounding (Sapir 1911; Mithun 1984; Di Sciullo and Williams 1987; Rosen 1989; Anderson 2000). Di Sciullo and Williams (1987: 68) claim that the status of incorporated nouns as parts of words predicts that they do not play a syntactic role and do as such not have a referential function and cannot be modified. This claim is in agreement with Mithun (1984), who also states that incorporated nouns do not refer and cannot combine with modifiers. Mithun (1984: 866–867, 871) argues, for instance, that incorporated nouns cannot introduce discourse referents, which suggests that they correspond to nouns that are considered pragmatically non-referential in the present study. In addition, she does not regard external modifiers as modifiers of incorporated nouns but as separate noun phrases that are independent of incorporation, and she maintains that only noun stems without inflectional definiteness or number marking can be incorporated (Mithun 1984: 847, 849, 859, 865–866, 870). The lexical approach thus appears to focus on incorporated nouns that match the −R/−M type.

By contrast, in syntactic approaches incorporated nouns are generally taken to be referential and modifiable, i.e. of the +R/+M type. Firstly, Sadock (1985: 383–384, 1991: 100–101), adopting his autolexical syntax model, proposes that incorporation constructions show a mismatch in their morphological and syntactic representations. He argues that incorporated nouns combine with incorporating verbs morphologically but retain their syntactic reality, including their referentiality and modifiability (Sadock 1985: 398–409, 1991: 86–88, 91–100). The referential characteristics that he addresses are the ability to refer to a specific entity and the ability to introduce discourse topics, i.e. the referential characteristics of incorporated nouns that he observes correspond to the pragmatic notion of referentiality used in the present study. With respect to modifiability, he recognizes modifiers that appear external to the incorporation construction, as exemplified in example (38), (39) and (44) above.

Secondly, Baker (1988, 1996, 2009) analyzes noun incorporation as syntactic head movement: the head noun of the noun phrase in a verb’s complement moves to this verb to become incorporated in it. Based on the assumption that nouns that head noun phrases are referential, Baker (1988: 81) can account for the referential characteristics of incorporated nouns that he observes. These characteristics include the ability to refer to a specific entity, the ability to appear in a co-referential relation with another noun and the ability to introduce a new entity into the discourse (Baker 1988: 78–80, 1996: 287–291), i.e. characteristics that match the FDG notion of referentiality. Moreover, Baker’s head-movement analysis can explain the occurrence of external modifiers as modifiers of incorporated nouns that are left behind when the nouns are moved to the incorporating verbs (Baker 1988: 92–105, 1996: 308).

Thirdly, Barrie and Mathieu (2016) propose that noun incorporation constructions result from phrasal movement of nominal projections. Correspondingly, they can explain the referentiality of incorporated nouns, such as their ability to function as antecedents in anaphoric reference, and the modifiability of incorporated
nouns in terms of their ability to combine with external modifiers in the same way as Baker (1988, 1996). In addition, they can account for the incorporation of nouns together with their modifying inflection or lexical modifiers. Barrie and Mathieu (2016) thus identify incorporated nouns corresponding to +R/+M nouns as defined in the present study just like Baker (1988, 1996), but recognize even more modification possibilities.

Several theoretical approaches thus appear to focus on either −R/−M incorporated nouns or +R/+M incorporated nouns, even though both are found cross-linguistically, as shown in the present study. Interestingly, this restricted focus might be related to another pattern that the current study reveals: the data suggest that the incorporation of −R/−M nouns and the incorporation of +R/+M nouns are independent of each other. Although eight of the 21 sample languages show −R/−M incorporated nouns as well as +R/+M incorporated nouns, eight other languages limit incorporation to −R/−M nouns and five languages allow the incorporation of +R/+M nouns but not the incorporation of −R/−M nouns. The incorporation of −R/−M nouns and the incorporation of +R/+M incorporated nouns could thus be two distinct processes that may, but need not occur in the same language. The lexical and syntactic approaches to noun incorporation described above then relate to only one of these processes each.

The proposal that there are two distinct noun incorporation processes, one involving +R/+M nouns and one involving −R/−M nouns, is actually supported by several theoretical approaches to incorporation, including the syntactic ones just discussed. Firstly, Sadock (1985: 398–415, 1991: 86–88, 91–99) mainly addresses the incorporation of “highly referential” and modifiable nouns in Kalaallisut and Southern Tiwa, but he also explicitly states that in other languages incorporated nouns may have different characteristics (Sadock 1986, 1991: 82–83, 99–100). Secondly, Baker (1988: 78–80, 1996: 307–308) and Barrie and Mathieu (2016: 4–5, 23) acknowledge that in addition to noun incorporation constructions created via a syntactic movement procedure, languages may show lexical noun-verb compounding constructions in which the nouns have a non-referential function and cannot be modified.23 Although these authors exclude these constructions from the domain of incorporation based on their definition of incorporation as syntactic movement, according to the definition of the present study these constructions involve incorporation as well, as long as they are semantically transparent and the process is reasonably productive. Thirdly, several semantic approaches to noun incorporation note that incorporation processes may be of different types. For instance, Chung and Ladusaw (2003) focus on the incorporation of referentially used nouns in Chamorro, but suggest that in other languages incorporated nouns may function non-referentially because the constructions are

---

23 Barrie and Mathieu (2016: 4, 23), taking a Distributed Morphology approach, consider these compounds to be the result of root-root merger.
formed via “compounding or some other morpholexical process” (Chung and Ladusaw 2003: 127).

On closer inspection, the occurrence of both $-R/-M$ incorporated nouns and $+R/+M$ incorporated nouns is also consistent with the lexical approach proposed by Mithun (1984). Mithun (1984) distinguishes between four functional types of noun incorporation: lexical compounding, incorporation that affects the argument structure of the incorporating verb, incorporation that manipulates the discourse structure of the clause and classificatory incorporation. She considers each of these to involve non-referential and non-modifiable nouns. However, Mithun’s (1984) notion of referentiality does not completely overlap with the FDG one on which the present study is based. She does not take anaphoric reference as evidence for the referential status of an incorporated noun, maintaining that the relevant anaphoric pronoun may simply refer to an entity that is not mentioned explicitly (Mithun 1984: 871). In addition, she does not analyze the occurrence of a noun with a referential function that appears to relate to the same entity as an incorporated noun, i.e. what is called a co-referential noun in the present study, as evidence for the referential function of the incorporated noun (Mithun 1984: 866, 867, 871). Moreover, as described above, Mithun (1984: 865–866, 870) does not interpret external modifiers as modifiers of incorporated nouns. These considerations are relevant for nouns in classificatory incorporation constructions, which may be followed by anaphoric pronouns, may appear to designate the same entity as a preceding or following referentially used noun phrase and may occur with external modifiers (Mithun 1984: 863–871). According to the FDG approach in the present study, classificatory incorporation may thus be analyzed as the incorporation of $+R/+M$ nouns, such that Mithun’s approach nevertheless captures $+R/+M$ as well as $-R/-M$ incorporated nouns.

On the other hand, Mithun (1984: 848, 874) also proposes an implicational relationship between the four functional types of noun incorporation which does not match the variation between $+R/+M$ noun incorporation and $-R/-M$ noun incorporation attested in the present study. According to Mithun, all incorporating languages show lexical compounding. Languages may additionally show incorporation that affects the argument structure of the incorporating verb and if they do, they may also allow incorporation that manipulates the discourse structure. Classificatory incorporation occurs only in languages that also show all other three types. In terms of the pragmatic-semantic types of incorporated nouns distinguished in the present study, this prediction means that languages only show the incorporation of $+R/+M$ incorporated nouns, i.e. classificatory incorporation, if they also show $-R/-M$ incorporated nouns, i.e. the other three types. This prediction is, however, not confirmed by the present study, which shows that $-R/-M$ noun incorporation and $+R/+M$ noun incorporation are independent of each other and that several languages show $+R/+M$ incorporated nouns without showing $-R/-M$ incorporated nouns.
While the incorporation of $+R/+M$ nouns and the incorporation of $-R/-M$ nouns appear to be independent of each other, the occurrence of $-R/+M$ incorporated nouns and $+R/+M$ incorporated nouns seems to be implicationally related: all sample languages that allow $-R/+M$ incorporated nouns also show incorporation of $+R/+M$ nouns, while the opposite pattern does not hold. This finding gives the impression that there is a cross-linguistic preference for $+R/+M$ incorporated nouns over $-R/+M$ incorporated nouns. This impression is further supported by the observation that $-R/+M$ nouns, found in only six of the sample languages, are much less frequent than $+R/+M$ nouns as well as $-R/-M$ nouns.

On the basis of the attested dependency relation between $+R/+M$ incorporated nouns and $-R/+M$ incorporated nouns, it seems suitable to consider $+R/+M$ noun incorporation and $-R/+M$ noun incorporation to belong to a single incorporation type, i.e. the incorporation of modifiable nouns, which then contrasts with the incorporation of non-modifiable nouns. This reasoning also matches the account of noun incorporation in Niuean presented by Massam (2001), one of the few studies that explicitly discusses non-referentially used modifiable nouns. Massam (2001: 192) analyzes both the incorporation of referentially used modifiable nouns and the incorporation of non-referentially used modifiable nouns in Niuean as a syntactic type of incorporation that she calls pseudo-incorporation and that involves the base generation of a noun phrase adjacent to the verb.

Another important finding of the study concerns the distribution of the different types of modifiers of incorporated nouns. Eight of the 21 sample languages only show $-R/-M$ incorporated nouns, i.e. no modification is allowed. Eight languages allow the incorporation of nouns with their modifying inflection and/or lexical modifiers. Importantly, many traditional approaches to incorporation, including the lexical approach proposed by Mithun (1984: 847, 849, 859) and the syntactic approach argued for by Baker (1988), limit incorporation to simple, uninflected stems and are as such in disagreement with the findings of the present study. Finally, seven sample languages show incorporation constructions with external modifiers. Interestingly, some languages show both incorporated and external modifiers, while others use only incorporated modifiers and again others allow only external ones. Thus, the data do not reveal any direct relationship between the appearance of external and incorporated modifiers. Note also that both incorporated and external modifiers are found in the context of $+R/+M$ noun incorporation as well as $-R/+M$ noun incorporation.

---

24 This finding contrasts with Smit’s (2005: 125–126) proposal that the incorporation of $-R/+M$ nouns is the basic type of noun incorporation that is found in all incorporating languages.

25 Another explanation for the relatively low number of languages with $-R/+M$ incorporated nouns is that such incorporated nouns may simply have received little attention in studies on incorporation because only few linguistic theories explicitly distinguish $-R/+M$ nouns (Smit 2005: 131).
Finally, the data indicate a few possible factors that may affect the possibility to use a +R/+M, −R/+M or −R/−M noun in an incorporation construction. Firstly, in some languages that allow incorporated nouns of more than one pragmatic-semantic type, there seems to be a relation between the incorporating verb and the referentiality and/or modifiability of the incorporated noun. For instance, in Kalaallisut only the verb -kar ‘go to’ and a few others allow the incorporation of +R/+M nouns with modification in the form of nominal inflection (Kristoffersen 1992: 154), in Niuean only incorporation constructions with the verb fai ‘have’ show −R/+M nouns (Massam 2001: 173–177) and in Ute-Southern Paiute only the verbs -ga ‘have’ and -’a ‘not have’ incorporate +R/+M nouns (Givón 2011: 336–340, 2018: p.c.). In these languages, lexical properties of the incorporating verbs seem to determine the referentiality and modifiability of incorporated nouns, as is also proposed by Massam (2001: 185–186, following Johns 1999) and Chung and Ladusaw (2003: 128).

Secondly, the status of an incorporated noun as semantic argument or modifier of the incorporating verb on the one hand or as semantic predicate of the incoorporating verb on the other hand plays a role here. Nominal predicates function non-referentially, such that incorporated nominal predicates are always −R/+M nouns or −R/−M nouns. In Bininj Kun-Wok, −R/−M incorporated nouns are all secondary nominal predicates (Evans 1999: 261, 2017: p.c.) and also all attested examples of −R/−M incorporated nouns in Kalaallisut are secondary nominal predicates (Fortescue 1984: 323; Kristoffersen 1992: 156). The examples of −R/+M noun incorporation in Kalaallisut (Fortescue 1984: 71), Ket (Vajda 2017: 917–920) and Mapudungun (Salas 1992: 197) also all involve predicatively used nouns.

### 4.6 Conclusion

This study has investigated the pragmatic referentiality and semantic modifiability of incorporated nouns. In order to tease apart the different views on these issues in the literature on noun incorporation, the chapter examined the cross- and intra-linguistic variation regarding the referential potential and modification possibilities of incorporated nouns in a systematic and consistent way. The FDG approach to referentiality and modifiability was applied to incorporated nouns in a sample of 21 incorporating languages, in order to determine whether languages show the incorporation of +R/+M nouns, −R/+M nouns and/or −R/−M nouns.

The data revealed a large variation between and within languages regarding the referentiality and modifiability of incorporated nouns. Both referentially and non-referentially used incorporated nouns were found and it was shown that incorporated nouns may both be modifiable and non-modifiable. More specifically, +R/+M nouns, −R/+M nouns as well as −R/−M nouns occur in incorporation constructions in the sample languages, and languages differ as to whether they show all three types of incorporated nouns, only show modifiable incorporated nouns, limit incorporation to
Incorporation: Constraints on variation

+R/+M nouns, use both +R/+M and −R/−M incorporated nouns or only incorporate −R/−M nouns. In addition, the occurrence of +R/+M incorporated nouns and the appearance of −R/−M incorporated nouns seem to be independent of each other, in that a single language may show both but may also restrict incorporation to either +R/+M or −R/−M nouns. By contrast, the incorporation of −R/+M nouns was found to be dependent on the incorporation of +R/+M nouns, i.e. languages may only show −R/+M incorporated nouns if they also allow +R/+M incorporated nouns.

The attested cross- and intra-linguistic variation with respect to the referential potential and modification possibilities of incorporated nouns may partly explain the conflicting views on the referentiality and modifiability of incorporated nouns in the literature. Because languages may limit incorporation to +R/+M nouns or to −R/−M nouns, studies on particular incorporating languages may only be able to identify +R/+M incorporated nouns or may only encounter −R/−M incorporated nouns. In addition, most theoretical approaches to noun incorporation seem to concentrate on one of the types of incorporated nouns only. Whereas lexical approaches to noun incorporation tend to describe incorporated nouns as non-referential and as being unable to take modification, most syntactic approaches to noun incorporation emphasize that incorporated nouns can be used to refer and can combine with modifiers. Moreover, differences between theoretical approaches with respect to the criteria they use for referentiality and modifiability play a role here. For instance, Mithun (1984) evaluates anaphoric pronouns that appear to relate to incorporated nouns differently from Baker (1988, 1996) and Barrie and Mathieu (2016), which affects their answers to the question if incorporated nouns function referentially. Furthermore, the definition of noun incorporation used in a theoretical approach is important. Because several syntactic approaches equate noun incorporation with a particular head-movement process, they directly exclude constructions with −R/−M nouns that are considered incorporation constructions in several other studies from the domain of noun incorporation.

The apparent independency between +R/+M incorporated nouns and −R/−M incorporated nouns on the one hand and the identified dependency between −R/+M incorporated nouns and +R/+M incorporated nouns on the other hand also have theoretical implications. Firstly, based on their independency, incorporation constructions with +R/+M nouns and incorporation constructions with −R/−M incorporated nouns seem to result from two separate incorporation processes. Secondly, because −R/+M incorporated nouns only occur in languages that also show +R/+M incorporated nouns, these two types of incorporated nouns may be classified together as the incorporation of modifiable nouns. Thus, a distinction can be made between two incorporation processes, one involving modifiable nouns and another involving non-modifiable nouns, which may, but do not have to co-occur in a single language.
As the present study has shown that the possibilities regarding the referentiality and modifiability of incorporated nouns show a large degree of cross- and intra-linguistic variation, that this cross- and intra-linguistic variation seems to be an important factor in the conflicting perspectives on the referentiality and modifiability of incorporated nouns in the literature and that the attested variation leads to the hypothesis that there are two independently occurring incorporation processes, i.e., the incorporation of modifiable nouns and the incorporation of non-modifiable nouns, it can be concluded that cross- and intra-linguistic variation are highly relevant for the understanding of the pragmatic referentiality and semantic modifiability of incorporated nouns.
5 Verb-based restrictions on noun incorporation across languages

5.1 Introduction

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

Verb-based restrictions on noun incorporation may, however, be highly relevant for theoretical approaches to noun incorporation. One of the main questions addressed in the literature on noun incorporation concerns the status of incorporation as a lexical or syntactic process (Mithun 2000: 923–925; Massam 2009: 1083–1086, 2017; Haugen 2015: 414–421). Is incorporation a lexically restricted type of word formation? Or is it rather a productive process that can be described by purely syntactic principles? Based on the characteristics and possibilities of incorporated nouns, arguments supporting each of these alternatives have been put forward. For instance, some studies state that incorporation, in contrast to most syntactic processes, is sensitive to the semantic roles of potentially incorporated nouns (Mithun 1984: 875; Anderson 2000: 16), while other works emphasize that only nouns in particular syntactic positions can be incorporated (Baker 1988: 81, 88, 90). In addition, some researchers have addressed the formal properties of incorporated nouns, demonstrating that in some languages not only noun stems but also inflected nouns and noun phrases can be incorporated (Barrie and Mathieu 2016; see also Chapter 3), which may be regarded as evidence for the syntactic status of noun incorporation in these languages. By contrast, others show that in some languages incorporation is limited to specific semantic types of nouns, such as body-part nouns (Aikhenvald 2007: 20; Caballero et al. 2008: 391).

Knowledge about possible verb-based restrictions on noun incorporation may also provide important insights concerning the question to what extent incorporation is a lexical or a syntactic process and whether and how languages may vary in this respect. Several studies have suggested that verb-based restrictions are relevant for

---

1 This chapter is a slightly adapted version of: Olthof, Marieke, Eva van Lier, Tjeu Claeissen, Swintha Danielsen, Katharina Haude, Nico Lehmann, Maarten Mous, Elisabeth Verbheven, Eline Visser, Marine Vuillermet & Arok Wolvengrey. forthc. Verb-based restrictions on noun incorporation across languages. Linguistic Typology.
noun incorporation. Caballero et al. (2008: 392), for instance, state that “it is not unexpected that some verbs might be able to select for a nominal object that will incorporate into them”, i.e. for any transitive verb it may be specified lexically whether or not it can incorporate its object noun. Moreover, for certain languages it has been noted that noun incorporation is only likely or possible with particular verbs. For instance, Mithun (2010: 52) notes that in Mohawk, noun incorporation is restricted in terms of both the nouns and verbs that can be involved in incorporation constructions, in that “some stems occur exclusively in such constructions, some often, some occasionally, some rarely, and some never”. For Ket, it is known that “[o]nly two transitive bases allow incorporation of their patient-role noun object with any productivity” (Vajda 2017: 911). In addition, verb-based restrictions have been observed for several voice- and valency-affecting alternations (Tsunoda 1985: 391–392; Kemmer 1993: 42–74; Näss 2007: 124–141; Polinsky 2013; Vigus 2018: 370–371; Say in prep.). For example, in some languages antipassives may only be formed on the basis of “a certain subset of transitive predicates” (Polinsky 2013). Such restrictions may also be relevant for noun incorporation, which in many languages has a valency-changing effect in that incorporated nouns may lose their morphosyntactic status of core argument (Mithun 1984: 856, 859; Rosen 1989: 310–311; Gerdts 1998: 88).

This chapter therefore aims to investigate to what extent languages restrict noun incorporation to particular verbs and what types of restrictions appear to be relevant cross-linguistically. Section 5.2 introduces the definition of noun incorporation used in the study, discusses earlier studies relevant for verb-based restrictions on noun incorporation and formulates the research questions. Section 5.3 presents the results of the first part of the study, which consists of an explorative typological survey based on descriptive sources of 50 incorporating languages (cf. Olthof and Van Lier 2018). Section 5.4 discusses the second part of the study, which investigates verb-based restrictions on noun incorporation more systematically in a sample of eight languages, guided by a questionnaire and based on data from spoken language corpora. Finally, in Section 5.5 we conclude that noun incorporation is indeed restricted in terms of which verbs allow this construction within and across languages. The likelihood that a verb can incorporate appears to be partly determined by its degree of morphosyntactic transitivity, but the attested variation across verbs and across languages shows that purely lexical restrictions play an important role as well.
5.2 Definition, theoretical background and research questions

5.2.1 Defining noun incorporation

While many different definitions of noun incorporation are used throughout the literature (see e.g. Massam 2009; Johns 2017), this study follows Caballero et al. (2008: 385) in defining noun incorporation as a construction in which a noun occurs “between parts of the inflected verbal complex”. The relevant “parts” are, in addition to the verbal stem, typically affixes, but they may also be clitics, particles or other separate words that appear strictly adjacent to verbs (Caballero et al. 2008: 385).

Thus, in example (1b) from Chukchi, the noun utt ‘stick’ is considered to be incorporated because it is preceded by the first part of the verbal person marking circumfix t-...-γʔek and followed by the verbal stem and the second part of the circumfix.

(1) Noun incorporation in Chukchi

a. γəm-nan t-ə-mle-γʔen-Ø ott-ə-lyən
     I-ERG 1SG.A-E-break-3SG.P-PST stick-E-ABS.SG
     ‘I broke the stick.’

b. γəm-Ø t-utt-ə-mle-γʔek-Ø
     I-ABS 1SG.S-stick-E-break-1SG.S-PST
     ‘I broke a stick.’

(Kurebito 2001: 79)

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

---

2 Note that in the questionnaire-based case studies of noun incorporation, discussed in Section 5.4, a few additional language-specific criteria are used (see Section 5.4.1.2 and http://dx.doi.org/10.21942/uva.12161751 for more details).


4 The vowel difference between ott in (1a) and utt in (1b) is due to a vowel harmony rule (Kurebito 2001: 66).
Incorporation: Constraints on variation

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

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

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

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

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

a. *De kapper begint te hier-knipp-en*
   The barber begins to hair-cut-INF
   ‘The barber begins to cut the hair.’
   (Dijk 1997: 44)

b. *Ik sil him hier-knipp-e*
   I will him hair-cut-INF
   ‘I will cut his hair.’
   (Dijk 1997: 41)

It is also important to mention that the definition we use does not delimit incorporation in terms of the semantic or syntactic role of the incorporated noun. Nor does it pose any restrictions on the formal characteristics of incorporated nouns and incorporating verbs. Although incorporated nouns in most languages are identical or at least very similar in form to corresponding independently occurring nouns, in a few languages such as Halkomelem (Gerds 2003: 345–346) some incorporated nouns have non-incorporated counterparts that are formally completely unrelated. Such nouns may, however, be seen as suppletive versions of non-incorporated nouns and the relevant constructions are not excluded from the study (cf. Caballero et al. 2008: 387–388). In addition, in some languages, including Movima (Haude 2006: 72–73) and Washo (Bochnak and Rhomieux 2013), some or all incorporated nouns are bound in the sense that they never occur without an additional morpheme or are even obligatorily incorporated into a verb. Cases with such incorporated nouns are included in the study as well. Similarly, we include constructions with obligatorily incorporating elements that are sometimes called affixes in languages like Eastern Ojibwa (see e.g. Mathieu 2013) and Kalaallisut (see e.g. Fortescue 1980), as long as these elements have action semantics and the resulting constructions conform to the definition of inclusion of the noun inside the inflected verbal complex. For instance, the Kalaallisut construction in example in (5) is considered to be a noun incorporation construction, even though the element *isur ‘fetch’* cannot occur independently without a noun (Fortescue 1980: 274, 1984: 322).

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

The inclusion of obligatorily incorporating or bound verbs is only relevant for the typological survey presented in Section 5.3; in the languages studied in the corpus-based case studies discussed in Section 5.4, such verbs only play a very marginal role. Overall, the meanings of the bound verbs included in the typological survey do not seem to be very different from the meanings of the other incorporating verbs. Therefore, we do not distinguish between bound verbs and other incorporating verbs in the discussion of the results of the typological survey in Section 5.3.
Noun incorporation in Kalaallisut

\textit{titurvi-isur-put}

cup-fetch-3PL.IND

‘They fetched (the) cups.’

(Fortescue 1984: 322)

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

5.2.2 Theoretical background

5.2.2.1 Noun incorporation and semantic transitivity

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

From a semantic argument-structure perspective, Mithun (1984: 875) argues that transitive verbs are more likely to incorporate their patient arguments than intransitive verbs. More specifically, she proposes that all incorporating languages at least allow the incorporation of patient arguments of transitive verbs, i.e. P-arguments, such that intransitive verbs can only incorporate their patient arguments, i.e. Sp-arguments, in languages that also show the incorporation of P-arguments (see also Haspelmath 2018: 318, fn. 9).

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

Secondly, Mithun (1984: 863) maintains that verbs that tend to take inanimate, non-agentive and non-individuated patient arguments are more suitable for incorporation than those with animate, agentive and individuated patient arguments.

\[\text{According to Mithun (1984: 875), languages that both allow the incorporation of P-arguments into transitive verbs and the incorporation of Sp-arguments into intransitive verbs may additionally show the incorporation of instruments and/or locations. Instrument incorporation and location incorporation thus appear to be more marginal types of noun incorporation.}\]
Importantly, non-individuated P-arguments are characteristic of verbs that are semantically low in transitivity (Hopper and Thompson 1980: 252–253; Tsunoda 1981: 393). Thus, the non-individuated status of the P-arguments of frequently incorporating verbs suggests that semantically low-transitive verbs are most likely to incorporate.

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

Note also that the noun incorporation process itself has often been regarded as a way to mark the referent of a noun as low in individuation. For instance, Hopper and Thompson (1980: 257) consider P-argument incorporation to correlate with low individuation of P-arguments. Similarly, for the seven P-incorporating languages included in her study, Vigus (2018: 360) finds that the function of incorporation is to indicate the lower individuation of P-arguments. Interestingly, if the function of noun incorporation is to mark a P-argument as low in individuation, we may expect that highly transitive verbs, which tend to take highly individuated P-arguments, are most likely to show incorporation. These verbs would then be used in an incorporation construction when their P-argument is, unexpectedly, less individuated.

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

---

8 According to Baker (1988: 86–87), adjuncts can never be incorporated. Note also that there is some similarity between Baker’s claim that only internal arguments can be incorporated and Mithun’s (1984: 875) proposal that all incorporating languages show the incorporation of patient arguments. However, whereas Mithun predicts that transitive verbs are more likely to incorporate nouns than intransitive ones, Baker does not make a distinction between transitive and unaccusative verbs but argues that incorporation
Secondly, based on studies using data from the Valency Patterns Leipzig (ValPaL) Project (Hartmann et al. 2013), a possible relation between the incorporation of P-arguments and degree of morphosyntactic transitivity can be identified. The ValPaL project investigates the argument-coding properties of the translational equivalents of 80 verb meanings in 36 languages. As part of this project, Haspelmath (2015: 143) assigns a so-called “transitivity prominence” score to 70 of the 80 verb meanings. This means that for these verb meanings he calculates the percentage of transitively encoded verbs among all translational equivalents across the sample languages. Transitive encoding is defined as the coding used for the A- and P-arguments of the verb meaning ‘break’ in a particular language. A score of 1 for a particular verb meaning indicates that its translational equivalents use basic transitive coding in all 36 languages, whereas a score of 0 means that the verb meaning does not have a translational equivalent with basic transitive coding in any of the 36 languages.

In the context of the same project, Malchukov (2015) and Wichmann (2015) study the ability of the 80 verb meanings to participate in object-demoting and object-deleting alternations, including P-argument incorporation, across the sample languages. Their findings lead to a hierarchy for object-demoting and object-deleting alternations in which verb meanings at the top are cross-linguistically most likely to undergo object-demoting and object-deleting alternations and verb meanings at the bottom are cross-linguistically least likely to undergo such alternations. This hierarchy, henceforth called the object dem/del hierarchy, is represented in (6) below. Note that in (6), only the 70 verb meanings also studied by Haspelmath (2015) are included and the transitivity prominence scores of these verbs are given between parentheses. into unergative verbs is impossible. In addition, Mithun (1984: 875) does not exclude the incorporation of nouns that are not arguments. The number of incorporation constructions included in their work is in fact fairly limited (see Wichmann 2015: 178). However, P-argument incorporation constructions share functional and sometimes also formal characteristics with other object-demoting constructions, such as antipassives (Heaton 2017: 17; Vigus 2018), such that the hierarchy may nevertheless be quite relevant for the study of noun incorporation. Note that this hierarchy also includes intransitive verb meanings. Although these verb meanings all appear in the lower part of the hierarchy, only one verb meaning, FEEL PAIN, appears at the lowest end of the hierarchy. This means that, surprisingly, translational equivalents of most of these intransitive verb meanings are able to undergo object-demoting and/or object-deleting alternations in at least some languages. On closer inspection, however, it appears that translational equivalents of many of the verb meanings in the lower part of the hierarchy can only be used in object-demoting and/or object-deleting alternations in very few of the languages studied and/or if they are first transitivized. For instance, translational equivalents of the verb meaning DIE can only undergo alternations that Wichmann (2015) classifies as object-demoting or object-deleting in Russian and Sliammon. In addition, in Bezhta the translational equivalents of BE DRY and BOIL can undergo an antipassive alternation, but only if they are combined with the causative suffix -l. In (6), as in the remainder of this chapter, we follow the ValPaL practice to write comparative verb meanings in small caps. Yet, for the sake of terminological simplicity, we will use the term “verb” as a shortcut for “(comparative) verb meaning”, unless we think it is important to explicitly differentiate between the comparative verb meaning and its translational equivalents, i.e. the actual language-specific lexical items, for which we also use the word “verb”. The lexical items will be written in italics and their translations given between single quotation marks, in accordance with general typological practice.
According to the methods developed by Wichmann (2015, 2016; Aldai and Wichmann 2018), the object dem/del hierarchy can be interpreted as statistically implicational. Thus, verbs lower on this hierarchy can usually only be involved in object-demoting and object-deleting alternations in a particular language, if verbs higher on the hierarchy allow these alternations as well. Given that P-incorporation is also an object-demoting construction, we expect to find overlap between high-ranking verbs on this hierarchy and verbs that incorporate frequently across languages in our study.

(6) Object dem/del hierarchy (adapted from Malchukov 2015: 105–106; Wichmann 2015: 166–167). For each verb, the number between parentheses indicates its transitivity prominence score as measured by Haspelmath (2015).

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

Crucially, it appears that the verbs higher on the hierarchy generally show higher transitivity prominence scores than the verbs lower on the hierarchy, i.e. they show transitive coding in more languages. The ranking of the verbs based on their
transitivity prominence scores and their position on the object dem/del hierarchy show a strong and statistically significant correlation, as demonstrated by their Spearman’s rank order correlation coefficient ($\rho = 0.78$, $p < 1e^{-14}$). We gather from this that morphosyntactic transitivity as measured by Haspelmath’s (2015) transitivity prominence scores is at least an important factor underlying the object dem/del hierarchy, which suggests that it is also relevant for P-argument incorporation, in that verbs with higher transitivity prominence scores may be expected to be more likely to incorporate their P-arguments.

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

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

---

12 Note that the object dem/del hierarchy results from a procedure based on Guttmann scaling (see Wichmann 2015, 2016; Aldai and Wichmann 2018), while the ranks in Haspelmath’s scale are based on simple counting. Yet, as shown in Aldai and Wichmann (2018: 270), although based on a smaller data set, the results are quite similar.
Verb-based restrictions on noun incorporation across languages

(7) Research questions

a. Which verbs are most likely to incorporate nouns across languages?

b. To what extent are verb-based restrictions on noun incorporation determined by morphosyntactic transitivity?

c. What other factors affect the likelihood that a verb is able to incorporate nouns?

d. To what extent do languages differ in terms of how many and which verbs allow noun incorporation and how frequently these verbs show noun incorporation?

Note that the research question in (7b) specifically focuses on morphosyntactic transitivity as discussed in Section 5.2.2.2 rather than on the semantic characteristics that can be related to transitivity presented in Section 5.2.2.1. Whereas the effect of and relation between these semantic characteristics remains somewhat unclear, morphosyntactic transitivity can be measured systematically on the basis of Haspelmath’s (2015) notion of transitivity prominence and, as shown in Section 5.2.2.2, is also involved in the object dem/del hierarchy proposed by Malchukov (2015) and Wichmann (2015). In addition, morphosyntactic transitivity presumably reflects some aspects of semantic transitivity.

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

5.3 Typological survey

5.3.1 Method and data

The typological survey of incorporating verbs makes use of a sample of 50 languages drawn from a list of 259 languages that are described as incorporating languages in the literature on incorporation. This list, which can be found in Appendix 1, includes languages from 82 language families, and the 50 languages in our sample are all from different families.13 We selected those languages for which most data could be obtained. The data are extracted from reference grammars and from articles on noun incorporation in the relevant languages. The sample languages and the data sources are included in Appendix 5.

13 There is one exception: both Panamint and Ute-Southern Paiute belong to the Uto-Aztecan family. Note also that, since noun-incorporating language families are not evenly distributed around the world (Velupillai 2012b), geographical distribution was not taken into account in the sampling procedure.
For each language we listed the meanings of all verbs for which it is mentioned or shown in the sources that they can be used in noun incorporation. Note that, in accordance with the definition of noun incorporation introduced in Section 5.2.1, we did not restrict our search to P- or Sp-incorporation constructions, i.e. verbs allowing the incorporation of locations and instruments were included as well. However, as the most frequently incorporating verbs in the languages of our sample, to be presented in the next subsection, generally show P- or Sp-incorporation in our data, we focus on P- and Sp-incorporation in the remainder of this section and only mention other types of noun incorporation where they are particularly relevant.

Considering the exact glosses used for the incorporating verbs in the sources, the data collection resulted in a list of 808 different verb meanings. However, this list included many near-synonyms. For instance, a verb glossed as ‘roast’ was found in one language and a verb glossed as ‘toast’ in some others. Further examples include ‘look after’ and ‘take care of’ or ‘happen’ and ‘occur’. For this reason, the next step of the data collection involved combining the near-synonyms into single entries in our list. To further reduce the number of meanings, we also merged certain stative verbs that are expressed in English by means of the verb ‘to be’ combined with an adjective, grouping them in accordance with the property concept classes distinguished by Van Lier (2017). For example, the meaning BE + PHYSICAL PROPERTY represents glosses such as ‘be dry’ and ‘be dirty’, while the meaning BE + EXPERIENTIAL STATE includes e.g. ‘be tired’ and ‘be hungry’. All in all, this merging procedure reduced the original list of 808 verb meanings to 526.

The method of data collection for the typological survey has some obvious limitations. Firstly, as can already be inferred from the list of sources included in Appendix 5, there are large differences in the amount of available data for the different languages. This means that the data gathered for some languages are much more likely to represent a substantial proportion of all verbs that can be used in noun incorporation than the data for other languages. Especially in the case of languages in which noun incorporation is a very productive process, our data necessarily cover only a small subset of the possibilities. Secondly, because the data were collected primarily on the basis of reference grammars, the data include information about verbs that can incorporate nouns, but very little information about which verbs cannot be used in noun incorporation. The latter form of evidence is found only in an indirect way in those few cases where incorporation is restricted to a very limited set of verbs (cf. the example of Ket in Section 5.1). Thirdly, the method does not take into account the frequency with which the relevant verbs are used in noun incorporation. The data thus only state that a particular verb can incorporate nouns and do not include any

---

14 The classes used in Van Lier (2017) are similar to the ones proposed by Dixon (2004 and earlier work), but experiential states are called “corporeal properties” by Dixon and treated as a subclass of “physical properties”. See Van Lier (2017) for a more detailed description and motivation of the various classes.
information about how often the verb occurs in noun incorporation, compared to other constructions.

5.3.2 Results and discussion

5.3.2.1 Frequently incorporating verbs

Table 1 shows the meanings of the verbs that are found as incorporating verbs most frequently across the sample. Each of these verbs is found in at least ten different languages. The number of languages in which a verb is found to incorporate is given in the second column. For those verbs which are also studied by Haspelmath (2015), we give the transitivity prominence scores between parentheses. Note that most of the verb meanings in Table 1 in fact represent mergings of multiple glosses, as explained and illustrated in the previous section. The exact set of glosses included under each verb meaning can be found on http://dx.doi.org/10.21942/uva.12161724.

Table 1. Verbs found as incorporating verbs in at least ten out of 50 languages in the survey. The numbers between parentheses indicate their transitivity prominence scores as measured by Haspelmath (2015).

<table>
<thead>
<tr>
<th>Verb meaning</th>
<th>Number of languages (of total 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C\textsc{ut} (1.00)</td>
<td>22</td>
</tr>
<tr>
<td>MAKE/DO</td>
<td>21</td>
</tr>
<tr>
<td>E\textsc{at} (0.93)</td>
<td>19</td>
</tr>
<tr>
<td>SEARCH \textsc{for} (0.88)</td>
<td>17</td>
</tr>
<tr>
<td>GIVE (0.98)</td>
<td>16</td>
</tr>
<tr>
<td>WASH (0.94)</td>
<td>16</td>
</tr>
<tr>
<td>PUT (0.98)</td>
<td>14</td>
</tr>
<tr>
<td>BREAK (\textsc{tr.}) (1.00)</td>
<td>13</td>
</tr>
<tr>
<td>K\textsc{ill} (1.00)</td>
<td>13</td>
</tr>
<tr>
<td>BE + PHYSICAL PROPERTY (0.00)\textsuperscript{15}</td>
<td>12</td>
</tr>
<tr>
<td>\textsc{bu}y</td>
<td>12</td>
</tr>
<tr>
<td>FEEL PAIN (0.12)</td>
<td>12</td>
</tr>
<tr>
<td>GO (.05)</td>
<td>12</td>
</tr>
<tr>
<td>HAVE</td>
<td>12</td>
</tr>
<tr>
<td>TAKE (1.00)</td>
<td>12</td>
</tr>
<tr>
<td>FALL</td>
<td>11</td>
</tr>
<tr>
<td>HIT (1.00)</td>
<td>11</td>
</tr>
<tr>
<td>HUNT</td>
<td>11</td>
</tr>
<tr>
<td>PUT \textsc{down}</td>
<td>11</td>
</tr>
<tr>
<td>REMOVE</td>
<td>11</td>
</tr>
<tr>
<td>SEE (0.93)</td>
<td>11</td>
</tr>
<tr>
<td>BE + EXPERIENTIAL STATE (0.00)\textsuperscript{16}</td>
<td>10</td>
</tr>
<tr>
<td>CATCH</td>
<td>10</td>
</tr>
<tr>
<td>DIE (0.00)</td>
<td>10</td>
</tr>
</tbody>
</table>

\textsuperscript{15} This is the transitivity prominence score for the meaning \textsc{be} dry in Haspelmath’s (2015) study.

\textsuperscript{16} This is the transitivity prominence score for the meanings \textsc{be} sad and \textsc{be} hungry in Haspelmath’s (2015) study.
5.3.2 Morphosyntactic transitivity

Table 1 shows that almost all of the most frequently incorporating verbs either have a (di)transitive meaning or are patientive intransitive verbs. This is in accordance with Baker’s (1988) claim that only transitive verbs and unaccusative verbs can incorporate nouns. Nevertheless, Table 1 also includes an exception to this pattern: the verb GO is generally an agentive intransitive. In addition, among the other, less frequently incorporating verbs (not included in Table 1) we also find, for instance, COME (in four languages), JUMP (in four languages) and RUN (in three languages). However, with these agentive intransitive verbs it is typically a goal or location that is incorporated rather than the agent (Sa) argument, as for example in (8) from Southern Tiwa.\(^{17}\)

(8) Noun incorporation in Southern Tiwa

\begin{verbatim}
Te-fiesta-mi-ban
1SG.S-party-go-PST
‘I went to the party.’
\end{verbatim}

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

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

The hierarchy in (9) shows the object dem/del hierarchy introduced in Section 5.2.2.2 and indicates the number of languages in which each of the verbs included in this hierarchy is found as an incorporating verb in our typological survey. It can be observed in (9) that some of the verbs that were expected to be prone to noun incorporation because they are high on the object dem/del hierarchy are indeed found as incorporating verbs in many of the sample languages. These include, for instance, EAT, WASH, GIVE, CUT, SEARCH FOR, KILL and BREAK (TR.). Correspondingly, some of the verbs that are low on this hierarchy, such as LAUGH, SIT, SIT DOWN, BE HUNGRY and FEEL COLD, are found as noun-incorporating verbs in few or none of the sample

\(^{17}\) In a few cases, the semantic role of the incorporated noun was hard to determine. Note, though, that both the incorporation of adjuncts and the incorporation of Sa-arguments are not predicted by Baker’s (1988) theory, whereas Mithun (1984) does recognize the incorporation of instruments and locations.
languages. Thus, there seems to be some overlap between verbs that are unlikely to undergo different types of object-demoting and object-deleting alternations and verbs that are unlikely to undergo noun incorporation. The data of the present study are, however, not completely in line with the expectations based on the object dem/del hierarchy. This will be discussed further in the next subsection.

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

\[
\begin{align*}
\text{EAT} & (19), \text{WASH} (16), \text{GIVE} (16) \supset \\
\text{STEAL} & (4), \text{TEACH} (1), \text{SHAVE} (2), \text{COOK} (6) \supset \\
\text{CUT} & (21), \text{WIPE} (5), \text{SEARCH FOR} (17), \text{HIT} (11) \supset \\
\text{KILL} & (13), \text{ASK FOR} (2), \text{TAKING} (12), \text{BEAT} (4) \supset \\
\text{SEE} & (11), \text{THROW} (8), \text{HEAR} (3), \text{TOUCH} (6), \text{LOOK AT} (0) \supset \\
\text{GRIND} & (2), \text{BREAK (TR.)} (13), \text{FILL} (2), \text{HUG} (2), \text{COVER} (6), \text{POUR} (4), \text{THINK} (1), \text{LOAD} (4) \supset \\
\text{TELL} & (4), \text{KNOW} (4), \text{TEAR} (3), \text{HELP} (1), \text{TIE} (7), \text{SHOW} (2), \text{CARRY} (7) \supset \\
\text{SING} & (2), \text{DIG} (6), \text{DRESS} (0) \supset \\
\text{CLIMB} & (2), \text{BUILD} (6), \text{FEAR} (1) \supset \\
\text{SMELL (TR.)} & (0), \text{PUSH} (3), \text{PUT} (14), \text{SEND} (3), \text{LEAVE (TR.)} (6) \supset \\
\text{FEEL} & (2), \text{BLINK} (0), \text{SAY} (2), \text{TALK} (2), \text{SHOUT AT} (0), \text{NAME} (1), \text{RUN} (3) \supset \\
\text{JUMP} & (4), \text{HIDE} (2), \text{FRIGHTEN} (1), \text{LIKE} (7), \text{PLAY (TR.)} (3), \text{FOLLOW} (7), \text{LIVE} (0), \text{BE DRY} (1) \supset \\
\text{BRING} & (6), \text{ROLL} (0), \text{LAUGH} (0), \text{BURN (INTR.)} (1), \text{SCREAM} (0), \text{GO} (12), \text{SINK (INTR.)} (0) \supset \\
\text{MEET} & (0), \text{DIE} (10), \text{COUGH} (0), \text{BOIL (INTR.)} (0), \text{BE A HUNTER} (0) \supset \\
\text{FEEL PAIN} & (12), \text{SIT} (5) \supset \\
\text{BE SAD} & (0) \supset \\
\text{SIT DOWN} & (1), \text{BE HUNGRY} (1) \supset \\
\text{RAIN} & (1) \supset \\
\text{FEEL COLD} & (2)
\end{align*}
\]

5.3.2.3 Other factors

Some verbs that appear high on the object dem/del hierarchy are not found as noun-incorporating verbs in many of our sample languages. For a few of these cases we can offer a tentative explanation. Firstly, verbs such as THINK, TELL and KNOW may be

\[\text{In our data, we could not distinguish between the meanings } \text{SEE versus LOOK AT and we included all examples under SEE. Hence, the “0” reported here is due to this choice, rather than to the fact that LOOK AT does not incorporate in any of the sample languages.}\]
unlikely candidates for incorporation because they potentially or even typically take clausal complements rather than nominal objects. It may be the case that these verb meanings more easily allow object deleting alternations, which may explain their different behavior in our data compared to the object dem/del hierarchy. Secondly, a few semantically quite specific verbs such as WIPE and GRIND we may have found in relatively few languages as noun-incorporating verbs simply because they are not used in many data sources. More generally, of course, the fact that a particular verb does not appear often in our data could well be a side effect of the method of data collection.

On the other hand, there are also some verbs that are low on the object dem/del hierarchy, yet are found as incorporating verbs in relatively large numbers of languages. Three of these, GO, DIE and FEEL PAIN, correspond to intransitive verbs, which explains why they are unlikely candidates for object-demoting and object-deleting alternations but do occur as incorporating verbs. Another verb that is found more frequently as an incorporating verb than expected based on this hierarchy is PUT. The relatively high number of languages that show noun incorporation with this verb may be related to the observation made by Mithun (1984: 863) that verbs with very general semantics are likely to incorporate nouns. Note here that we also find MAKE/DO, HAVE and TAKE, all of which may be considered to be general in semantic scope, among the most frequently incorporating verbs, as can be seen in Table 1.

5.3.2.4 Variation across languages

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

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

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

5.4.1 Method and data
5.4.1.1 Questionnaire design

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

The inventory of verb meanings in the questionnaire takes into account previous research on the role of morphosyntactic transitivity in incorporating verbs (see the discussion in Section 5.2.2.2) as well as the results of the typological survey discussed in Section 5.3. First, the questionnaire verbs cover the full range of transitivity prominence scores calculated by Haspelmath (2015), in order to verify the finding from the typological survey that verbs with a relatively high transitivity score and verbs with a very low transitivity score, i.e. intransitive verbs, are most likely to show noun incorporation. Second, verb meanings representing each of the levels of the object dem/del hierarchy (Malchukov 2015; Wichmann 2015) are selected because P-incorporation is expected to pattern with other object-demoting and object-deleting alternations, such that verbs ranking high on this hierarchy would also be frequently used in P-incorporation. Thirdly, the questionnaire contains both typically patientive and agentive intransitive verbs, in order to test the idea from earlier literature (and to a certain extent supported by the typological survey) that many patientive intransitive verbs show noun incorporation, while agentive intransitive verbs do not or only rarely allow it. Fourth, most verbs found most frequently as incorporating verbs in the typological survey (listed in Table 1) are also part of the questionnaire. Finally, we take into account expectations based on other factors, namely that verbs with a very general meaning are often used in incorporation, as well as verbs expressing habitual activities in combination with certain incorporated nouns. In contrast, verbs that can take complement clause objects are not expected to be prone to incorporation. Verb meanings representing each of these verb types are also part of the questionnaire.

---

21 We tried to balance the representativeness of the range of verb meanings in the questionnaire with general concerns of feasibility of the data collection. Therefore, we did not, for instance, include all the verb meanings from the ValPaL project.
22 We excluded REMOVE and PUT DOWN, since these are semantically similar to TAKE and PUT, which we did include.
Incorporation: Constraints on variation

In addition to the list of verb meanings, the questionnaire includes some meta-questions on the corpus on which its answers are based as well as some general questions about restrictions on incorporation in the relevant language and about other verbs than the selected 47 that allow incorporation in that language. In addition, the questionnaire asks for at least one example of each incorporating verb used in a noun incorporation construction. The complete questionnaire is included on http://dx.doi.org/10.21942/uva.12161748.

5.4.1.2 Language sample and data
The questionnaire was filled out by the expert authors, for one (or two, in Danielsen’s case) of the eight noun-incorporating languages, as represented in Table 2.

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

<table>
<thead>
<tr>
<th>Language</th>
<th>Glottocode</th>
<th>Language family</th>
<th>Country</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baure</td>
<td>baur1253</td>
<td>Arawakan</td>
<td>Bolivia, Plurinational State of</td>
<td>Swintha Danielsen</td>
</tr>
<tr>
<td>Ese Ejja</td>
<td>ese1248</td>
<td>Pano-Taçanan</td>
<td>Bolivia, Plurinational State of</td>
<td>Marine Vuillermet</td>
</tr>
<tr>
<td>Guarayo</td>
<td>guar1292</td>
<td>Tupian</td>
<td>Bolivia, Plurinational State of</td>
<td>Swintha Danielsen</td>
</tr>
<tr>
<td>Iraqw</td>
<td>iraq1241</td>
<td>Afro-Asiatic</td>
<td>Tanzania, United Republic of</td>
<td>Tjeu Claessen and Maarten Mous</td>
</tr>
<tr>
<td>Kalamang</td>
<td>kara1499</td>
<td>West Bomberai</td>
<td>Indonesia</td>
<td>Eline Visser</td>
</tr>
<tr>
<td>Movima</td>
<td>movi1243</td>
<td>Isolate</td>
<td>Bolivia, Plurinational State of</td>
<td>Katharina Haude</td>
</tr>
<tr>
<td>Plains Cree</td>
<td>plai1258</td>
<td>Algic</td>
<td>Canada, United States</td>
<td>Arok Wolfengrey</td>
</tr>
<tr>
<td>Yucatec Maya</td>
<td>yuca1254</td>
<td>Mayan</td>
<td>Belize, Guatemala, Mexico</td>
<td>Nico Lehmann and Elisabeth Verhoeven</td>
</tr>
</tbody>
</table>

The data gathered in the questionnaire-based case studies are mostly from electronic corpora of spoken language data assembled during fieldwork, typically as part of documentation projects and often in the context of language endangerment. Details about the respective language corpora can be found on http://dx.doi.org/10.21942/uva.12161751. In some cases, the corpus data were supplemented by information from dictionaries, reference grammars and other published sources as well as by elicitation or volunteering by native speakers and/or by personal language knowledge of the respective expert authors. Roughly, the corpora range in size between ca. 29,000 and 160,000 words and consist mostly of (spoken) narrative and

---

23 Note that the size and composition of the sample is merely a matter of convenience. The fact that four languages are from Bolivia is purely coincidental. We are not aware of any direct contact between (some of) these languages, but we cannot exclude the possibility that there are similarities between them due to areal effects. Of the languages in the questionnaire-based study, the following also figure in the 50-language sample used for the typological survey: Ese Ejja, Iraqw, Movima and Yucatec Maya.
It goes without saying that these corpora are relatively small compared to corpora of many Indo-European and other well-studied languages often used in corpus linguistics, and this may impact the reliability of the frequency data extracted from them. Despite this limitation, we consider the corpus-based methodology advantageous, especially because it allows for a much more systematic search for particular verbs compared to the typological survey.

The document on http://dx.doi.org/10.21942/uva.12161751 explains how noun incorporation constructions are identified in the sample languages. In all but one case, namely Kalamang, these identification criteria match the general definition employed in the typological survey (see Section 5.2.1), in the sense that there are at least some conditions under which incorporation involves the inclusion of the noun inside the inflected verbal complex. In Kalamang, noun incorporation is defined by the absence of an object-marker on the incorporated noun in combination with a phonological criterion: the noun and verb have a single prosodic contour and thus form a single phonological word. Also in other sample languages the main definition is supplemented by various additional diagnostics, which may be morphosyntactic and/or phonological in nature. In Ese Ejja, for instance, they include the lack of the e-marker on incorporated nouns from the e-class, which always take this marker when they occur independently, as well as the phonological word-status of the incorporation construction in terms of stress assignment (Vuillermet 2012: 514, 515). For concrete examples of noun incorporation constructions in all sample languages we refer to http://dx.doi.org/10.21942/uva.12161751.

In some languages, noun incorporation constructions may be nominalized. Nominalized incorporation constructions are included in the study, except when the relevant nominalization strategy makes it impossible to verify if the relevant constructions really involve noun incorporation or not, as is the case in Iraqw. In this language patients of nominalized verbs immediately precede the nominalized verb with no additional marking just as an incorporated noun would precede an inflected verb. However, this patient requires an object pronoun in the verbal complex of the inflected main verb, suggesting that the patient is a separate constituent and not an incorporated noun (Mous and Qorro 2010: 73–75). Such a criterion is absent in nominal clauses containing nominalized verbs with a patient noun. For this reason, such constructions were not counted as noun incorporation here.

On a final note, it should be stressed that not only the form, but also the function of noun incorporation may differ between the sample languages. For

---

24 For Baure, Ese Ejja and Guarayu the number of words in the corpus is unknown. Information about the number of hours of recorded speech for these languages is included on http://dx.doi.org/10.21942/uva.12161751. For the other languages included in the questionnaire-based study, the number of words in the corpus can be found there.

25 In addition, we believe it is important to capitalize on the available resources of lesser studied languages and to stimulate collaboration between native speakers, fieldworkers and typologists.
example, in Baure, so-called “Ground incorporation” and “classifying incorporation” portray the referent of the incorporated noun as generic, non-individuated and backgrounded (Danielsen 2007: 99). In Ese Eja, by contrast, incorporation does not have this function, which is rather served by the antipassive construction. Incorporation in this language always involves possessed nouns and serves to promote the possessor to argument status (Vuillermet 2012: 514, 518–519). As some languages show more than one type of noun incorporation, there are also differences within languages. In fact, the different functions of noun incorporation are expected to play an important role in determining the verbs’ incorporating potential. However, differentiating between these different functional types of incorporation must be left for a future study.

5.4.1.3 Method
For each of the 47 verb meanings in the questionnaire we checked whether or not the translational equivalent in each sample language appears with noun incorporation in the relevant corpus. Just as in the typological survey, not only cases of P- and Sp-incorporation but also cases of incorporation of nouns with other semantic roles are included. The document on http://dx.doi.org/10.21942/uva.12161751 contains the results of this query. For verb meanings with more than one translational equivalent, the translational equivalents are numbered as (i), (ii), et cetera, and data are included for each of them (see further below on the selection of translational equivalents). Those translational equivalents that are found in the relevant corpus at least once with an incorporated noun are counted as incorporating verbs in the corresponding language. The frequency of occurrence in the corpus of these incorporating verbs, both with and without noun incorporation, is also given in the document on http://dx.doi.org/10.21942/uva.12161751. Translational equivalents that are not attested with incorporation in the corpus of a language are interpreted as non-incorporating verbs in the relevant language. Of course, however, the absence of noun incorporation with a particular verb in a (relatively small) corpus does not prove that noun incorporation is impossible. Therefore, we also included, as much as possible, information from published sources and native speakers (see Section 5.4.1.2), in order to verify whether or not noun incorporation is (im)possible for a verb. For verbs that are not found in the corpus at all (neither with nor without an incorporated noun) and for which additional sources are not conclusive either, the question whether or not they allow noun incorporation cannot be answered and these cases are treated as missing data points. Finally, for some verb meanings no verbal translational equivalent was found in one or more of the sample languages. In these cases the question whether the verb

---

*An anonymous reviewer points out that different functions of incorporation, especially syntactic versus semantic ones, also influence the type of incorporated nouns. For instance, in Ese Eja, where incorporation serves the syntactic function of possessor raising, body-part nouns are preferentially incorporated.*
Verb-based restrictions on noun incorporation across languages

allows noun incorporation is irrelevant, and these cases are also considered missing data points.

One further methodological issue is important to interpret the results of the case studies and their comparison. As pointed out by Haspelmath and Hartmann (2015: 51–53), finding translational equivalents of particular verb meanings is not always a trivial matter. Mosel (in prep.) illustrates this problem in the context of a corpus investigation similar to ours, focusing on the Austronesian language Teop and using the verb CUT as an example. Looking for translational equivalents of CUT in this language, she finds the following: (i) Teop words translated by ‘cut’; (ii) Teop words that in addition to ‘cut’ are translated by other English words (e.g. ‘carve, cut, shave’); (iii) Teop words that are not translated by ‘cut’, but by a word that shares an English translation with another Teop word that is also translated by ‘cut’ (e.g. ‘shave’). Our study is mostly limited to cases like (i). In addition, when there is a choice within such cases, we choose the verb with the semantically least specific translation. Thus, if we find, for instance, two verbs glossed as ‘cut’ and ‘cut with knife’, respectively, we choose the former. Only if we cannot make a motivated choice between two alternative verbs, for instance between ‘cut with knife’ and ‘cut with machete’, which are equally semantically specific, we take both (or in rare cases all three or four) verbs into account. Cases with multiple translations, as in Mosel’s scenario (ii), are considered only when there is no candidate of type (i). Cases of type (iii) are not considered at all: a verb glossed as ‘shave’ would not be counted under ‘cut’, but rather under ‘shave’. Despite this procedure, however, decisions were not always straightforward to make. In the document on http://dx.doi.org/10.21942/uva.12161751 we therefore provide the verbs we chose as translational equivalents for the verb meanings in the questionnaire in each of the sample languages.

5.4.2 Results and discussion

5.4.2.1 Frequently incorporating verbs

Table 3 shows the 47 verb meanings included in the questionnaire, ordered according to their incorporation scores across the eight sample languages: the higher the incorporation score, the more frequently the verb meaning’s translational equivalents are used in noun incorporation in the data from the eight sample languages. The data on which Table 3 is based are included in the document on http://dx.doi.org/10.21942/uva.12161769. Note that those verb meanings that were found to be noun-incorporating in ten or more languages in the typological survey are presented in bold (cf. Table 1).
Table 3. Verbs included in the questionnaire ordered based on their cross-linguistic incorporation scores calculated on the basis of the data from the eight languages. The verbs that were found to be noun-incorporating in ten or more languages in the typological survey are presented in bold.

<table>
<thead>
<tr>
<th>Verb meaning</th>
<th>Cross-linguistic incorporation score</th>
</tr>
</thead>
<tbody>
<tr>
<td>WASH</td>
<td>1.00</td>
</tr>
<tr>
<td>CUT</td>
<td>0.94</td>
</tr>
<tr>
<td>EAT</td>
<td>0.88</td>
</tr>
<tr>
<td>CATCH</td>
<td>0.79</td>
</tr>
<tr>
<td>KILL</td>
<td>0.79</td>
</tr>
<tr>
<td>FEEL PAIN</td>
<td>0.75</td>
</tr>
<tr>
<td>HAVE</td>
<td>0.75</td>
</tr>
<tr>
<td>THROW</td>
<td>0.75</td>
</tr>
<tr>
<td>BE DRY</td>
<td>0.71</td>
</tr>
<tr>
<td>PUT</td>
<td>0.71</td>
</tr>
<tr>
<td>BREAK (TR.)</td>
<td>0.69</td>
</tr>
<tr>
<td>HIT</td>
<td>0.69</td>
</tr>
<tr>
<td>SHAVE</td>
<td>0.67</td>
</tr>
<tr>
<td>GIVE</td>
<td>0.63</td>
</tr>
<tr>
<td>SEARCH FOR</td>
<td>0.63</td>
</tr>
<tr>
<td>COVER</td>
<td>0.56</td>
</tr>
<tr>
<td>ASK FOR</td>
<td>0.50</td>
</tr>
<tr>
<td>COOK</td>
<td>0.50</td>
</tr>
<tr>
<td>FALL</td>
<td>0.50</td>
</tr>
<tr>
<td>HUNT</td>
<td>0.50</td>
</tr>
<tr>
<td>TAKE</td>
<td>0.50</td>
</tr>
<tr>
<td>BUY</td>
<td>0.43</td>
</tr>
<tr>
<td>TELL</td>
<td>0.43</td>
</tr>
<tr>
<td>FEEL COLD</td>
<td>0.40</td>
</tr>
<tr>
<td>MAKE/DO</td>
<td>0.38</td>
</tr>
<tr>
<td>SEE</td>
<td>0.38</td>
</tr>
<tr>
<td>FRIGHTEN</td>
<td>0.33</td>
</tr>
<tr>
<td>SIT DOWN</td>
<td>0.33</td>
</tr>
<tr>
<td>SIT</td>
<td>0.31</td>
</tr>
<tr>
<td>DIE</td>
<td>0.25</td>
</tr>
<tr>
<td>JUMP</td>
<td>0.25</td>
</tr>
<tr>
<td>NAME</td>
<td>0.25</td>
</tr>
<tr>
<td>SING</td>
<td>0.25</td>
</tr>
<tr>
<td>RUN</td>
<td>0.21</td>
</tr>
<tr>
<td>SAY</td>
<td>0.21</td>
</tr>
<tr>
<td>BE SAD</td>
<td>0.20</td>
</tr>
<tr>
<td>SINK (INTR.)</td>
<td>0.20</td>
</tr>
<tr>
<td>BURN (INTR.)</td>
<td>0.13</td>
</tr>
<tr>
<td>KNOW</td>
<td>0.13</td>
</tr>
<tr>
<td>LEAVE</td>
<td>0.13</td>
</tr>
<tr>
<td>PLAY</td>
<td>0.13</td>
</tr>
<tr>
<td>GO</td>
<td>0.06</td>
</tr>
<tr>
<td>BE HUNGRY</td>
<td>0</td>
</tr>
<tr>
<td>FEAR</td>
<td>0</td>
</tr>
<tr>
<td>HELP</td>
<td>0</td>
</tr>
<tr>
<td>MEET</td>
<td>0</td>
</tr>
<tr>
<td>THINK</td>
<td>0</td>
</tr>
</tbody>
</table>
For each verb meaning, the cross-linguistic incorporation score is the average of the language-specific incorporation scores for this verb meaning. These language-specific incorporation scores can have one of three values: 1 if the only or all translational equivalents of the relevant verb meaning allow noun incorporation; 0 if the only or all translational equivalents of the verb meaning do not allow noun incorporation; and 0.5 if there is at least one translational equivalent that allows noun incorporation and at least one that does not allow noun incorporation. No language-specific score was assigned to a verb meaning for which the language does not have any translational equivalents or for which it is unclear if the translational equivalent(s) allow noun incorporation. The cross-linguistic incorporation scores were calculated by dividing the sum of the scores for a particular verb across the eight languages by the number of languages for which a score could be calculated, resulting in a cross-linguistic incorporation score between 0 and 1.

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

In order to evaluate to what extent the ranking of verbs in terms of their cross-linguistic ability to incorporate can be considered a statistically implicational hierarchy, just as the verbs in the object dem/del hierarchy, we applied the method developed by Wichmann (2015, 2016; Aldai and Wichmann 2018) to determine its Guttmann’s coefficient. The coefficient is 81.98, which is below the conventional 85%. Also the p-value is not statistically significant: $p = 0.15$. This means that the

---

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

eight languages in our sample are not sufficient to determine a reliable ordering of the verbs with some form of implicational power. This result can be visualized with NeighborNet (Huson and Bryant 2006), as in Figure 1: while, as expected, the verbs that have comparable cross-linguistic incorporation scores appear in the same areas of the tree, they are connected by boxes rather than lines, indicating non-treelike or non-implicational behavior.28 Yet, as will be discussed in the next subsection, the ordering of the verbs in our study does correlate with the verb rankings found in the ValPaL project.

![Figure 1. NeighborNet visualization of the incorporation ability of the verbs in the eight sample languages.](image)

5.4.2.2 Morphosyntactic transitivity

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

languages have multiple translational equivalents with language-internally distinct values for incorporation potential. These transformed data were also used to create the NeighborNet visualization in Figure 1.

28 We thank Alena Witzlack-Makarevich for creating the NeighborNet visualization.
Table 4. Verb meanings included in the questionnaire and studied by Haspelmath (2015) ordered based on their cross-linguistic incorporation score. Transitivity prominence scores are given between parentheses.

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

The Spearman’s rank order coefficient for the ranking of the verb meanings based on their cross-linguistic incorporation scores and their ranking based on their transitivity prominence scores shows that there is a moderate, but statistically significant positive correlation between these rankings: $\rho = 0.56$ ($p < 0.001$). It can be seen in Table 4 that verbs with cross-linguistic incorporation scores of 0.50 and higher have transitivity prominence scores that range between 0.88 and 1.00, with the exceptions of FEEL PAIN.
Incorporation: Constraints on variation

and BE DRY. These latter two verb meanings have very low transitivity scores; hence, they can be regarded as intransitive verbs cross-linguistically and seem to favor incorporation for that reason. FEEL COLD and DIE are two other verbs that belong to the patientive intransitive class, with transitivity prominence scores of 0.00, showing noun incorporation with a considerable cross-linguistic frequency across the sample languages. By contrast, some other members of the patientive intransitive class, such as BE SAD, SINK (INTR.), BURN (INTR.) and BE HUNGRY, appear unexpectedly near the bottom of the table.

The agentive intransitive verbs with a very low transitivity prominence score are generally found to be unlikely to incorporate. This finding is expected based on the preference for incorporation into unaccusative rather than unergative verbs, identified in earlier work (Baker 1988: 81–82, 87–90).29 Thus, JUMP, RUN, PLAY and GO have transitivity prominence scores of 0.10 and lower and are only very rarely used as incorporating verbs in our data. Note also that, when they do incorporate, they incorporate goals or instruments rather than agents. The only agentive intransitive verb that has a relatively high cross-linguistic incorporation score is SIT DOWN. However, the examples of incorporation into verbs meaning SIT DOWN involve an incorporated goal or an incorporated body-part noun, with its possessor expressed as the subject of the construction, rather than an incorporated agent argument.

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

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

29 Notably, Baker (1988) claims that unergative verbs cannot incorporate at all. Here, we rather look at semantically defined agentive and patientive intransitive verbs.
Verb-based restrictions on noun incorporation across languages

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

EAT (0.88), WASH (1.00), GIVE (0.63)
SHAVE (0.67), COOK (0.50)
CUT (0.94), SEARCH FOR (0.63), HIT (0.69)
KILL (0.79), ASK FOR (0.50), TAKE (0.50)
SEE (0.38), THROW (0.75)
BREAK (TR.) (0.69), COVER (0.56), THINK (0.00)
TELL (0.43), KNOW (0.13), HELP (0.00)
SING (0.25)
FEAR (0.00)
PUT (0.71), LEAVE (0.13)
SAY (0.21), NAME (0.25), RUN (0.21)
JUMP (0.25), FRIGHTEN (0.33), PLAY (0.13), BE DRY (0.71)
BURN (INTR.) (0.13), GO (0.06), SINK (INTR.) (0.20)
MEET (0.00), DIE (0.25)
FEEL PAIN (0.75), SIT (0.31)
BE SAD (0.20)
SIT DOWN (0.33), BE HUNGRY (0.00)
FEEL COLD (0.40)

The Spearman’s rank order coefficient for the ranking of the verbs based on their cross-linguistic incorporation score and their ranking on the object dem/del hierarchy (Malchukov 2015; Wichmann 2015) shows a moderate, but statistically significant positive correlation for these rankings ($\rho = 0.54$, $p < 0.001$), just as for the ranking of the verbs based on their cross-linguistic incorporation score and their ranking in terms of transitivity prominence. This is expected given the high correlation between the object dem/del hierarchy and the ranking of the verbs based on their transitivity prominence score reported in Section 5.2.2.2.

5.4.2.3 Other factors
The data summarized in Table 4 and in (10) above suggest that a few factors other than morphosyntactic transitivity also affect the likelihood that a verb allows noun incorporation. Firstly, verbs like KNOW, THINK and SAY have relatively low cross-linguistic incorporation scores compared to other verbs with similar transitivity prominences scores and on similar positions on the object dem/del hierarchy. This finding matches the results for these verbs in the typological survey (see Section 5.3.2.3) and it suggests again that these verbs do not incorporate frequently because
their P-arguments may be complement clauses rather than nouns. Secondly, for the verbs NAME, HELP, MEET and, to a lesser extent, LEAVE and FEAR, which have relatively low cross-linguistic incorporation scores even though they have quite high transitivity prominence scores and appear quite high on the object dem/del hierarchy, it may be noted that their meanings imply a P-argument which is typically human or at least animate. As discussed in Section 5.2.2.1, it has been claimed that verbs that tend to take animate patient arguments are not very likely to incorporate, and animacy may thus also play a role for these verbs. Thirdly, it can be seen in (10) that PUT appears in the lower half of the object dem/del hierarchy but has a quite high cross-linguistic incorporation score. As mentioned in Section 5.3.2.3, this finding matches the outcome of the typological survey and corroborates Mithun’s (1984: 863) observation that verbs with general semantics are likely to allow incorporation. Fourthly, it has often been argued that body-part nouns are frequently incorporated, and indeed we find that many verbs that often involve body-part nouns, like FEEL PAIN and FEEL COLD, incorporate relatively frequently.

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

Interestingly, information about the relative frequency with which verbs appear in incorporation cross-linguistically also reflects some of these additional morphosyntactic, semantic and lexical factors. So far, we have considered the ordering of the verbs in our questionnaire based on their absolute incorporation potential across the sample languages, i.e. on whether or not they are found to allow incorporation, irrespective of the relative frequency with which the verbs actually appear in incorporation constructions in the corpora. Yet, we may also focus on the question how often verbs that allow incorporation actually appear with incorporated nouns. For each verb meaning, we added up all attested occurrences of each translational equivalent that allows noun incorporation across languages and all instances of incorporation constructions with these translational equivalents across languages. A cross-linguistic incorporation ratio can be calculated for each verb meaning by dividing the frequency of the incorporation constructions by the total
frequency of these verbs (see also http://dx.doi.org/10.21942/uva.12161769). While it is hard to interpret a verb scale based on these ratios, since some of the ratios are based on far fewer data points than others, some general observations still seem worthwhile.

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

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

Finally, there are verb meanings whose translational equivalents seem to display widely varied behavior across, but also within, languages, suggesting idiosyncratic restrictions. As an example of cross-linguistic variation, consider FALL. This verb is found to incorporate in four languages: Baure, Movima, Plains Cree and Yucatec Maya. In Baure and Movima it does so every time it occurs in the corpus, i.e. 45 out of 45 cases for Baure and 98 out of 98 cases for Movima. By contrast, although FALL can incorporate in Plains Cree and Yucatec Maya (as attested on the basis of

---

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

dictionary data and speaker elicitation), none of the 24 occurrences of this verb in the Plains Cree corpus and none of the 72 occurrences of this verb in the Yucatec Maya corpus actually involves incorporation. As for intra-linguistic variation, recall that for some verb meanings in some languages we considered multiple translational equivalents. We have already mentioned that such near-synonyms may show differential behavior in terms of their absolute (in)ability to incorporate. Such differences may also show up in relative terms when two translational equivalents can both incorporate, but do so to different degrees. For instance, in Guarayu the verbs -\textit{nupa} ‘hit (hard)’ and -\textit{ipete} ‘hit’ show incorporation in one out of 11 and six out of seven cases, respectively. It must be emphasized again, however, that in many cases the low absolute numbers prevent meaningful comparisons and firm conclusions.

5.4.2.4 Variation across languages

The previous subsections primarily discussed the attested variation from the perspective of the different verbs, but we may also focus on variation across languages. As the document on http://dx.doi.org/10.21942/uva.12161769 shows, the eight languages can be ordered in terms of their language-specific incorporation score, i.e. the number of verbs attested with incorporation, relative to the total number of verb meanings in that language for which data points are available. These scores lead to the ranking in (11).

(11) Language ranking based on incorporation scores across verbs

\begin{align*}
\text{Guarayu} & \ (0.59) \ > \ \text{Movima} \ (0.51) \ > \ \text{Kalamang} \ (0.45) \ > \ \text{Baure} \ (0.43) \ > \\
\text{Yucatec Maya} & \ (0.40) \ > \ \text{Ese Ejja} \ (0.38) \ > \ \text{Iraqw} \ (0.32) \ > \ \text{Plains Cree} \ (0.31)
\end{align*}

Based on the available frequency information, language-specific incorporation ratios across verbs can be calculated as well, by dividing all instances of incorporation found in the corpus of a language across all incorporating verbs in that language by the total number of occurrences of these verbs taken together. According to these language-specific incorporation ratios, which are also included in the document on http://dx.doi.org/10.21942/uva.12161769, the ordering of the languages is as follows:

(12) Language ranking based on incorporation ratios across verbs

\begin{align*}
\text{Movima} \ (43.21\%) & \ > \ \text{Baure} \ (19.99\%) \ > \ \text{Guarayu} \ (13.38\%) \ > \ \text{Ese Ejja} \\
(13.03\%) & \ > \ \text{Plains Cree} \ (13.00\%) \ > \ \text{Kalamang} \ (8.16\%) \ > \ \text{Yucatec Maya} \\
(4.88\%) & \ > \ \text{Iraqw} \ (3.61\%)
\end{align*}

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

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

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

5.5 Discussion and conclusion

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

First of all, the results of the two sub-parts of the study, i.e. the typological survey and the questionnaire-based case studies, show considerable overlap and largely confirm the predictions based on earlier studies reviewed in Section 5.2.2.2: morphosyntactically highly transitive verbs and – to a lesser extent – patientive intransitive verbs are cross-linguistically most likely to incorporate, while agentive intransitive verbs are unlikely to do so.

Furthermore, although in the questionnaire-based part of the study both the sample size and the size of the individual language corpora is relatively restricted, the
Incorporation: Constraints on variation

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

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

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

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

6.1 Theoretical approaches to incorporation

This thesis focuses on the cross- and intra-linguistic variation found in incorporation and on the ways in which this variation is constrained. The four studies included in Chapter 2 to 5 show that the variation between incorporation constructions with respect to their pragmatic, semantic, morphosyntactic, phonological and lexical properties is pervasive. At the same time, the variation within each of the domains investigated is constrained. While some of the constraints appear to be relatively arbitrary and language-specific, others are cross-linguistically systematic, taking the form of implicational hierarchies.

The thesis thus provides new insights into the range of variation in incorporation constructions across languages, and these insights have important implications for theoretical approaches to incorporation. A comprehensive account of incorporation should be able to describe the full range of variation occurring in incorporation constructions, while at the same time explaining its limits. In this chapter I therefore examine to what extent the different theoretical accounts of incorporation proposed in the literature are able to capture or predict the variation and constraints presented in the thesis. In this way, the explanatory potential of these theoretical accounts can be assessed.

In the literature on incorporation, a general distinction is made between two types of theoretical approaches to incorporation or, more specifically, noun incorporation, which differ in the component of the linguistic system where incorporation constructions are assumed to be formed (Mithun 1994: 5025; Massam 2009: 1083–1086, 2017; Štekauer et al. 2012: 43–47; Murasugi 2014: 286–288; Haugen 2015: 415–421). On the one hand, studies including Sapir (1911), Mithun (1984, 1986a), Di Sciullo and Williams (1987), Rosen (1989) and Anderson (2000, 2005) take a lexical approach to incorporation. Very generally speaking, these studies claim that incorporation is a lexical word-formation process similar or identical to lexical compounding. According to these lexical accounts, incorporated nouns and incorporating verbs do not have any independent syntactic status. Sadock (1985, 1986, 1991), Baker (1988, 1996, 2009) and Barrie and Mathieu (2016), on the other hand, maintain that incorporation is a syntactic process, arguing that incorporated nouns and incorporating verbs are syntactic constituents. In addition, some frameworks, such as Distributed Morphology, do not allow for lexical processes creating new words but instead propose that all word formation is syntactic (Harley 2009: 129). Consequently, accounts of incorporation in these frameworks automatically predict that compounding and incorporation are both syntactic in nature (Harley 2009; Witschko 2009). Finally, the Functional Discourse Grammar (FDG, Hengeveld and Mackenzie
Incorporation: Constraints on variation

2008) approach to incorporation taken in this thesis can also be described as syntactic. It proposes that incorporation takes place in FDG’s Grammatical Component, rather than in its Lexicon. Thus, in contrast to the lexical approaches but in the same way as the other syntactic approaches just discussed, the FDG approach proposes that incorporated nouns and incorporating verbs have a syntactic status.

In the next sections I consider the results of the studies presented in Chapter 2 to 5 in light of the different theoretical accounts of incorporation. The major findings from Chapter 2 to 5 are presented in the summary included at the end of the thesis. I discuss the findings of each chapter in turn, comparing their compatibility with the lexical approaches to incorporation, on the one hand, and the syntactic approaches, on the other hand. Finally, Section 6.6 concludes to what extent the thesis’ findings support the lexical and syntactic approaches overall.

6.2 Basic settings and implicational hierarchies

Chapter 2 captures the pragmatic, semantic and phonological variation found in previous literature on noun incorporation in two types of constraints. In some cases, it is necessary to use basic settings, while in others the observed patterns of variation can be caught in implicational hierarchies. In FDG, a basic setting specifies a limited number of options from which each individual language makes a choice. It accounts for the type of variation where languages simply vary with respect to the particular option they select. A basic setting differs from an implicational hierarchy, which expresses a cross-linguistic pattern in some domain of variation in which two or more options are implicationally related. Basic settings and implicational hierarchies both result in constraints on the variation in incorporation constructions.

Basic settings are in line with lexical as well as syntactic approaches to incorporation. Basic settings only state that some languages select one option, while others choose another, without making any claims about how the distribution may be explained. As such, they occur, in some form, in all linguistic theories. Correspondingly, the basic settings regarding noun incorporation formulated in Chapter 2 can be accounted for both in the lexical and in the syntactic approaches to incorporation. Note that syntactic accounts of incorporation within the generative framework tend to describe basic settings such as the ones discussed in Chapter 2 as parameters that are relevant for noun incorporation constructions (Baker 1996: 283; Harley 2009: 141; Barrie and Mathieu 2016: 8).

Implicational hierarchies may also be expected on the basis of both the lexical and the syntactic approaches to incorporation. Since implicational hierarchies indicate that the variation in a particular domain shows a cross-linguistically systematic pattern, they fit well with syntactic accounts of incorporation that consider

---

1 Technical terms as applied in FDG are capitalized (see Hengeveld and Mackenzie 2008: 44).
incorporation a regulated productive process, such as the ones proposed by Sadock (1985: 399, 1991: 85) and Baker (1988: 80). Moreover, in FDG it is explicitly claimed that interface conditions tend to take the form of implicational hierarchies (Hengeveld and Mackenzie in prep.), and implicational hierarchies are proposed for various domains of grammar (Hengeveld and Mackenzie 2008: 31–37).

At the same time, constraints in the form of implicational hierarchies can also be found in lexical accounts of incorporation. One of the primary claims made by Mithun (1984: 848, 874, 890; see also Section 4.5) is that there are four functional types of noun incorporation, i.e. “lexical compounding” (Mithun 1984: 848), “the manipulation of case” (Mithun 1984: 856), “the manipulation of discourse structure” (Mithun 1984: 859) and “classificatory noun incorporation” (Mithun 1984: 863), that are implicationally related. In addition, Mithun (1984: 875) posits that the incorporation of “patients of intransitive V’s [verbs]” is dependent on the incorporation of “patients of transitive V’s”, while incorporated “instruments and/or locations” only occur in languages that also show the incorporation of patients of transitive and intransitive verbs.

**6.3 Morphosyntactic forms of incorporated elements**

In Chapter 3 I focus on the variation in the morphosyntactic forms of incorporated elements and a possible constraint on this variation. The main findings are that incorporated elements may be single lexical morphemes, i.e. bound roots and unbound stems, derived stems, inflected words and phrases, and that these forms of incorporated elements are implicationally related. This implicational relationship indicates that the distribution of the different forms is cross-linguistically systematically constrained.

The findings concerning the morphosyntactic forms of incorporated elements seem to present a challenge to the lexical and syntactic accounts of incorporation proposed in earlier literature. Proponents of lexical accounts of incorporation assume that incorporated nouns always take the form of a stem (Sapir 1911: 257; Mithun 1984: 847, 1986a: 32; Anderson 2000: 13–14, 2005: 258), just as compounding is claimed to pertain to stems (Mithun 1984: 848). Incorporated nouns are thus said not to take inflectional marking for number, definiteness or case (Mithun 1984: 849, 859). The idea that incorporated nouns can only be stems is in correspondence with the “thesis of the atomicity of words” (Di Sciullo and Williams 1987: 89), also known as the lexicalist hypothesis, which states that “[t]he internal structure of words is not accessible to the syntax” (Ackema and Neeleman 2002: 93; see also Anderson 2005: 264). As the formation of phrases is a syntactic process and inflection also involves syntax, phrase formation and inflection are not expected to affect internal components
Incorporation: Constraints on variation

of words. Because, in lexical approaches, lexical processes are assumed to take place before syntax, these approaches predict that elements that have undergone inflection or phrase formation are not available for incorporation. Nevertheless, Chapter 3 shows that not only simple and derived stems can be incorporated, but incorporated inflected words and incorporated phrases occur as well. Therefore, the lexical accounts’ assumption that incorporation is limited to stems does not reflect the variation in forms attested in this thesis.

Even though the syntactic accounts of incorporation do not support the thesis of the atomicity of words, their predictions about the possible forms of incorporated elements are also quite restricted. In Sadock’s (1985: 383–384, 1991: 100–101) autolexical syntax model, incorporation is assumed to result from a mismatch between morphological and syntactic representations, such that incorporation creates constructions “in which some of the morphological constituents can be shown to have independent syntactic reality” (Sadock 1991: 100). More specifically, however, Sadock (1991: 105) states that incorporation concerns the combination of a lexeme, i.e. an incorporating verb, with a stem. In this way, he limits incorporation to stems. Sadock (1985, 1991) also discusses constructions in which lexemes are combined with inflected words. In the present thesis, such constructions are considered incorporation constructions with incorporated inflected words. However, Sadock (1991: 105) regards these constructions as involving cliticization (see also Anderson 2005), which he describes as a phenomenon similar to but distinct from incorporation. Analyzing the findings from Chapter 3 according to Sadock’s model thus entails that the incorporation of stems and the incorporation of inflected words are two separate processes, i.e. incorporation and cliticization respectively. From a typological perspective, however, there is reason to assume that this separation is undesirable. The data in Chapter 3 show that the occurrence of incorporated simple stems and derived stems, on the one hand, and the occurrence of incorporated inflected words, on the other hand, are implausibly related, which suggests that the incorporation of elements of different forms result from a single process.

Baker’s (1988: 71–72; see also Barrie and Mathieu 2016: 36) syntactic account of incorporation also limits the incorporation process to simple and derived stems. Baker (1988, 1996, 2009) proposes that incorporation constructions are formed via syntactic head movement, and because this movement procedure is restricted to

2 Some theories may consider inflection as a lexical process. If such a theory is adopted, the incorporation of inflected words may be explained, but the incorporation of phrases remains difficult to account for.

3 Note also that none of the lexical accounts discussed (Sapir 1911; Mithun 1984, 1986a; Di Sciullo and Williams 1987; Rosen 1989; Anderson 2000, 2005) pays much attention to incorporated elements of parts of speech other than nouns.

4 Although Sadock (1986, 1991: 82–83, 99–100) focuses on incorporation constructions showing a mismatch between morphological and syntactic representations, he also recognizes a purely morphological type of incorporation. In this type of incorporation, incorporated inflected words and phrases are not expected either, as incorporated elements in this morphological type of incorporation do not have any syntactic reality.
uninflected stems, his account automatically predicts that the stem is the only possible form of incorporated elements. Barrie and Mathieu (2016), by contrast, present an analysis of incorporation as phrasal movement, which also allows for the incorporation of more complex morphosyntactic forms. Barrie and Mathieu (2016: 9) argue that incorporation involves phrases of the types “nP (categorized/nominalized stems), dP (modified N-stem), DP (possessor DPs, demonstratives), KP (case-marked nominals), and CP (relative clauses)”. In this way, they can easily account for the incorporation of derived stems, inflected words and phrases. In addition, because nPs may consist of an acategorical root and a non-overt categorizing n, their analysis of incorporation as phrasal movement also allows for the incorporation of single lexical morphemes that may be argued to correspond to acategorical roots with non-overt categorizing ns.

Barrie and Mathieu (2016) thus seem able to account for all types of forms attested as incorporated elements. However, their analysis of incorporation as phrasal movement also makes some predictions that are not necessarily or fully consistent with the data from Chapter 3. Firstly, the claim that incorporated nouns are minimally nPs leads to the expectation that all incorporating languages showing derived stems consisting of a root and an overt categorizing n, i.e. derived stems corresponding to nPs, allow the incorporation of these derived stems. However, many languages, including seven of the 30 sample languages studied in Chapter 3, limit incorporation to single lexical morphemes. In these languages, it is not obvious that incorporated nouns are minimally nPs and not acategorical roots. Barrie and Mathieu (2016: 4, 23) do discuss incorporation-like constructions involving acategorical roots, but distinguish these from incorporation constructions formed via phrasal movement by considering them cases of lexical root-root merger. In this way, they exclude them from the domain of incorporation and regard them as resulting from a separate process. However, the distinction between these two processes involving phrasal material vs. acategorical roots does not predict the implicational relationship found in Chapter 3 between incorporated single lexical morphemes on the one hand and more complex incorporated elements on the other hand. This relationship indicates that incorporated elements of different forms are interdependent, which suggests that they belong to a single linguistic phenomenon. Secondly, because n is taken to provide referentiality (Barrie and Mathieu 2016: 14, 29, 40), the phrasal-movement account of incorporation is restricted to referential nouns, while acategorical roots involved in lexical root-root merger must be non-referential. In languages such as Iraqw and Movima, however, incorporated nominal elements all function non-referentially, even though they may take the form of derived stems. This finding poses an additional challenge to the separation between incorporated nouns corresponding to nPs and

---

5 Barrie and Mathieu’s (2016: 34) examples of CP incorporation involve headless relative clauses, which in the analysis of Chapter 3 would be considered phrases.
incorporated nouns corresponding to lexically merged acategorical roots that follows from Barrie and Mathieu’s (2016) account.

In contrast to Barrie and Mathieu (2016), Wiltschko (2009: 220–221) argues that both nPs and acategorical roots may be incorporated syntactically. In her analysis, there is no strict separation between the incorporation of single lexical morphemes corresponding to acategorical roots and other incorporated elements. Thus, her analysis can capture the existence of an implicational relationship between incorporated single lexical morphemes and incorporated elements of more complex forms. Like Barrie and Mathieu (2016: 14, 29, 40), however, Wiltschko (2009: 210) proposes that nPs are referential on the basis of their categorizing n, such that incorporated derived stems, which are necessarily minimally nPs, must be referential. As described above, languages like Iraqw and Movima show that this relation between referentiality and incorporated derived stems does not always hold.

Finally, the FDG approach to incorporation differs from the syntactic accounts discussed so far, in being able to capture all forms of incorporated elements attested as well as their implicational relationship. FDG’s maximal template for morphosyntactic words allows for incorporated lexical morphemes, derived stems, words, phrases and clauses, and all these forms, except for incorporated clauses, are found. In addition, because only a single incorporation process is recognized, i.e. the insertion of elements of different forms in a morphosyntactic template, the implicational hierarchy describing the distribution of incorporated elements with different forms is not unexpected.

6.4 Pragmatic referentiality and semantic modifiability
In Chapter 4 I investigate the pragmatic referentiality and semantic modifiability of incorporated nouns. The data show that incorporated nouns may be referentially used modifiable nouns (+R/+M nouns), non-referentially used modifiable nouns (−R/+M nouns) and non-referentially used non-modifiable nouns (−R/−M nouns). While the incorporation of +R/+M nouns and the incorporation of −R/−M nouns appear to be independent phenomena, the incorporation of −R/+M nouns is demonstrated to be implicationally related to the incorporation of +R/+M nouns.

Examining to what extent the different accounts of incorporation proposed in earlier literature predict the attested variation between these different pragmatic-semantic types of incorporated nouns is not a straightforward matter. As mentioned in Chapter 4, researchers focusing on incorporation use the notions of referentiality and modifiability in different ways. Consequently, the correspondence between theoretical accounts of incorporation and the pragmatic-semantic variation between incorporated nouns found cannot be evaluated directly. Instead, their notions of referentiality and modifiability first have to be translated to the notions of referentiality and modifiability as used in the present thesis.
For instance, in contrast to the analysis used in Chapter 4, Mithun (1984: 865–867, 870; see also Di Sciullo and Williams 1987: 65–66; Rosen 1989: 297–301 [for nouns in "Classifier NI"] does not consider modifiers that occur external to classificatory noun incorporation constructions as modifiers of the incorporated nouns. Instead, she views these modifiers as independent noun phrases without a lexical head. As she also states that incorporated nouns always take the form of stems (Mithun 1984: 847, 1986a: 32; see also Section 6.3 above), her account seems to predict that incorporated nouns are non-modifiable. In addition, Mithun (1984: 865–867, 869–871) does not regard anaphoric reference or co-reference, which are possible for nouns in classificatory noun incorporation, as evidence for the referentiality of incorporated nouns. Moreover, her analysis of modifiers external to incorporation constructions entails that demonstratives appearing outside incorporation constructions of the classificatory noun incorporation type are not taken to indicate the referentiality of incorporated nouns either. And as incorporated nouns in the other functional types of noun incorporation identified by Mithun (1984), i.e. lexical compounding, the manipulation of case and the manipulation of discourse structure, do not show any referential characteristics (Mithun 1984: 849–850, 855–856, 859), Mithun appears to analyze all incorporated nouns as non-referential.

In this way, Mithun’s (1984) account seems to be limited to the incorporation to −R/−M nouns. However, applying the criteria for referentiality and modifiability employed in Chapter 4 to Mithun’s types of incorporation leads to a different picture. Importantly, because classificatory noun incorporation constructions can be combined with external modifiers and their incorporated nouns can be involved in anaphoric reference and co-reference, they can be considered +R/+M incorporated nouns. Mithun (1984) can then account for both +R/+M incorporation, corresponding to classificatory noun incorporation, and −R/−M incorporation, corresponding to her other three functional types of incorporation. Nevertheless, Mithun’s (1984: 848, 874, 890) claim that classificatory noun incorporation occurs only in languages that also show the other three types of noun incorporation appears to be in conflict with the finding from Chapter 4 that +R/+M incorporated nouns and −R/−M incorporated nouns appear in languages independently of each other. In addition, −R/+M incorporated nouns are not recognized in Mithun’s (1984) work.

In contrast to Mithun (1984), Anderson (2000: 18–19, 2005: 266–267) explicitly acknowledges the occurrence of referential incorporated nouns. He claims that the analysis of incorporation as a lexical compounding process has to be adapted to accommodate the incorporation of such referential nouns. According to Anderson (2000: 18, 2005: 267), a lexical compounding analysis of incorporation generally entails that an incorporated noun “saturates” or “satisfies” one of the arguments of the verb, typically the argument that otherwise would be expressed as a direct object. The resulting verb then becomes intransitive and the incorporated noun internally binds its “R-role”. In this way, it becomes impossible for the incorporated noun to be
Incorporation: Constraints on variation

In order to explain the occurrence of incorporated referential nouns, Anderson (2000: 18–19, 2005: 267) proposes that it is also possible that incorporation only “‘unifies’ the semantics of the noun with the argument position of the verb, but without saturating the argument itself” (Anderson 2005: 267). In this case, the incorporating verb can still take a regular direct object noun phrase external to the incorporation construction, which may be referential. The referential function of the incorporated noun is then due to the referentiality of the direct object. Importantly, the direct object may be a full external noun phrase, but may also remain unexpressed or lack a lexical head. In this way, Anderson (2000, 2005) analyzes the modifiers appearing external to incorporation constructions as independent noun phrases without a lexical head, just like Mithun (1984).

Anderson’s unifying noun incorporation thus shows similarities to Mithun’s classificatory noun incorporation: the constructions can co-occur with external modifiers and the incorporated nouns can be associated with referentiality. This type of incorporation thus corresponds to the incorporation of +R/+M nouns in the analysis of the present thesis. As saturating noun incorporation does not allow external modifiers and contrasts with unifying noun incorporation with respect to referentiality, this type of noun incorporation matches −R/−M incorporation. As noted for Mithun’s account above, this means that the occurrence of −R/+M incorporated nouns is not predicted by Anderson’s lexical account.

Just like the lexical accounts of incorporation, most syntactic accounts capture +R/+M incorporated nouns and −R/−M incorporated nouns, but not −R/+M ones. Sadock (1985: 398–399, 1986: 22–28, 1991: 86–99) observes that at least some incorporated nouns in languages like Kalaallisut and Southern Tiwa introduce discourse referents, which may function as antecedents, and can combine with external modifiers. Primarily based on such observations, he claims that incorporated nouns may have syntactic reality in these languages. In other languages, however, incorporation is a purely morphological word-formation process, such that incorporated nouns are necessarily non-referential and cannot be combined with modifiers (Sadock 1986: 19, 30, 1991: 82–83, 99–100). Thus, while incorporated nouns in Sadock’s syntactic incorporation are +R/+M nouns, his morphological incorporation matches −R/−M incorporated nouns.

According to Baker’s (1988, 1996, 2009) syntactic head-movement analysis of incorporation, noun incorporation involves referential nouns that can be modified by modifiers external to the incorporation constructions. Baker (1988: 78–79, 1996: 287–291) notes that an incorporated noun may refer to a specific entity, have a definite interpretation, function as the antecedent for anaphoric reference and be co-referential

---

6 Anderson’s (2000, 2005) distinction between unifying and saturating noun incorporation also resembles Rosen’s (1989) distinction between Classifier NI and “Compound NI” (Anderson 2000: 23, 2005: 267, 272), except that Rosen does not relate this distinction to the potential referentiality of incorporated nouns.
with another instance of the same noun. Because Baker (1988: 81) assumes that nouns are referential when they function as the head of a noun phrase, his analysis of incorporation as head movement can straightforwardly explain these referential characteristics of incorporated nouns. In addition, the head-movement account predicts that incorporated nouns may leave modifiers behind (Baker 1988: 92–105, 1996: 308), which entails that incorporated nouns may be modified by external modifiers. Baker’s head-movement analysis of incorporated nouns thus corresponds to the incorporation of +R/+M nouns. As noun-verb constructions formed via head movement are the only ones that Baker considers to involve incorporation, his account seems to restrict noun incorporation to +R/+M nouns. Nevertheless, he acknowledges that lexical noun-verb compounding exists in languages as well (Baker 1988: 78–80, 1996: 307, 329). Noun-verb compounds look similar to incorporation constructions but contain nouns that are non-referential and cannot be modified by external modifiers (Baker 1988: 78, 96, 1996: 307, 329), i.e. they involve −R/−M nouns. Although Baker (1988: 78, 96, 1996: 329–330) excludes these constructions from the domain of incorporation because he assumes that they do not involve syntactic movement, the identification of these constructions entails that Baker can account for both the +R/+M and the −R/−M incorporated nouns attested in Chapter 4. Moreover, he identifies both languages that only show lexical noun-verb compounding, such as English, and languages that use both syntactic noun incorporation and lexical noun-verb compounding, such as Mohawk (Baker 1988: 78, 1996: 307, 329). In this way, his analysis is also consistent with the intra-linguistic variation between +R/+M incorporated nouns and −R/−M incorporated nouns. The occurrence of −R/+M is, however, not expected to occur.

Barrie and Mathieu’s (2016: 40) phrasal-movement analysis of incorporation is similar to Baker’s head-movement analysis in that it predicts that incorporated nouns are referential and can be modified. In Barrie and Mathieu’s (2016: 9, 14, 40) analysis, incorporated nouns are assumed to be referential because they are minimally nPs, i.e. they contain a nominalizing n that introduces referentiality. Incorporated nouns may also be dPs, DPs, KPs and CPs (Barrie and Mathieu 2016: 9), which entails that incorporated nouns may be incorporated together with their modifiers (see e.g. Barrie and Mathieu 2016: 18–19, 31–34, 36–37). Additionally, external modifiers are recognized in Barrie and Mathieu’s (2016: 4, 40) account. Incorporated nouns thus appear to be of the +R/+M type. As noted in Section 6.3, Barrie and Mathieu (2016: 4, 23) also identify constructions that look like incorporation but which they consider to involve lexical root-root merger. Because they do not involve syntactic movement, they are not regarded as incorporation by Barrie and Mathieu (2016). Nevertheless, the acategorial roots in these constructions do match −R/−M incorporated nouns, because they cannot be modified and lack referentiality. In this way, Barrie and Mathieu’s (2016) work can account for +R/+M incorporated nouns as well as for −R/−M incorporated nouns.
Importantly, none of the accounts discussed seems to recognize −R/+M incorporated nouns such as the ones shown in example (1) from Ket and in example (2) from Kalaallisut.\(^7\)

(1) Incorporation of a −R/+M noun in Ket  
\(\text{tab-ay-t-o-n-aq}\)  
\(\text{dog.PL-3PL.AN.SBJ-TC-PST-PST-become}\)  
‘They turned into dogs.’  
(Vajda 2017: 918)

(2) Incorporation of a −R/+M noun in Kalaallisut  
\(\text{savaatili-ngur-putin=nguuq pikkuris-suq}\)  
\(\text{sheep.herder-become-2SG.IND=QUOT be.capable-INTR.PTCP}\)  
‘They say you’ve become a capable sheep-herder.’  
(Fortescue 1984: 71)

In example (1) and (2), the incorporated nouns ascribe an entity, i.e. they are used non-referentially, and they are modified by grammatical plural marking and an external participle respectively. One plausible reason why the various lexical and syntactic approaches do not seem to account for −R/+M incorporated nouns is that most incorporation research has focused on incorporated nouns functioning as semantic arguments. Prototypically, incorporated nouns correspond to arguments that in a context without incorporation would be expressed as the direct object of a transitive verb or as the subject of an intransitive verb (Gerdts 1998: 86–87; Mithun 2000: 917; Lehmann and Verhoeven 2005: 918; Sadock 2006: 585; Aikhenvald 2007: 19; Borik and Gehrke 2015: 2; Haugen 2015: 414–415). Some studies have also discussed incorporated nouns matching instruments and locations (Gerdts 1998: 86–87; Mithun 2000: 917; Lehmann and Verhoeven 2005: 918; Sadock 2006: 585; Aikhenvald 2007: 19; Murasugi 2014: 284; Haugen 2015: 414–415), but little attention has been paid to incorporated nouns functioning as predicates. Nevertheless, predicatively used nouns can be incorporated, as shown by examples (1) and (2) and the findings in Section 4.4, and these are often of the −R/+M type. The thesis thus indicates that, in order to capture the full range of variation in incorporated nouns, theoretical work on incorporation should not only concentrate on incorporated arguments, instruments and locations, but include incorporated nominal predicates as well.

In contrast to the lexical and syntactic approaches discussed above, the FDG approach taken in the present thesis does predict the occurrence of −R/+M

\(^7\) Glosses in the examples are adapted to the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/resources/glossing-rules.php).

incorporated nouns. The thesis investigates the referentiality and modifiability of incorporated nouns using FDG’s three-way typology of pragmatic-semantic usages of common nouns which includes +R/+M nouns, −R/+M nouns and −R/−M nouns. On the basis of this typology, −R/+M incorporated nouns are expected just as well as +R/+M and −R/−M incorporated nouns. Consequently, the FDG approach can capture the full variation with respect to the referentiality and modifiability of incorporated nouns.

6.5 Verb-based lexical restrictions
Chapter 5 shows that morphosyntactically highly transitive verbs and patientive intransitive verbs are cross-linguistically most likely to be able to incorporate nouns. However, this pattern appears to be a tendency only, i.e. it cannot be used reliably to predict whether a particular verb allows noun incorporation or not. Instead, the ability of a particular verb to incorporate nouns seems to be largely idiosyncratically determined.

Idiosyncratic restrictions on noun incorporation are in line with lexical accounts of incorporation. Mithun (1984: 858, 860, 863–864, 889) notes that noun incorporation, like other word-formation processes, is not fully productive. Some nouns and verbs are more prone to incorporation than others, and incorporation seems to be restricted to particular nouns and/or verbs in a partly arbitrary way. According to Mithun (1984: 872, 889), noun incorporation constructions tend to be lexicalized and this then explains how speakers know which constructions are possible and which are not. Anderson (2000: 15–16, 2005: 265) further claims that lexical restrictions on incorporation are easily understood if incorporation is assumed to be a lexical rule. This lexical rule would then only apply to lexical items for which it is specified, in the lexicon, that they are amenable to incorporation.

Lexical accounts can thus easily explain the lexical restrictions found in Chapter 5. However, syntactic accounts seem to be able to capture these as well. Sadock (1985: 399, 410, 1991: 85) emphasizes for syntactic incorporation in Kalaallisut that some verbs incorporate obligatorily, some verbs incorporate optionally and some verbs do not allow incorporation at all. Correspondingly, for each verbal bound root in Kalaallisut, he claims that “lexical features […] determine its applicability” (Sadock 1985: 399). Incorporation can thus be described as a regular, productive process applying to all verbs for which it is specified lexically that they are available for incorporation. Sadock (1991: 83–85) further supports this view by arguing that, at least in Kalaallisut, the number of possible incorporation constructions is so high that it is unlikely that all incorporation constructions are formed and stored lexically, as this would result in an enormous lexicon.

Baker (1988, 1996, 2009) does not discuss restrictions on the verbs that occur in incorporation constructions but rather emphasizes that incorporation is highly
Incorporation: Constraints on variation

productive (Baker 1988: 80). Although this characterization of incorporation does not exclude the possibility that the lexicon contains limitations on incorporation for a few verbs, the large number of lexical restrictions in some incorporating languages is not directly predicted by a general characterization of incorporation as a very productive process. On the other hand, as described in Section 6.4, Baker (1988: 78–80, 1996: 307, 329) assumes that lexical noun-verb compounding may exist next to syntactic incorporation. Importantly, in the case of lexical compounding idiosyncratic restrictions may be expected, just as in the lexical accounts.

Barrie and Mathieu (2016: 10, 12, 16, 24, 33) also mention the productivity of the incorporation process. At the same time, they observe that incorporation “is typically never fully productive in any language” and that “[t]here are always lexical exceptions in many areas of the grammar” (Barrie and Mathieu 2016: 37). In addition, they show that the productivity of incorporation may be different for lexical verbs on the one hand and light verbs on the other (Barrie and Mathieu 2016: 10). Lexical verbs optionally incorporate nouns, i.e. they are unbound stems, while light verbs obligatorily incorporate nouns, i.e. they correspond to bound roots as defined in this thesis. According to Barrie and Mathieu (2016: 10), in a particular language incorporation may be productive with both types of verbs, may be free with lexical verbs and restricted with light verbs or may be limited to light verbs. They suggest that the possibilities with respect to the obligation or optionality to incorporate into lexical verbs and light verbs are specified lexically.

Finally, the FDG approach to incorporation resembles the other syntactic approaches mentioned here in arguing that incorporation is a productive process without excluding the possibility that this process is to some extent lexically restricted.

6.6 Conclusion

This thesis demonstrates that incorporation constructions show much variation across and within languages, while, at the same time, this variation is clearly constrained.

Firstly, several pragmatic, semantic and phonological domains of variation addressed in previous research can be captured in basic settings and implicational hierarchies.

Secondly, with respect to the morphosyntactic forms of incorporated elements, there is variation between lexical morphemes, derived stems, inflected words and phrases, but these different forms are implicationally related.

Thirdly, pragmatically and semantically, variation can be found between +R/+M, −R/+M and −R/−M incorporated nouns. While +R/+M and −R/−M incorporated nouns occur in languages independently of one another, the occurrence of −R/+M incorporated nouns is dependent on the occurrence of +R/+M incorporated nouns in the same language. Finally, noun incorporation may be possible with many different incorporating verbs, but is restricted lexically in that, in most languages, some but not all verbs can incorporate nouns. Morphosyntactically highly transitive verbs and patientive verbs
are generally most likely to allow for noun incorporation, but the restrictions are mostly idiosyncratic.

These findings have some clear implications for theoretical approaches to incorporation. Although the FDG approach to incorporation generally seems to be consistent with the data concerning each of the domains of incorporation investigated, the other lexical and syntactic incorporation accounts discussed can explain some of the findings more easily than others. On the one hand, the use of basic settings and implicational hierarchies to describe domains of variation and the existence of both +R/+M incorporated nouns and −R/−M incorporated nouns are in agreement with each of these accounts. On the other hand, findings such as the occurrence of −R/+M incorporated nouns in various languages and the interdependency between the incorporation of −R/+M nouns and the incorporation of +R/+M nouns present challenges to the lexical as well as the syntactic approaches. Importantly, some domains of variation and their constraints pose problems for the lexical and syntactic approaches to different degrees. Such a difference can be found primarily with respect to the morphosyntactic forms of incorporated elements. At least part of the variation in the forms of incorporated elements found in the thesis can be captured by some syntactic approaches, while none of the lexical accounts predicts any variation in this domain. For this reason, the syntactic approaches to incorporation seem better suited to account for the findings of the present thesis than the lexical ones.

All in all, the thesis shows that some parts of the variation found in incorporation constructions and some of the constraints on this variation cannot be fully explained by any of the traditional lexical and syntactic approaches to incorporation. The thesis thus indicates that theoretical work on incorporation may generally benefit from paying more attention to the full range of variation in incorporation constructions and the constraints that govern this variation.
References

Incorporation: Constraints on variation


References


180  Incorporation: Constraints on variation


Hartmann, Iren, Martin Haspelmath & Bradley Taylor (eds.). 2013. 
Valency Patterns 


Haspelmath, Martin. 2008. Creating economical morphosyntactic patterns in language change. In Jeff Good (ed.), 

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. 

Haspelmath, Martin. 2015. Transitivity prominence. In Andrej Malchukov & Bernard Comrie (eds.), 
Valency classes in the world’s languages, vol. 1: Introducing the framework, and case studies from Africa and Eurasia (Comparative Handbooks of Linguistics 1(1)), 131–147. Berlin: De Gruyter Mouton.

Haspelmath, Martin. 2018. The last word on polysynthesis: A review article. 
Linguistic Typology 22(2). 307–326.

Valency classes in the world’s languages, vol. 1: Introducing the framework, and case studies from Africa and Eurasia (Comparative Handbooks of Linguistics 1(1)), 41–71. Berlin: De Gruyter Mouton.


Haugen, Jason D. 2008. Denominal verbs in Uto-Aztecan. 


Incorporation: Constraints on variation


Incorporation: Constraints on variation


References


Incorporation: Constraints on variation


Olthof, Marieke & Kees Hengeveld. subm. Noun incorporation in Functional Discourse Grammar.
References


Say, Sergey. in prep. The antipassive derivation and the lexical meaning of the verb.


Appendix 1 List of incorporating languages

This table presents the list of incorporating languages that was compiled as described in Section 1.3, 3.4.1 and 4.3.1. The names of the languages, their family classifications, macro-areas and countries are based on Glottolog (Hammarström et al. 2017). 

<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</td>
</tr>
<tr>
<td>Gawwada</td>
<td>Afro-Asiatic</td>
<td>Africa</td>
<td>Ethiopia</td>
</tr>
<tr>
<td>Iraqiw</td>
<td>Afro-Asiatic</td>
<td>Africa</td>
<td>Tanzania, United Republic of</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; 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>Guimuz</td>
<td>Africa</td>
<td>Ethiopia; Sudan</td>
</tr>
<tr>
<td>//Ani</td>
<td>Khoek-Kwadi</td>
<td>Africa</td>
<td>Botswana</td>
</tr>
<tr>
<td>Kxoe</td>
<td>Khoek-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>Soninke¹</td>
<td>Mande</td>
<td>Africa</td>
<td>Gambia; Guinea; Guinea-Bissau; Mali; Mauritania; Senegal</td>
</tr>
<tr>
<td>Anindilyakwa</td>
<td>Guinwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Bininj Kun-Wok</td>
<td>Guinwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Ngalakan</td>
<td>Guinwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Ngalkbin</td>
<td>Guinwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Ngandi</td>
<td>Guinwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Rembarrnga</td>
<td>Guinwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Warray</td>
<td>Guinwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Wubuy</td>
<td>Guinwinyguan</td>
<td>Australia</td>
<td>Australia</td>
</tr>
<tr>
<td>Gurunj</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>Puntupi-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>Marathi</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; Palestine, State of; Syrian Arab Republic</td>
</tr>
<tr>
<td>Language</td>
<td>Region</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Hokkaido Ainu</td>
<td>Eurasia</td>
<td>Japan</td>
<td></td>
</tr>
<tr>
<td>Bahnar</td>
<td>Austroasiatic</td>
<td>Viet Nam</td>
<td></td>
</tr>
<tr>
<td>Bodo Gadaba</td>
<td>Austroasiatic</td>
<td>India</td>
<td></td>
</tr>
<tr>
<td>Bondo</td>
<td>Austroasiatic</td>
<td>India</td>
<td></td>
</tr>
<tr>
<td>Bugan</td>
<td>Austroasiatic</td>
<td>China</td>
<td></td>
</tr>
<tr>
<td>Car Nicobarese</td>
<td>Austroasiatic</td>
<td>India</td>
<td></td>
</tr>
<tr>
<td>Central Nicobarese</td>
<td>Austroasiatic</td>
<td>India</td>
<td></td>
</tr>
<tr>
<td>Gata'</td>
<td>Austroasiatic</td>
<td>India</td>
<td></td>
</tr>
<tr>
<td>Gorum-Parenga</td>
<td>Austroasiatic</td>
<td>India</td>
<td></td>
</tr>
<tr>
<td>Ho</td>
<td>Austroasiatic</td>
<td>India</td>
<td></td>
</tr>
<tr>
<td>Juang</td>
<td>Austroasiatic</td>
<td>India</td>
<td></td>
</tr>
<tr>
<td>Kharia</td>
<td>Austroasiatic</td>
<td>India; Nepal</td>
<td></td>
</tr>
<tr>
<td>Khasi</td>
<td>Austroasiatic</td>
<td>Bangladesh; India</td>
<td></td>
</tr>
<tr>
<td>Mundari</td>
<td>Austroasiatic</td>
<td>Bangladesh; India; India; Nepal</td>
<td></td>
</tr>
<tr>
<td>Sora</td>
<td>Austroasiatic</td>
<td>India</td>
<td></td>
</tr>
<tr>
<td>Temiar</td>
<td>Austroasiatic</td>
<td>Malaysia</td>
<td></td>
</tr>
<tr>
<td>Alutor</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia; Russian Federation</td>
<td></td>
</tr>
<tr>
<td>Chukchi</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia; Russian Federation</td>
<td></td>
</tr>
<tr>
<td>Itelmen</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia; Russian Federation</td>
<td></td>
</tr>
<tr>
<td>Koryak</td>
<td>Chukotko-Kamchatkan</td>
<td>Eurasia; Russian Federation</td>
<td></td>
</tr>
<tr>
<td>Tamil</td>
<td>Dravidian</td>
<td>India; Sri Lanka</td>
<td></td>
</tr>
<tr>
<td>Telugu</td>
<td>Dravidian</td>
<td>Eurasia; India</td>
<td></td>
</tr>
<tr>
<td>Central Siberian Yupik</td>
<td>Eskimo-Aleut</td>
<td>Eurasia; Russian Federation; United States</td>
<td></td>
</tr>
<tr>
<td>Kalaallisut</td>
<td>Eskimo-Aleut</td>
<td>Greenland</td>
<td></td>
</tr>
<tr>
<td>Catalan</td>
<td>Indo-European</td>
<td>Andorra; France; Italy; Spain</td>
<td></td>
</tr>
<tr>
<td>Danish</td>
<td>Indo-European</td>
<td>Denmark; Germany</td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>Indo-European</td>
<td>Belgium; Brazil; French Guiana; Germany; Guyana; Netherlands; Suriname</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>Indo-European</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>
<td></td>
</tr>
<tr>
<td>French</td>
<td>Indo-European</td>
<td>Andorra; Belgium; Canada; France; French Guiana; Italy; Luxembourg; Monaco; Saint Pierre and Miquelon; Switzerland; United States</td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>Indo-European</td>
<td>Austria; Belgium; Czechia; Denmark; France; Germany; Hungary; Italy; Liechtenstein; Luxembourg;</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>Language Group</td>
<td>Region</td>
<td>Countries and Regions</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>---------------------------------------------------------------------------------------</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>Europe</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</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>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>Europe</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>Russia</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-Tibetian</td>
<td>Eurasia</td>
<td>China</td>
</tr>
<tr>
<td>Lahu</td>
<td>Sino-Tibetian</td>
<td>Eurasia</td>
<td>China; Lao People's Democratic Republic; Myanmar; Thailand; Viet Nam</td>
</tr>
<tr>
<td>Mandarin Chinese</td>
<td>Sino-Tibetian</td>
<td>Eurasia</td>
<td>China; Korea; Democratic People's Republic of; Lao People's Democratic Republic; Mongolia; Myanmar; Russian Federation; Taiwan, Province of China; Viet Nam</td>
</tr>
<tr>
<td>Tangut</td>
<td>Sino-Tibetian</td>
<td>Eurasia</td>
<td>China</td>
</tr>
<tr>
<td>Tibetan</td>
<td>Sino-Tibetian</td>
<td>Eurasia</td>
<td>Bhutan; China; India; Nepal</td>
</tr>
<tr>
<td>Udihe</td>
<td>Tungusic</td>
<td>Eurasia</td>
<td>Russia</td>
</tr>
<tr>
<td>Language</td>
<td>Family</td>
<td>Region</td>
<td>Countries</td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Turkish</td>
<td>Turkic</td>
<td>Eurasia</td>
<td>Albania; Armenia; Azerbaijan; Bulgaria; Cyprus; Georgia; Greece; Iran, Islamic Republic of; Iraq; Macedonia, Republic of; Syrian Arab 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>Canada; United States</td>
</tr>
<tr>
<td>Severn Ojibwa</td>
<td>Algic</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Shawnee</td>
<td>Algic</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Sakska</td>
<td>Algic</td>
<td>North America</td>
<td>Canada; United States</td>
</tr>
<tr>
<td>Ahêna</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>
<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>Biríri</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>Quechua</td>
<td>Cochina-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 Inuktitut</td>
<td>Eskimo-Aleut</td>
<td>North America</td>
<td>Canada</td>
</tr>
<tr>
<td>Haida²</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>Language</td>
<td>Type</td>
<td>Region</td>
<td>Countries</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>-----------</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</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>Mixteco</td>
<td>Misumalpan</td>
<td>North America</td>
<td>Honduras; Nicaragua</td>
</tr>
<tr>
<td>Chamalapa 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</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Chucahuaxtal</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Coatec1,2</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>
<tr>
<td>Mixtulete3,4</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>San Jerónimo</td>
<td>Otomanguean</td>
<td>North America</td>
<td>Mexico</td>
</tr>
<tr>
<td>Tecóatl Maratec</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>Palaihmihan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Atrigewi1</td>
<td>Palaihmihan</td>
<td>North America</td>
<td>United States</td>
</tr>
<tr>
<td>Central Pomo</td>
<td>Pomnoan</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;</td>
</tr>
<tr>
<td>Language</td>
<td>Family</td>
<td>Region</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------</td>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>Assiniboine</td>
<td>Siouan</td>
<td>North America</td>
<td></td>
</tr>
<tr>
<td>Biloxi</td>
<td>Siouan</td>
<td>North America</td>
<td></td>
</tr>
<tr>
<td>Crow</td>
<td>Siouan</td>
<td>North America</td>
<td></td>
</tr>
<tr>
<td>Hidatsa</td>
<td>Siouan</td>
<td>North America</td>
<td></td>
</tr>
<tr>
<td>Ho-Chunk</td>
<td>Siouan</td>
<td>North America</td>
<td></td>
</tr>
<tr>
<td>Lakota</td>
<td>Siouan</td>
<td>North America, Canada</td>
<td></td>
</tr>
<tr>
<td>Mandan</td>
<td>Siouan</td>
<td>North America</td>
<td></td>
</tr>
<tr>
<td>Takelma</td>
<td>Takelma (Isolate)</td>
<td>North America</td>
<td></td>
</tr>
<tr>
<td>Purepecha</td>
<td>Tarascan</td>
<td>North America, Mexico</td>
<td></td>
</tr>
<tr>
<td>Tonkawa</td>
<td>Tonkawa (Isolate)</td>
<td>North America</td>
<td></td>
</tr>
<tr>
<td>Upper Necaxa</td>
<td>Tononac</td>
<td>North America</td>
<td></td>
</tr>
<tr>
<td>Gitsxan</td>
<td>Tsimshian</td>
<td>North America, Canada</td>
<td></td>
</tr>
<tr>
<td>Nisga’a</td>
<td>Tsimshian</td>
<td>North America, Canada</td>
<td></td>
</tr>
<tr>
<td>Southern-Coastal</td>
<td>Tsimshian</td>
<td>North America, Canada</td>
<td></td>
</tr>
<tr>
<td>Classical Nahuatl</td>
<td>Uto-Aztecan</td>
<td>North America, Mexico</td>
<td></td>
</tr>
<tr>
<td>Comanche</td>
<td>Uto-Aztecan</td>
<td>North America, United States</td>
<td></td>
</tr>
<tr>
<td>Eastern Huasteca</td>
<td>Nahuatl</td>
<td>North America, Mexico</td>
<td></td>
</tr>
<tr>
<td>Hopi</td>
<td>Uto-Aztecan</td>
<td>North America, United States</td>
<td></td>
</tr>
<tr>
<td>Huchol</td>
<td>Uto-Aztecan</td>
<td>North America, Mexico</td>
<td></td>
</tr>
<tr>
<td>Panamint</td>
<td>Uto-Aztecan</td>
<td>North America, United States</td>
<td></td>
</tr>
<tr>
<td>Puma Bajo</td>
<td>Uto-Aztecan</td>
<td>North America, Mexico</td>
<td></td>
</tr>
<tr>
<td>Pipil</td>
<td>Uto-Aztecan</td>
<td>North America, El Salvador</td>
<td></td>
</tr>
<tr>
<td>Shoshoni</td>
<td>Uto-Aztecan</td>
<td>North America, United States</td>
<td></td>
</tr>
<tr>
<td>Tarahumaran²</td>
<td>Uto-Aztecan</td>
<td>North America, Mexico</td>
<td></td>
</tr>
<tr>
<td>Tlaxcala-Puebla-Central</td>
<td>Uto-Aztecan</td>
<td>North America, Mexico</td>
<td></td>
</tr>
<tr>
<td>Ute-Southern</td>
<td>Paute</td>
<td>Uto-Aztecan, North America, United States</td>
<td></td>
</tr>
<tr>
<td>Yaqui</td>
<td>Uto-Aztecan</td>
<td>North America, Mexico, United States</td>
<td></td>
</tr>
<tr>
<td>Kwak’wala</td>
<td>Wakashan</td>
<td>North America, Canada, United States</td>
<td></td>
</tr>
<tr>
<td>Nuu-chab-nalh</td>
<td>Wakashan</td>
<td>North America, United States</td>
<td></td>
</tr>
<tr>
<td>Washo</td>
<td>Washo (Isolate)</td>
<td>North America, United States</td>
<td></td>
</tr>
<tr>
<td>Yana</td>
<td>Yana (Isolate)</td>
<td>North America, United States</td>
<td></td>
</tr>
<tr>
<td>Zuni</td>
<td>Zuni (Isolate)</td>
<td>North America, United States</td>
<td></td>
</tr>
<tr>
<td>Aneityum</td>
<td>Austronesian</td>
<td>Papuasia, Vanuatu</td>
<td></td>
</tr>
<tr>
<td>Chamorro</td>
<td>Austronesian</td>
<td>Papuasia, Guam, Northern Mariana Islands</td>
<td></td>
</tr>
<tr>
<td>Chuukese</td>
<td>Austronesian</td>
<td>Papuasia, Micronesia, Federated States of</td>
<td></td>
</tr>
<tr>
<td>Dehu</td>
<td>Austronesian</td>
<td>Papuasia, New Caledonia</td>
<td></td>
</tr>
<tr>
<td>East Futuna</td>
<td>Austronesian</td>
<td>Papuasia, Wallis and Futuna</td>
<td></td>
</tr>
<tr>
<td>Fijian</td>
<td>Austronesian</td>
<td>Papuasia, Fiji</td>
<td></td>
</tr>
<tr>
<td>Iaai</td>
<td>Austronesian</td>
<td>Papuasia, New Caledonia</td>
<td></td>
</tr>
<tr>
<td>Indonesian</td>
<td>Austronesian</td>
<td>Papuasia, Indonesia, Netherlands, Philippines, Saudi Arabia, Singapore, United States</td>
<td></td>
</tr>
<tr>
<td>Kambara</td>
<td>Austronesian</td>
<td>Papuasia, Indonesia</td>
<td></td>
</tr>
<tr>
<td>Kosraean</td>
<td>Austronesian</td>
<td>Papuasia, Micronesia, Federated States of</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>Family</td>
<td>Region</td>
<td>Country/Region</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------</td>
<td>-------------------------------</td>
<td>----------------------------------------</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>Marshallense</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>Nuean</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Niue</td>
</tr>
<tr>
<td>Paawan</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>
<tr>
<td>Pohnpeian</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Micronesia, Federated States of</td>
</tr>
<tr>
<td>Sakai</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'abaiata</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 Besi²</td>
<td>Austronesian</td>
<td>Papunesia</td>
<td>Indonesia</td>
</tr>
<tr>
<td>Warendbori</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>Amelé</td>
<td>Nuclear Trans New Guinea</td>
<td>Papunesia</td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td>Alamblak</td>
<td>Sepik</td>
<td>Papunesia</td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td>Kalamang³</td>
<td>West Bomberai</td>
<td>Papunesia</td>
<td>Indonesia</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>Baure¹</td>
<td>Arawakan</td>
<td>South America</td>
<td>Bolivia, Plurinational State of</td>
</tr>
<tr>
<td>Palikúr</td>
<td>Arawakan</td>
<td>South America</td>
<td>Brazil; French Guiana</td>
</tr>
<tr>
<td>Wayuu</td>
<td>Arawakan</td>
<td>South America</td>
<td>Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>Paumari</td>
<td>Arawan</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Galibi Carib</td>
<td>Cariban</td>
<td>South America</td>
<td>Brazil; French Guiana; Guyana;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Suriname; Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>Kukúro-Kalapalo¹</td>
<td>Cariban</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Panare</td>
<td>Cariban</td>
<td>South America</td>
<td>Venezuela, Bolivarian Republic of</td>
</tr>
<tr>
<td>Wari¹</td>
<td>Chapacuran</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Malayo</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>Murui Huitoto</td>
<td>Huitotoan</td>
<td>South America</td>
<td>Colombia; Peru</td>
</tr>
<tr>
<td>Katukina-Kanamari</td>
<td>Katukinan</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Wichí²</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>Dāw</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>
<tr>
<td>Southern Nambikuára</td>
<td>Nambiquaran</td>
<td>South America</td>
<td>Brazil</td>
</tr>
<tr>
<td>Language</td>
<td>Macro-Geographic Family</td>
<td>Region</td>
<td>Country/Region</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------</td>
<td>--------------</td>
<td>-----------------------------------------</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 Éjja</td>
<td>Pano-Tucanoan</td>
<td>South America</td>
<td>Bolivia, Plurinational State of, Peru</td>
</tr>
<tr>
<td>Ticuna</td>
<td>Ticuna-Juri</td>
<td>South America</td>
<td>Brazil, Colombia, Peru</td>
</tr>
<tr>
<td>Tanimuca-Retuará</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>Guarayu</td>
<td>Tupian</td>
<td>South America</td>
<td>Bolivia, Plurinational State of</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>Tapajé</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>
<tr>
<td>Yanomámi</td>
<td>Yanomamic</td>
<td>South America</td>
<td>Brazil; Venezuela, Bolivarian Republic of</td>
</tr>
</tbody>
</table>
Appendix 2 Proportions of language families

This table presents the proportion of language families from each macro-area in the list of incorporating languages, included in Appendix 1, and the corresponding number of languages from each macro-area in the 30-language sample used in Chapter 3 and 4. The table is based only on the 248 incorporating languages that were identified as described in Section 3.4.1 and 4.3.1, i.e. the languages marked with “+” in the list in Appendix 1 are not taken into account. 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>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>
<tr>
<td>Total</td>
<td>84</td>
<td>1</td>
<td>30</td>
</tr>
</tbody>
</table>
Appendix 3 Sources consulted for the sample languages in Chapter 3

**Bininj Kun-Wok**


**Chimalapa Zoque**

**Chukchi**

202  Incorporation: Constraints on variation


Crow


Appendix 3  203


**Eastern Ojibwa**


Halkomelem


Hokkaido Ainu


Appendix 3  205


**Iraqw**


**Kalaallisut**


**Ket**


Mapudungun
Zúñiga, Fernando. 2015. Valency classes in Mapudungun. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world’s languages*, vol. 2: *Case studies from Austronesia and the Pacific, the Americas, and theoretical outlook* (Comparative Handbooks of Linguistics 1(2)), 1515–1543. Berlin: De Gruyter Mouton.

Marithiel

Mohawk
Appendix 3  209


**Movima**


Incorporation: Constraints on variation


Nadèb


Niuean


Appendix 3


Northern Gumuz

Nuu-chah-nulth
212 Incorporation: Constraints on variation

Palikúr

Panare

Paraguayan Guaraní
Appendix 3


**Sora**


**South Slavey**


Incorporation: Constraints on variation


**Southern Tiwa**


**Ute-Southern Paiute**


**Washo**


Incorporation: Constraints on variation


Western Frisian


Western Highland Chatino


Yimas


Appendix 3  217


Yucatec Maya


Appendix 4 Sources consulted for the sample languages in Chapter 4

Bininj Kun-Wok

Hokkaido Ainu

Iraqw
Incorporation: Constraints on variation


**Kalaallisut**


**Ket**


**Mapudungun**


**Mohawk**


Incorporation: Constraints on variation


Movima

Nadēb

Niuean

Northern Gumuz

Nuu-chah-nulth


**Palikûr**


**Panare**


**Paraguayan Guarani**


**Sora**


**Ute-Southern Paiute**


**Washo**


**Western Frisian**


**Yimas**


Yucatec Maya
**Appendix 5 Language sample and sources consulted for the typological survey in Chapter 5**

This table includes the names, glottocodes and family classifications of the languages included in the typological survey from Chapter 5 as presented in Glottolog (Hammarström et al. 2017), as well as the data sources used for each of the sample languages. 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>Glottocode</th>
<th>Family</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alamblak</td>
<td>alam1246</td>
<td>Sepik</td>
<td>Bruce (1984: 170–176, 370)</td>
</tr>
<tr>
<td>Atsujewi</td>
<td>atsu1245</td>
<td>Palaihnihan</td>
<td>Palancar (1999: 155, 162)</td>
</tr>
<tr>
<td>Caddo</td>
<td>cadd1256</td>
<td>Caddoan</td>
<td>Melnar (1998)</td>
</tr>
<tr>
<td>Haida1</td>
<td>haid1248</td>
<td>Haida</td>
<td>Enrico (2003: 1263–1267)</td>
</tr>
<tr>
<td>Malay [Damana]</td>
<td>mala1522</td>
<td>Chibchan</td>
<td>Quesada (1999: 248)</td>
</tr>
<tr>
<td>Mandinka</td>
<td>mand1436</td>
<td>Mande</td>
<td>Creissels (2008: 82–87)</td>
</tr>
</tbody>
</table>

1 Haida is considered a language family rather than a single language in Glottolog (Hammarström et al. 2017).
<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Language Family</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nadeb</td>
<td>nade1244</td>
<td>Nadaup</td>
<td>Warr (1990: 324–327)</td>
</tr>
<tr>
<td>Southern Tiwa</td>
<td>sout2961</td>
<td>Kiowa-Tanoan</td>
<td>Allen et al. (1984)</td>
</tr>
<tr>
<td>Takelma</td>
<td>take1257</td>
<td>Takelma (Isolate)</td>
<td>Palancar (1999: 160)</td>
</tr>
<tr>
<td>Tiwi</td>
<td>tiwi1244</td>
<td>Tiwi (Isolate)</td>
<td>Osborne (1974: 47–50)</td>
</tr>
<tr>
<td>Warembori</td>
<td>ware1253</td>
<td>Austronesian</td>
<td>Donohue (1999: 43–46)</td>
</tr>
<tr>
<td>Washo</td>
<td>wash1253</td>
<td>Washo (Isolate)</td>
<td>Bochnak and Rhomieux (2013)</td>
</tr>
<tr>
<td>Zuni</td>
<td>zu1i1245</td>
<td>Zuni (Isolate)</td>
<td>Miner (1986: 251–253)</td>
</tr>
</tbody>
</table>

**Full references to the sources**


Incorporation: Constraints on variation


Incorporation: Constraints on variation


Summary

Incorporation: Constraints on variation

The aim of this thesis is to add to our understanding of the cross- and intra-linguistic variation in incorporation constructions by examining the range of variation regarding their pragmatic, semantic, morphosyntactic, phonological and lexical properties and by focusing specifically on constraints on this variation. After a general introduction in Chapter 1, the pragmatic, semantic and phonological variation in incorporation and constraints on this variation identified in previous literature are reviewed in Chapter 2. Three in-depth studies of particular domains of morphosyntactic, pragmatic-semantic and lexical variation follow in Chapter 3, 4 and 5, respectively. These domains of variation are analyzed in detail, based on data from a varied sample of languages. The variation and constraints investigated in Chapter 2 to 5 are subsequently discussed in light of several theoretical approaches to incorporation in Chapter 6, in order to examine their implications for theoretical accounts of incorporation.

A broad review of the pragmatic, semantic and phonological variation in noun incorporation constructions attested in previous literature is presented in Chapter 2. For each of the pragmatic, semantic and phonological domains discussed, it is examined whether the variation is constrained by a basic setting that specifies a limited number of options from which each individual language makes a choice or whether the variation shows a cross-linguistic pattern in the form of an implicational hierarchy. Both types of constraints appear to be relevant for the variation in noun incorporation. For a few domains of variation, including the phonological domains addressed, only basic settings can be identified. For most pragmatic and semantic domains of variation investigated, by contrast, the available properties form implicational hierarchies, in that a language can only show a particular property if it also, in other contexts, shows another. These implicational relationships indicate that there are systematic patterns in the pragmatic and semantic variation in noun incorporation across languages.

In Chapter 3, the morphosyntactic forms of incorporated elements are investigated. This chapter includes incorporated nouns as well as incorporated verbs, adjectives, adverbs and adpositions. The morphosyntactic possibilities for incorporated elements are examined on the basis of a systematic analysis of data from a sample of 30 incorporating languages. Both single lexical morphemes, i.e. unbound stems and bound roots, derived stems, inflected words and phrases are shown to be incorporated across languages. Moreover, it is found that these forms constitute an implicational hierarchy in which they are ordered from simple to more complex, i.e. more complex forms can only be incorporated in languages that also show the
incorporation of simpler forms. It is thus demonstrated that there is a large variation in the forms of incorporated elements, while this variation is at the same time constrained in a systematic way.

The focus of Chapter 4 is on the pragmatic referentiality and semantic modifiability of incorporated nouns. These domains have already attracted considerable attention in the incorporation literature, but this chapter studies the variation with respect to the referentiality and modifiability of incorporated nouns between and especially within languages in a balanced sample of languages. It is shown that incorporated nouns may be referentially used modifiable nouns (+R/+M nouns), non-referentially used modifiable nouns (−R/+M nouns) and non-referentially used non-modifiable nouns (−R/−M nouns). Regarding the incorporation of +R/+M nouns and −R/−M nouns, it is found that some languages display one of the two while others allow both. Whether or not noun incorporation is restricted to either of the two types is a language-specific issue. By contrast, the incorporation of −R/+M nouns is demonstrated to be dependent on the incorporation of +R/+M nouns, i.e. the occurrence of −R/+M incorporated nouns in a particular language implies the occurrence of +R/+M incorporated nouns in the same language. In this way, the variation regarding the pragmatic referentiality and semantic modifiability of incorporated nouns is also subject to a constraint in the form of a cross-linguistic implicational relationship.

In Chapter 5, the question which types of verbs do and do not allow the incorporation of nouns is addressed. The chapter consists of a typological survey of incorporating verbs in descriptive sources for a sample of 50 incorporating languages and a more detailed set of case studies on incorporating verbs in corpus data from eight languages. Based on the typological survey and case studies, the degree of morphosyntactic transitivity of a verb is found to play a role: morphosyntactically highly transitive verbs are most likely to show noun incorporation, and patientive intransitive verbs often allow the incorporation of nouns as well. However, the relation between a verb’s degree of morphosyntactic transitivity and its ability to incorporate nouns seems to reflect a tendency rather than a strict constraint in the form of an implicational pattern. Moreover, whether or not a particular verb allows noun incorporation in a language to a large extent seems to be an idiosyncratic issue: lexical properties of verbs often appear to be relevant, and the possibility or impossibility to use a particular verb in noun incorporation thus seems to be partly determined by lexical constraints.

In Chapter 6, it is examined to what extent the different theoretical accounts of incorporation proposed in the literature are able to capture or predict the variation and constraints presented in Chapter 2 to 5. In this way, the explanatory potential of these theoretical accounts can be assessed. The chapter follows the general distinction made in the literature on incorporation between lexical and syntactic accounts of incorporation. The syntactic approach to incorporation taken in the thesis, which is
based on Functional Discourse Grammar, generally seems to be consistent with the data concerning each of the domains of incorporation investigated. By contrast, the other lexical and syntactic accounts discussed cannot fully explain all variation and constraints attested in the thesis. Importantly, some domains of variation and their constraints pose problems for the lexical and syntactic approaches to different degrees. At least part of the variation in the forms of incorporated elements found in the thesis can be captured by some syntactic approaches, while none of the lexical accounts predicts any variation in this domain. For this reason, the syntactic approaches to incorporation seem better suited to account for the findings of the present thesis than the lexical ones.

All in all, the four studies included in this thesis show that the variation between incorporation constructions with respect to their pragmatic, semantic, morphosyntactic, phonological and lexical properties is pervasive. At the same time, constraints can be formulated regarding all domains investigated. While the Functional Discourse Grammar approach taken in the thesis can generally explain the variation and constraints identified, some domains of variation in incorporation and some of their constraints pose challenges to the traditional lexical and syntactic approaches to incorporation. The thesis thus indicates that theoretical work on incorporation may generally benefit from paying more attention to the full range of variation in incorporation constructions and the constraints that govern this variation.
Samenvatting (summary in Dutch)

Incorporatie: Beperkingen op variatie

Het doel van dit proefschrift is om bij te dragen aan onze kennis over de cross- en intra-linguïstische variatie in incorporatiestructuren. Hiervoor bestudeer ik hoeveel variatie hun pragmatische, semantische, morfosyntactische, fonologische en lexicale eigenschappen vertonen en neem ik in het bijzonder de beperkingen op deze variatie onder de loep. Na een algemene introductie in hoofdstuk 1 behandelt hoofdstuk 2 de pragmatische, semantische en fonologische variatie in incorporatie, en de beperkingen hierop, die al uit eerder onderzoek zijn gebleken. Drie uitvoeriger studies van specifieke morfosyntactische, pragmatisch-semantische en lexicale variatiedomeinen volgen respectievelijk in hoofdstuk 3, 4 en 5. Deze domeinen worden in detail geanalyseerd op basis van data van een gevarieerd talensample. De variatie en beperkingen die ik onderzoek in hoofdstuk 2 tot 5 bespreek ik vervolgens in hoofdstuk 6 in het licht van verschillende theoretische benaderingen van incorporatie om zo hun implicaties voor incorporatietheorieën te onderzoeken.

De pragmatische, semantische en fonologische variatie in de incorporatie van zelfstandige naamwoorden die in eerdere literatuur is geïdentificeerd vormt het onderwerp van hoofdstuk 2. Voor elk pragmatische, semantische en fonologische domein dat in dit hoofdstuk aan bod komt, wordt ook bekeken hoe de variatie in het domein beperkt wordt: lijkt het domein onderworpen aan een instelling die simpelweg een aantal opties specificereert waar elke taal een keuze uit maakt of vertoont de variatie een cross-linguïstisch patroon in de vorm van een implicationele hiërarchie? Beide typen beperkingen blijken relevant voor de variatie in de incorporatie van zelfstandige naamwoorden. Voor een paar van de behandelde variatiedomeinen, waaronder de fonologische domeinen, kunnen alleen eenvoudige instellingen worden herkend. Voor de meeste van de pragmatische en semantische variatiedomeinen die aan bod komen, vormen de mogelijke eigenschappen daarentegen implicationele hiërarchieën: een taal kan alleen een bepaalde eigenschap vertonen als hij ook, in andere contexten, een bepaalde andere eigenschap laat zien. Deze implicationele relaties geven aan dat er systematische patronen zijn in de pragmatische en semantische variatie in de incorporatie van zelfstandige naamwoorden over talen heen.

In hoofdstuk 3 onderzoek ik de morfosyntactische vormen van geïncorporeerde elementen. Dit hoofdstuk richt zich niet alleen op geïncorporeerde zelfstandige naamwoorden maar ook op geïncorporeerde werkwoorden, bijvoeglijke naamwoorden, bijwoorden en adposities. Ik bestudeer de morfosyntactische mogelijkheden voor geïncorporeerde elementen op basis van een systematische analyse van data van een sample van 30 incorporerende talen. De data laten zien dat zowel lexicale morfemen, d.w.z. ongebonden stammen en gebonden wortels, als
afgeleide stammen, verbogen en vervoegde woorden en frases geïncorporeerd kunnen worden, in verschillende talen. Bovendien toont de analyse aan dat deze vormen deel uitmaken van een implicationele hiërarchie waarin ze geordend zijn van simpel naar complex: complexere vormen kunnen alleen geïncorporeerd worden in talen die ook de incorporatie van simpeler vormen toestaan. Op deze manier blijkt dat er grote variatie bestaat in de vormen van geïncorporeerde elementen, terwijl deze variatie tegelijkertijd op een systematische manier beperkt is.

De focus van hoofdstuk 4 ligt op de pragmatische referentialiteit en semantische modificeerbaarheid van geïncorporeerde zelfstandige naamwoorden. Deze domeinen hebben al veel aandacht gekregen in de incorporatieliteratuur, maar dit hoofdstuk bestudeert de variatie in de referentialiteit en modificeerbaarheid van geïncorporeerde zelfstandige naamwoorden tussen en vooral ook binnen talen in een gebalanceerd talensample. De data laten zien dat zowel referentieel gebruikte modificeerbare zelfstandige naamwoorden (+R/+M zelfstandige naamwoorden), niet-referentieel gebruikte modificeerbare zelfstandige naamwoorden (−R/+M zelfstandige naamwoorden) als niet-referentieel gebruikte niet-modificeerbare zelfstandige naamwoorden (−R/−M) kunnen worden geïncorporeerd. Wat betreft de incorporatie van +R/+M zelfstandige naamwoorden en −R/−M zelfstandige naamwoorden blijkt uit de data dat sommige talen slechts een van de twee laten zien terwijl andere beide toestaan. Of de incorporatie van zelfstandige naamwoorden beperkt is tot een van de twee is een taalspecifieke zaak. Tegelijkertijd blijkt de incorporatie van −R/+M zelfstandige naamwoorden afhankelijk te zijn van de incorporatie van +R/+M geïncorporeerde zelfstandige naamwoorden, d.w.z. het voorkomen van −R/+M geïncorporeerde zelfstandige naamwoorden impliceert het voorkomen van +R/+M geïncorporeerde zelfstandige naamwoorden in dezelfde taal. De variatie in de pragmatische referentialiteit en semantische modificeerbaarheid van geïncorporeerde zelfstandige naamwoorden is zo dus ook onderhevig aan een beperking in de vorm van een cross-linguïstische implicationele relatie.

Hoofdstuk 5 gaat over de vraag welke typen werkwoorden wel en niet de incorporatie van zelfstandige naamwoorden toestaan. Dit hoofdstuk bestaat uit een typologisch overzicht van incorporerende werkwoorden in descriptieve bronnen voor een sample van 50 incorporerende talen en een set meer gedetailleerde casestudies van incorporerende werkwoorden in corpusdata van acht talen. Op basis van het typologische overzicht en de casestudies blijkt dat de mate van morfosyntactische transitiviteit van een werkwoord een rol speelt: morfosyntactisch hoog transitieve werkwoorden zijn de meest waarschijnlijke kandidaten voor de incorporatie van zelfstandige naamwoorden en intransitieve werkwoorden met een patiens als argument staan de incorporatie van zelfstandige naamwoorden ook vaak toe. De relatie tussen de mate van morfosyntactische transitiviteit van een werkwoord en zijn geschiktheid voor incorporatie representeert echter eerder een tendens dan een strikte beperking in de vorm van een implicationeel patroon. Of een bepaald werkwoord wel
Samenvatting (summary in Dutch)

of geen incorporatie van zelfstandige naamwoorden toestaat lijkt bovendien in belangrijke mate een idiosyncratische zaak: lexicale eigenschappen van werkwoorden zijn medebepalend en de mogelijkheid of onmogelijkheid om een bepaald werkwoord te gebruiken in zelfstandig-naamwoordincorporatie lijkt dus deels bepaald te zijn door lexicale beperkingen.

In hoofdstuk 6 bestudeer ik in welke mate de verschillende theoretische benaderingen van incorporatie uit de literatuur in staat zijn om de variatie en beperkingen uit hoofdstuk 2 tot 5 te beschrijven of voorspellen. Op deze manier kan het verklarend potentieel van deze theoretische benaderingen beoordeeld worden. Het hoofdstuk volgt het onderscheid dat gewoonlijk in de literatuur over incorporatie wordt gemaakt tussen lexicale en syntactische benaderingen van incorporatie. Dit proefschrift stelt zelf een syntactische benadering voor die is gebaseerd op Functional Discourse Grammar en deze benadering blijkt over het algemeen consistent te zijn met de data voor elk van de onderzochte variatiedomeinen van incorporatie. De andere lexicale en syntactische benaderingen lijken daarentegen niet in staat alle variatie en beperkingen die in het proefschrift geïdentificeerd zijn volledig te verklaren. Belangrijk hierbij is dat de variatiedomeinen en hun beperkingen op verschillende manieren en in verschillende mate problematisch zijn voor de lexicale en syntactische benaderingen. Zo kan in ieder geval een deel van de in dit proefschrift gevonden variatie in de vormen van geïncorporeerde elementen door sommige syntactische benaderingen beschreven worden, terwijl geen van de lexicale benaderingen enige variatie voorspelt in dit domein. De syntactische benaderingen van incorporatie lijken daarom geschikter om de bevindingen van dit proefschrift te verklaren dan de lexicale benaderingen.

Alles bij elkaar genomen laten de vier studies die in dit proefschrift zijn opgenomen zien dat de variatie tussen incorporatieconstructies met betrekking tot hun pragmatische, semantische, morfosyntactische, fonologische en lexicale eigenschappen wijdverbreid is. Tegelijkertijd kunnen er beperkingen worden geformuleerd op de variatie in alle onderzochte domeinen. Terwijl de Functional Discourse Grammar-benadering van dit proefschrift de geïdentificeerde variatie en beperkingen over het algemeen goed kan verklaren, stellen sommige domeinen van variatie in incorporatie en sommige van hun beperkingen de traditionele lexicale en syntactische benaderingen voor uitdagingen. Het proefschrift laat dus zien dat theoretische werk op het gebied van incorporatie er over het algemeen baat bij zou hebben om meer aandacht te schenken aan de totale variatie in constructies met incorporatie en de beperkingen die deze variatie bepalen.
Curriculum Vitae

Marieke Olthof was born on May 27, 1991 in Zaandam. After graduating from the Zaanlands Lyceum in 2009, she started her studies at the University of Amsterdam. She completed a BA in Scandinavian languages & cultures in 2012 and then took a BA in linguistics, which she finished in 2014. Subsequently, she was a student in the research MA in linguistics and graduated from this program in 2016. She then started her PhD project at the same university, which was funded by the Dutch Research Council (NWO) in the context of the research program PhDs in the Humanities. The project resulted in this thesis. During her PhD, Marieke presented her work at several conferences and participated in various summer and winter schools. In addition, she was a member of the Faculty of Humanities’ PhD council and taught courses in the BA program in Scandinavian languages & cultures and in the BA program in linguistics.