PREDICATE INITIALITY IN SPACE AND TIME

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

LINGUISTICS

May 2022

By

Kevin Baetscher

Dissertation committee:
Lyle Campbell, Co-Chairperson
Gary Holton, Co-Chairperson
Shelece Easterday
Alexander Mawyer
William O’Grady

Keywords: typology, syntax, constituent order
Abstract

This dissertation consolidates data from available areal studies and hundreds of grammars on languages from across the globe to compile the most comprehensive atlas of predicate initiality in the world’s languages. It approaches predicate initiality (P1) from different perspectives: from a functional-theoretical perspective, from an areal-typological perspective, and from a historical perspective.

This study identifies 811 languages that allow the placement of the predicate before the two core arguments of a transitive clause in pragmatically neutral statements, which represents around 11% of the world’s languages. The resulting typological picture is that P1 is not only relatively rare, but its global distribution is highly uneven. 88% of P1 languages are found in one of the five major P1 hotspots: on Pacific Islands (including Taiwan and the Philippines), in the Pacific Northwest coast of America, Mesoamerica, the Maghreb, or the East African Rift. In these areas, P1 order is the norm, either within a single language family (e.g. Austronesian in Pacific Islands), or across various neighboring language families in a linguistic area (e.g. in Mesoamerica). By contrast, P1 is virtually absent from all of Eurasia and Australia, and the linguistically very diverse regions of West Africa and New Guinea.

P1 can be traced back to the earliest reconstructible stages for many P1 languages. Adding just the P1 languages from the six language families Austronesian, Otomanguean, Afroasiatic, Mayan, Salish, and Nilotic, for which the reconstruction of P1 is generally accepted, accounts for 80% of P1 languages. The Austronesian family alone accounts for almost half (346, or 43%) of the world’s P1 languages. Independent emergence of P1 is theoretically possible, but practically highly unlikely. Perhaps the only case where the historical development of P1 can be reconstructed in detail—thanks to a long literary tradition—are the Celtic languages and Modern Greek.

Put simply, the reason for the rare independent emergence of P1 is that the common syntactic processes that alter constituent order lead away from P1: fronting operations (topicalization, focalization) typically target participants rather than actions, because the former are more salient and continuous in narratives. Hence, there are only few syntactic pathways for P1 to develop language-internally across time. P1 is not a diachronic target but rather tends toward a subject-predicate-object order over time. As a result, P1 almost exclusively spreads through diachronic diversification of languages and language contact. This is the reason why P1 forms such distinct areal clusters, with relatively few typological outliers, compared to predicate-initial or predicate-final constituent orders.
To my late early mentors
Al & Bob
Acknowledgements

ʻO ka mua, e mahalo aʻe au i ke Akua, nāna nō i hoʻoulu mai iaʻu e ʻimi noʻiʻi i ka moʻokūʻauhau o nā ʻōlelo i ahu mai ai ke kumu o kō lākou mau hiʻoi‘ona i kēia au.

I raise my hands in gratitude to each of my committee members for the guidance they have given me in the composition of my dissertation. A special thanks is due to Lyle Campbell, as this dissertation started as a term paper in his typology class a few years back and was certainly inspired even earlier than that by his typology class at the Summer Institute of Linguistics in Ann Harbour in 2013, before I joined UH. In fact, Lyle played a big role in luring me to Hawaiʻi. Lyle, I’ve profited tremendously from your classes, conversations, and support. Kiitoksia paljon for your many layers of feedback to various drafts of this thesis. AwA’ahdah to Gary Holton, who paved my way for many technical aspects of this dissertation, ranging from data management to linguistic mapping and text formatting. Gary also drilled me to sharpen my definitions and terminology, for which I also thank William O’Grady, who generally had much influence in my understanding of, and approach to, syntax. Many thanks are due also to my other committee members, Shelece Easterday with her typologist frame of mind, and Alex Mawyer, who kept reminding me of the bigger questions a second-language speaker of P1 languages would raise; both of you joined my committee very spontaneously when vacancies arose, and I thank you for that. Māuruuru roa!

I also express my deepest gratitude to Robert Blust, who gave me valuable input during the early stages of my dissertation but unfortunately did not live to see its completion. By far my deepest insight and understanding of any P1 languages at this point of my life relates to the Austronesian language family, and a large part of this knowledge either springs directly from, or was in some way inspired by, the countless hours I spent in Bob’s classes or in his office. Terima kasih, e moe me ka maluhia.

I consider myself extremely lucky to have been able to pursue my doctorate studies at the University of Hawaiʻi, which enabled me to be guided by such a high-profile committee who contributed immensely to the quality of this dissertation. Any remaining errors and inconsistencies are entirely my own fault. I am in debt to the entire University of Hawaiʻi, and especially the Linguistics Department of Mānoa, with its very engaged body of professors and students. Mahalo ā nui also to Jen Kanda and John Kawahara for their continued support with administrative issues.

I must also thank Marine Vuillermet for speeding up the progress of my data collection by sharing her online access to linguistic publications.
This work is not only the product of academic endeavour, but it reflects many interactions I have had with native speakers and second-language learners of P1 speakers. I thank each and every one of you for sharing your thoughts and questions, and I hope that this dissertation answers some of them. My highest respect to all the language warriors who promote their ancestral language by learning and teaching against all odds!

In particular, I am indebted to the Hawaiian community for sharing their language, culture, and land with me. Not only did you inspire this work, but you provided an important counter-balance to my intellectual endeavour, thus keeping me grounded on your beautiful hōnua. Without providing a exhaustive list, I am obliged to mention: ku‘u mau hoa lumi ‘o Al Schütz me Keao NeSmith, ‘o ke kumu Michael Pili Pang me ka Hālau Hula Ka No‘eau, me ku‘u hoa aloha ‘o Kamealoha Forrest. Mahalo a nui loa iā ‘oukou āpau!

Finally, I thank all my friends and family in Switzerland who bore with me and spurred me on to keep my eyes on the horizon, while I doubled down on data collection and writing during the pandemic.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS</td>
<td>absolutive case</td>
</tr>
<tr>
<td>ACC</td>
<td>accusative case</td>
</tr>
<tr>
<td>AF</td>
<td>agent form</td>
</tr>
<tr>
<td>ANA</td>
<td>anaphorical pronoun</td>
</tr>
<tr>
<td>ANIM</td>
<td>animate</td>
</tr>
<tr>
<td>ART</td>
<td>article</td>
</tr>
<tr>
<td>AUX</td>
<td>auxiliary</td>
</tr>
<tr>
<td>AV</td>
<td>agent voice</td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
</tr>
<tr>
<td>CISLOC</td>
<td>cislocative</td>
</tr>
<tr>
<td>CL</td>
<td>classifier</td>
</tr>
<tr>
<td>COMPL</td>
<td>completive</td>
</tr>
<tr>
<td>CONN</td>
<td>connector</td>
</tr>
<tr>
<td>CONSTR</td>
<td>construct state</td>
</tr>
<tr>
<td>DECL</td>
<td>declarative</td>
</tr>
<tr>
<td>DET</td>
<td>determiner</td>
</tr>
<tr>
<td>DETRANS</td>
<td>detransitivizer</td>
</tr>
<tr>
<td>DIST</td>
<td>distal demonstrative</td>
</tr>
<tr>
<td>DUR</td>
<td>durative</td>
</tr>
<tr>
<td>ERG</td>
<td>ergative case</td>
</tr>
<tr>
<td>F</td>
<td>feminine gender</td>
</tr>
<tr>
<td>FOC</td>
<td>focus particle</td>
</tr>
<tr>
<td>FUT</td>
<td>future tense</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive case</td>
</tr>
<tr>
<td>HAB</td>
<td>habitual mode</td>
</tr>
<tr>
<td>INANIM</td>
<td>inanimate</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>INTR</td>
<td>intransitive</td>
</tr>
<tr>
<td>IPV</td>
<td>imperfective aspect</td>
</tr>
<tr>
<td>LNK</td>
<td>linker</td>
</tr>
<tr>
<td>LOC</td>
<td>locative case</td>
</tr>
<tr>
<td>M</td>
<td>masculine gender</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Meaning</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
</tr>
<tr>
<td>MED</td>
<td>medial demonstrative</td>
</tr>
<tr>
<td>NTR</td>
<td>neuter gender</td>
</tr>
<tr>
<td>N</td>
<td>noun marker</td>
</tr>
<tr>
<td>N. ...</td>
<td>non-...</td>
</tr>
<tr>
<td>NMLZ</td>
<td>nominalization</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative case</td>
</tr>
<tr>
<td>NP</td>
<td>noun phrase</td>
</tr>
<tr>
<td>L2</td>
<td>second language</td>
</tr>
<tr>
<td>OBJ</td>
<td>object</td>
</tr>
<tr>
<td>P1</td>
<td>predicate-initial / predicate initiality</td>
</tr>
<tr>
<td>PERF</td>
<td>perfect</td>
</tr>
<tr>
<td>PERS</td>
<td>personal marker</td>
</tr>
<tr>
<td>PF</td>
<td>patient form</td>
</tr>
<tr>
<td>PFV</td>
<td>perfective aspect</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>POS</td>
<td>predicate-object-subject constituent order</td>
</tr>
<tr>
<td>POSS</td>
<td>possessive form</td>
</tr>
<tr>
<td>PRET</td>
<td>preterite</td>
</tr>
<tr>
<td>PRG</td>
<td>progressive aspect</td>
</tr>
<tr>
<td>PROSP</td>
<td>prospective aspect</td>
</tr>
<tr>
<td>PROX</td>
<td>proximal demonstrative</td>
</tr>
<tr>
<td>PSO</td>
<td>predicate-subject-object constituent order</td>
</tr>
<tr>
<td>PST</td>
<td>past tense</td>
</tr>
<tr>
<td>PV</td>
<td>patient voice</td>
</tr>
<tr>
<td>REAL</td>
<td>realis mood</td>
</tr>
<tr>
<td>REC.PST</td>
<td>recent past tense</td>
</tr>
<tr>
<td>RED</td>
<td>reduplication</td>
</tr>
<tr>
<td>REFL</td>
<td>reflexive</td>
</tr>
<tr>
<td>REM.PST</td>
<td>remote past tense</td>
</tr>
<tr>
<td>SBJ</td>
<td>subject</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SOP</td>
<td>subject-object-predicate constituent order</td>
</tr>
<tr>
<td>SPO</td>
<td>subject-predicate-object constituent order</td>
</tr>
<tr>
<td>TOP</td>
<td>topic</td>
</tr>
<tr>
<td>TR</td>
<td>transitive</td>
</tr>
<tr>
<td>V2</td>
<td>verb-second word order</td>
</tr>
<tr>
<td>1s</td>
<td>first person singular</td>
</tr>
<tr>
<td>1p</td>
<td>first person plural</td>
</tr>
<tr>
<td>2s</td>
<td>second person singular</td>
</tr>
<tr>
<td>etc.</td>
<td>etc.</td>
</tr>
</tbody>
</table>
# Contents

Abstract i  
Acknowledgements iii  
Abbreviations v  
1 Introduction 1  
2 Defining basic syntactic functions and constituent order 4  
3 Previous global surveys 18  
4 The present survey 21  
4.1 Methodology and Sources 22  
4.2 Numerical and areal overview of results 26  
4.2.1 Pacific Islands 29  
4.2.2 Australia 30  
4.2.3 North America 31  
4.2.4 Mesoamerica 37  
4.2.5 South America 38  
4.2.6 Africa 43  
4.2.7 Europe 47  
4.2.8 Asia 48  
4.2.9 Sign languages 49  
5 Constituent order change and its reconstruction 50  
6 The history of P1 from a phylo-genetic perspective 58  
6.1 Language families with P1 in a majority of languages 60  
6.1.1 Otomanguean 60  
6.1.2 Mayan 62  
6.1.3 Salish 65  
6.1.4 Surmic 67
<table>
<thead>
<tr>
<th>Section</th>
<th>Language Family</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1.5</td>
<td>Totonacan</td>
<td>68</td>
</tr>
<tr>
<td>6.1.6</td>
<td>Enlhet-Enenlhet</td>
<td>70</td>
</tr>
<tr>
<td>6.1.7</td>
<td>Kadugli-Krongo (Kadu)</td>
<td>71</td>
</tr>
<tr>
<td>6.1.8</td>
<td>Wakashan</td>
<td>72</td>
</tr>
<tr>
<td>6.1.9</td>
<td>Xinkan</td>
<td>74</td>
</tr>
<tr>
<td>6.1.10</td>
<td>Chumash</td>
<td>75</td>
</tr>
<tr>
<td>6.1.11</td>
<td>Sahaptian</td>
<td>76</td>
</tr>
<tr>
<td>6.1.12</td>
<td>Chinookan</td>
<td>78</td>
</tr>
<tr>
<td>6.1.13</td>
<td>Kuliak</td>
<td>79</td>
</tr>
<tr>
<td>6.1.14</td>
<td>Tsimshianic</td>
<td>82</td>
</tr>
<tr>
<td>6.1.15</td>
<td>Coosan</td>
<td>83</td>
</tr>
<tr>
<td>6.2</td>
<td>Language families with P1 in a minority of languages</td>
<td>86</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Austronesian</td>
<td>86</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Afroasiatic</td>
<td>93</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Nilotic</td>
<td>100</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Uto-Aztecan</td>
<td>101</td>
</tr>
<tr>
<td>6.2.5</td>
<td>Arawakan</td>
<td>105</td>
</tr>
<tr>
<td>6.2.6</td>
<td>Austroasiatic</td>
<td>108</td>
</tr>
<tr>
<td>6.2.7</td>
<td>Indo-European</td>
<td>112</td>
</tr>
<tr>
<td>6.2.8</td>
<td>Mixe-Zoquean</td>
<td>116</td>
</tr>
<tr>
<td>6.2.9</td>
<td>Chapacuran</td>
<td>118</td>
</tr>
<tr>
<td>6.2.10</td>
<td>Cariban</td>
<td>121</td>
</tr>
<tr>
<td>6.2.11</td>
<td>Tupian</td>
<td>122</td>
</tr>
<tr>
<td>6.3</td>
<td>Language isolates and dialect networks with P1</td>
<td>124</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Alsea</td>
<td>124</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Guató</td>
<td>124</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Kariri</td>
<td>125</td>
</tr>
<tr>
<td>6.3.4</td>
<td>Kuot</td>
<td>125</td>
</tr>
<tr>
<td>6.3.5</td>
<td>Salinan</td>
<td>126</td>
</tr>
<tr>
<td>6.3.6</td>
<td>Yokuts</td>
<td>127</td>
</tr>
</tbody>
</table>

7 Tracing predicate initiality through time

8 Conclusion

References
List of Figures

3.1 WALS distribution of basic constituent order (P1 in yellow, SPO in red, SOP in blue) (Dryer and Haspelmath 2013) ........................................ 20

4.1 Distribution of predicate initiality in the world (Access to interactive map: https://rpubs.com/kbaetsch/P1) ........................................ 28

4.2 Major P1 hotspots (polygons) and smaller P1 cluster (circles) ............ 29

4.3 Distribution of predicate initiality in Pacific Islands (Papunesia) .......... 30

4.4 Distribution of predicate initiality in North America ........................ 32

4.5 Distribution of predicate initiality in Mesoamerica ........................ 37

4.6 Distribution of predicate initiality in South America ........................ 40

4.7 Distribution of predicate initiality in Africa .................................. 43

4.8 Distribution of predicate initiality in Europe ................................. 47

4.9 Distribution of predicate initiality in Asia .................................... 48

6.1 P1 in the Otomanguean language family ................................. 61

6.2 P1 in the Mayan language family ........................................ 63

6.3 P1 in the Salish language family ........................................ 66

6.4 P1 in the Surmic language family ........................................ 67

6.5 P1 in the Totonacan language family .................................... 69

6.6 P1 in the Enlhet-Enenhlet language family ............................ 70

6.7 P1 in the Kadugli-Krongo language family ............................ 72

6.8 P1 in the Wakashan language family .................................... 73

6.9 P1 in the Xinkan language family ........................................ 74

6.10 P1 in the Chumash language family ..................................... 75

6.11 P1 in the Sahaptian language family ..................................... 77

6.12 P1 in the Chinookan language family .................................... 78

6.13 P1 in the Kuliak language family ........................................ 81

6.14 P1 in the Tsimshianic language family .................................... 82

6.15 P1 in the Coosan language family ......................................... 85

6.16 P1 in the Austronesian language family ................................ 86

6.17 P1 in the Afroasiatic language family .................................... 94

6.18 P1 in the Nilotic language family ......................................... 100
6.19 P1 in the Uto-Aztecan language family .................................. 102
6.20 P1 in the Arawakan language family .................................... 105
6.21 P1 in the Austroasiatic language family ................................. 109
6.22 P1 in the Indo-European language family .............................. 113
6.23 P1 in the Mixe-Zoquean language family ............................... 117
6.24 P1 in the Chapacuran language family ................................. 120
6.25 P1 in the Cariban language family ...................................... 121
6.26 P1 in the Tupían language family ...................................... 123
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>WALS Features 81A and 81B: Order of Subject, Object and Verb (1,444 entries)</td>
<td>19</td>
</tr>
<tr>
<td>4.1</td>
<td>Non-P1 languages of the North American West Coast</td>
<td>35</td>
</tr>
<tr>
<td>4.2</td>
<td>Non-P1 languages of interior and Eastern North America</td>
<td>36</td>
</tr>
<tr>
<td>4.3</td>
<td>Constituent order in South America (roughly organized from north to south, families with P1 in <strong>bold</strong>)</td>
<td>41</td>
</tr>
<tr>
<td>4.4</td>
<td>Constituent order in South America (roughly organized from north to south, families with P1 in <strong>bold</strong>) [continued]</td>
<td>42</td>
</tr>
<tr>
<td>4.5</td>
<td>Constituent order in Africa (roughly organized from west to east, families with P1 in <strong>bold</strong>)</td>
<td>45</td>
</tr>
<tr>
<td>4.6</td>
<td>Constituent order in Africa (roughly organized from west to east, families with P1 in <strong>bold</strong>) [continued]</td>
<td>46</td>
</tr>
<tr>
<td>6.1</td>
<td>P1-majority language families</td>
<td>58</td>
</tr>
<tr>
<td>6.2</td>
<td>P1-minority language families</td>
<td>59</td>
</tr>
<tr>
<td>6.3</td>
<td>Basic constituent order in Austronesian languages according to Blust (adapted from Blust 2013: 461–2)</td>
<td>87</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

Probably all languages allow sentences where the verb, or at least part of it, comes first in the sentence. Trust me. Should you not believe me, have you considered English sentences such as these last two? But there is a difference between languages that simply allow verb-initial order in certain contexts, and those that usually place the verb first, or even require it to be first (in pragmatically neutral clauses, as clarified in Chapter 2). Thus, given the three elements Alexander, his teddy bear, and hugged, an English speaker would form the sentence *Alexander hugged his teddy bear*, whereas a Hawaiian speaker would say *Ua pūʻili ʻo Alexander i kāna pea kiʻi*, putting the verb phrase *ua pūʻili ‘hugged’* in first position. That is because the basic word order of Hawaiian is verb-initial, or more precisely, as shall be explained in Chapter 2, predicate-initial (P1).

Like Hawaiian, other Polynesian languages, such as Tahitian, Marquesan and Māori, follow the same basic syntax. Beside these, other P1 languages are Welsh and Halkomelem Salish, as well as the Austronesian languages of Taiwan and the Philippines, all languages that I have studied in one way or another over the years. Though I never consciously chose specifically such languages for study, this syntactic structure is comparatively rare among the world’s languages, as will be demonstrated in detail in this present study. What all of the above languages have in common, is that their speakers are generally aware of the “backward” order of their sentences with respect to the dominant international languages that they are familiar with: English, French and Mandarin Chinese all follow a basic predicate-medial syntax, where the predicate is usually preceded by a nominal element. Naturally, the speakers of such languages ask themselves about the reason for this unusual word order, and if there are any other languages of that ilk. As a linguist, I have been confronted with these questions by speakers of P1 languages on numerous occasions, and I developed an increasing interest in this topic myself. Unfortunately, the existing literature on the topic of P1 is quite sparse and mostly limited to single languages. The few publications that do deal with P1 from a cross-linguistic perspective are focused on the formal derivation of this word order, from a generative perspective. While such studies may have other merits, they are not directed or suited to provide satisfying answers to the questions posed by untrained speakers of P1 languages. Since my personal linguistic background lies in the descriptive-functional and historical tradition, in which P1 has received almost no attention as a distinct category worthy of dedicated research, I was convinced to dedicate my dissertation to the topic of predicate initiality. Hence, the impulse for this present study came to a large extent from linguistically untrained friends.
who developed an awareness of and curiosity about the typological peculiarity of their native languages. Their inquiry fomented my proclivity for syntactic typology, giving rise to this study.

This work is thus concerned with languages that regularly place the predicate at the beginning of the clause, known as predicate-initial (P1) languages. Its first objective is to provide as exhaustive a survey of the global distribution of predicate initiality (P1) as is possible given the current extent of linguistic documentation. Beyond that, it traces the occurrence of P1 back in time, thereby investigating its historical origins in its phylogenetic and areal contexts, to the extent that syntax is at all reconstructible beyond written records. In other words, the purpose of this thesis is to investigate the following two questions:

1. What is the distribution of P1 languages?

2. How did this distribution develop historically?

After laying a conceptual foundation to allow the cross-linguistic comparison of word order in Chapter 1 and reviewing the extent of previous research in this area in Chapter 3, an answer to the question of distribution and frequency of P1 will be provided in Chapter 4, where the topic is approached from a typological-areal perspective. The subsequent chapters approach P1 from a historical perspective: Chapter 5 discusses the methodology of syntactic reconstruction, while Chapter 6 investigates the occurrence of P1 from a phylogenetic-historical perspective, offering an in-depth description of P1 by language family, and reconstructing the syntactic history of the relevant families. Chapter 7, finally, provides a final overview of the rise and fall of P1 across time and space.

This survey consolidates data from available areal studies and hundreds of grammars on languages from across the globe, compiling the most comprehensive atlas of predicate initiality in the world’s languages. It confirms the common perception that predicate-initial order is significantly less frequent compared to predicate-medial and predicate-final word orders. With the parameters used in this study, only 811 languages, or around 11% of the world’s languages, make common use of P1. Further, this study reveals that predicate initiality is not evenly distributed across the world but occurs in a few hotspots: Pacific Islands (including Taiwan and the Philippines), the Pacific Northwest coast of America, Mesoamerica, the Maghreb, and the East African Rift. In these few areas, P1 order is the areal, either within a single language family (Austronesian on Pacific Islands, and Afroasiatic in the Maghreb), or across various neighboring language families in a linguistic area. Outside of the few above-mentioned areas, only a few dispersed P1 languages are found. P1 is virtually absent from all of Eurasia and Australia (with a few exceptions), and

---

1 There is no general consensus on the number of languages in the world, and often not even on the number of languages within a language family. For the simple reason of consistency, this study treats language varieties as they are encoded on Glottolog. Though Glottolog definitely lies on the split-friendly side of approaches—meaning that it tends to treat intelligible dialects as separate languages—it has the advantage of presenting a publicly accessible total list of the world’s languages.
it is notably absent from two of the major hotspots of linguistic diversity: West Africa and New Guinea. P1 thus forms obvious areal clusters. This study suggests that the reason for this heterogeneous distribution is that P1 rarely develops independently, as there are only few possible syntactic pathways for it to develop language-internally across time. It appears that P1 can be reconstructed to the earliest proto-stages of most families that currently feature this order, i.e. to the horizon of observable history. Many languages, especially in modern Indonesia and northern Africa, can be demonstrated to have lost an earlier P1 syntax over time. Yet, the history of P1 is not just one of syntactic erosion: there are a few cases where P1 verifiably entered a language family through intense language contact, the most obvious case being the Nahua branch of the Uto-Aztecan family, spoken by Aztecs. Speakers of Nahua moved into the P1-dominant area of Mesoamerica in recoverable history and adopted the local constituent order, as discussed in §6.2.4. Other examples of an obvious adoption of P1 through language contact are the Nicobarese languages (§6.2.6) and the language isolate Kuot (§6.3.4). The strong areal clustering of P1 in a few hotspots suggests that P1 spread by language contact in several places, though the direction of diffusion can seldom be determined due to the time depth and lack of relevant attestations.

As mentioned above, I intend this study to provide a satisfying answer to the inquiry of linguists and laymen alike about the distribution and origins of predicate-initial structures in the world’s languages. Beyond that, I hope that this work serve as a stepping stone into further research into P1 languages as a type. Although this survey does not delve into questions of generative representations or structural correlates of P1 syntax in general, research into these areas necessarily needs to cover the breadth of P1 languages with all their variation. In other words, a study based entirely on data from the most prominent P1 branches (the most commonly cited being Philippine, Mayan, and Celtic) is hardly qualified to extrapolate on the structure of P1 languages in general, just as one cannot make overarching statements on mammals based only on canines, felids and rodents. The present study demonstrates the breadth of P1 languages across areas and linguistic phyla, and it can thus serve as a point of departure for subsequent studies on P1.
Chapter 2: Defining basic syntactic functions and constituent order

Word order is one of the principal areas of research within the field of linguistic typology. Since Greenberg 1963, typologists have been on the quest for correlations between basic word order and other structural patterns in human language. Previous typological studies have been concerned primarily with the order of verb and object, and how that correlates to the word order inside other constituents, such as the order of head noun and adjective, head noun and possessor, and head noun and relative clause. Such correlations have been claimed to be stronger or weaker by different authors (see e.g. Hawkins 19833, Campbell et al. 1988, Dryer 1992, Bickel 2007, Song 2014).

Not too seldom, there is disagreement about the basic word order of a given language. Some have even questioned the very notion of basic word order itself, as a cross-linguistically valid category (Brody 1984), or for particular languages (Mithun 1987). Hence, it is important to work with an explicit and precise definition of word order. Typologists claim to investigate syntactic order by looking at the “subject” (S), “object” (O), and “verb” (V), but their definition of these terms often differs from how other linguists would use them. Usually, subjects and objects are understood to be syntactic constituents, i.e. they are defined by their function in clauses, whereas the verb is a word class (part of speech). Of course, we should not be comparing apples and oranges. A description of syntactic structure should not focus on word classes (nouns, adjectives, verbs), but primarily on syntactic constituents. Also, our description should ideally not be restricted to clauses with verbal predicates, but it should be more broadly applicable. The semantic nature of the typologists’ definition of the three core clausal constituents is apparent in such descriptions as “the entity which initiates an action (S), the entity at which that action is directed (O), and the action itself (V)” (Song 2012: 2). Typologists aptly describe these elements as semantic “macroroles” (Foley and Van Valin 1984, Van Valin 2005) or “prototypes” (Dowty 1991). Again, a description of syntactic structure should be concerned primarily with syntactic constituents, not with semantic roles. Furthermore, a typology of the order of semantic elements runs into the following two problems. First, it ignores the morpho-syntactic and pragmatic patterns of ergative languages, where the absolutive (i.e. subject of intransitive clause and patient-like argument of a transitive) forms a coherent category. Second, it is forced to limit the description to a single grammatical voice because semantic roles are assigned to different syntactic categories in different voices. This
limitation is particularly problematic in languages with symmetrical voice, because the
Typologist’s preference for one voice over the other will be more or less arbitrary.

In view of these shortcomings in the typologist tradition, this study is concerned only
with syntactic functions, particularly with the ordering of the predicate, the subject, and
the object in their traditional definitions as syntactic constituents, rather than word order
in other domains, such as the ordering of adjective and head noun, that of modal verb and
main verb, etc. It is not concerned with semantic roles and word classes. Consequently,
the symbols “V” for verb and “A” for agent are consistently avoided, unless in direct
quotation of another author. Instead, this study uses the symbols P, S, O for the basic
syntactic constituents. In an effort toward less ambiguous terminology, this study refers
to the ordering of the predicate, subject and object as constituent order, although the more
commonly used phrase in the literature for the same is “word order.” (Needless to say,
constituents often consist of several words.) This discussion may seem like terminological
detail, but the idiosyncratic definition of commonly used linguistic terms can lead to
serious confusion.

The predicate is understood here as the function defining valency, i.e. the number of
core arguments in a clause. It is in that sense the structural core of every clause. As such,
the predicate, as defined here, does not include the object, as some theoretical approaches
to syntax understand it. However, I do understand the predicate to include auxiliary
verbs. This is important to specify because some languages commonly split the predicate
complex in such a way that only part of the predicate precedes both core arguments. A
well-known example of such a structure are the Celtic languages (cf. §6.2.7), where only
a small set of main verbs, including the commonly used auxiliaries precede the subject,
whereas most main verbs follow the subject. Such languages are included in this study.
The predicate always contains a verb in European languages, but this is not true for all the
world’s languages: in languages without an overt copula, equational clauses lack a verb,
though their core is still the predicate. In some prominent P1 languages, especially those
of the Austronesian family and those of the Pacific Northwest Sprachbund, the existence
of nominal vs. verbal word classes is debated, which, if true, could mean that no predicate
actually contains a verb. Even without verbs, predicates can be identified in all of the
world’s languages: the predicate is the one constituent that determines the number of
core arguments in the clause, and it usually coincides with the pragmatic function of the
comment (or “rhema”) of an utterance. To illustrate this, the following examples from
Hawaiian demonstrate the clause-initial position of the predicate, regardless of its nature.

(2.1) Different types of predicates in Hawaiian [haw, Austronesian]

a. Stative verb:

\[ \text{Akamai} \quad \text{ʻo} \quad \text{Kaleo.} \]

smart \quad SBJ K.

‘Kaleo is smart.’

\footnote{Some readers may be familiar with “P” being used for the patient, as e.g. by Comrie. This is not how “P”
is used in this study.}
b. Intransitive verb:

\[ \textbf{Noho} \text{ o } \text{ Kaleo ma Mānoa} \]

sit sbj K. loc M.

‘Kaleo lives in Mānoa.’

c. Transitive verb:

\[ \textbf{Honi} \text{ o } \text{ Kaleo iā Keahi.} \]

kiss sbj K. obj K.

‘Kaleo kisses Keahi.’

d. Noun phrase:

\[ \textbf{Koʻu hoa aloha} \text{ o } \text{ Kaleo} \]

1s.poss friend love sbj K.

‘Kaleo is my friend.’

Thus, restricting the discussion on constituent order to verbs—or even a sub-type of verbs—neglects the commonality of predicate-initial constituent order in Hawaiian. Yet, for the purposes of the present study, predicates can informally be understood to refer to verbs, since this study is only concerned with data from transitive clauses (see below), in which predicates are usually instantiated by “verby” elements, as in (2.1c). However, in accordance with the discussion above, this informal practical definition should not be understood to imply that predicates and verbs are always synonymous, nor even that all languages have verbs. Though more sophisticated definitions are certainly necessary to define predicates in single languages, or even language families, a typology of the world’s languages requires a readily operationalized and unambiguous definition.

The categories of subject and object are less clearly defined, because their diagnostic characteristics can vary greatly across languages. Usually, the single argument of an intransitive clause is readily identified as the subject, but opinions often differ about transitive clauses. Usually, one argument of a transitive clause can be demonstrated to be treated like the single argument of an intransitive clause. This argument is universally considered a core argument, although there is not always a consensus on its analysis as a subject or object, especially in languages with ergative alignment, where patients are regularly assigned to that argument role. (Again, this demonstrates the conflation of syntactic and semantic roles discussed above.) The other core argument of a transitive clause can usually be diagnosed on the basis of a syntactic affinity with the predicate, be that substantiated through verbal agreement, constituent order, the syntactic option of promotion to subjecthood, or others. However, these diagnostics differ across languages, depending on the language’s morphosyntactic profile, and this variability has led many

\[\text{\textsuperscript{2}However, this is not the case for split-S systems, where even the single argument of an intransitive can be marked in different ways, depending on its semantic role. In some of these languages, word order may conceivably not be described without reference to certain semantic roles.}\]

\[\text{\textsuperscript{3}This is not true for a small set of languages, including Niimi’ipuutímt (Nez Percé) or Ainu, which are described as employing tripartite alignment. The overt morphology in such languages readily allows the identification of core arguments.}\]
to disagree on, or even doubt altogether, the cross-linguistic applicability of the notions of subject and object. Because of the seemingly deadlocked disagreement on the cross-linguistic notion of subjecthood, many researchers have abandoned the terms “subject” and “object” in favor of more neutral, less biased/controversial term for some languages, such as the “pivot” or “trigger” in Austronesian languages (cf. Schachter 1976, LaPolla and Poa 2006). Probably, the typologist tradition to work with semantic macro-roles arises from the same motivation to not get caught in theoretical discussions and language-specific traditions of terminology. Although I acknowledge the problem of a universally applicable concise definition of “subject,” I nevertheless adopt that term from the primary literature for this study, given that its usage for particular languages and language families is generally less controversial, and especially given that the differentiation between core arguments has no more than tangential bearing to the present study.\footnote{I also give preference to the terminology of the syntactic functions (subject and object) over the one of the macro-roles (A and P) where both are in use, for the reasons stated above.}

The concept of predicate initiality (P1) is in this study understood to refer to any order of the predicate, subject, and object in which the predicate is the first of these. In other words, this study includes both PSO and POS (traditionally labeled VSO and VOS), as relevant as this distinction may be for other studies, in particular for theoretical syntax (cf. Clemens and Polinsky 2017). There are three reasons for this conflation: (a) clauses with two overt (non-pronominal, as discussed below) arguments are rare in natural discourse in general, verging on ungrammaticality in some languages (Givón 1975, Chafe 1987, Du Bois 1987, Payne 1986, Gerdt and Hukari 2009); (b) many P1 languages allow free ordering of core arguments, as long as they follow the predicate (e.g. Hiligaynon, Lealao Chinantec [Rupp 1989: 39]);\footnote{This problem was also recognized by Dryer, who notes: “While the feature [relative order of S, O, V] shown on this map [WALS] is perhaps the single most frequently cited typological feature of languages, it is now recognized that it represents a clause type that does not occur especially frequently in spoken language; it is more common that at least one of the two arguments of a transitive clause will be pronominal, and in many languages pronominal subjects are expressed by verbal affixes.” (Dryer and Haspelmath 2013)} (c) treating PSO and POS languages as a coherent group allows us to investigate whether they share typological correlates, similar to other types of languages defined by word order, notably PO and OP (cf. Greenberg 1963, 1966).\footnote{It should be noted here that the conflation of SO and OS order for the present purposes of this study is by no means intended be interpreted that their relevant order is not relevant in general. Similarly, any correlates of PO vs. OP order are not doubted in this study, but they are simply not relevant here, because all the languages treated in this study are PO.} It is but a fortunate consequence of this conflation that the contentious issue of identifying subjects and objects across various language families can and will be avoided in this study, because the relative order of subject and object is irrelevant. Thus, where reference is made to subjects and objects in this study, these terms merely reflect the wording of the primary sources, but they can be understood as Argument 1 and Argument 2 for the purposes of this study.

To restate the research question, this study is concerned with languages where the predicate precedes its core arguments. Core arguments are understood to be the syntactic constituents that are subcategorized for by the predicate. This is usually taken to mean...
that they are *required*, as opposed to adjuncts, which are optional addenda. However, so-called pro-drop languages can freely drop any constituents beside the predicate. Thus, optionality is not a reliable diagnostic for argumenthood. Indeed, finding a cross-linguistically valid diagnostic for core arguments is difficult, if not impossible. However, each language has a distinct set of diagnostics that can be used to identify core arguments. Beside optionality, some common diagnostics are case marking, agreement, and eligibility to subject promotion by grammatical voice. This study follows the language-specific definitions for core arguments in the literature.

This study deals exclusively with the actual surface order of constituents, as opposed to some proposed deeper underlying order that could be derived with the assumption of movement and zero elements. One could argue that the concept of P1 in this study is a mere epiphenomenon which may be the result of divergent underlying structures, but this measure is taken because the potential analyses of a given surface pattern are manifold, depending on the theoretical approach, and they often conflict each other. The surface structure, on the other hand, is a given that everyone agrees on. Besides, what *isn't* an epiphenomenon? It shall also be stressed that this study does not deal with typological correlates of constituent order, such as the orientation of adpositions or the order of demonstrative and head noun. Although such patterns are often taken as evidence for the identification of a language's basic constituent order, such arguments are (where explicit) consciously ignored in this study. The reason for this is that while these typological correlates are certainly worthy of study—indeed I hope that this survey supports further study in this field—excessive top-down application of typological findings to the analysis of individual languages skews the typological perspective in that it strengthens the predominant patterns while explaining away minor patterns. Although typology may certainly inform and assist language documentation and the analysis of single languages, the general direction of dependence should be the opposite: typology is based on language descriptions, rather than vice versa. Therefore, the evidence for constituent order considered in this study shall come precisely from the surface order of constituents, with considerations on pragmatics and frequency of usage, as discussed below.

The syntactic definition of P1 employed in this study allows us to include a broad range of constructions, notably not requiring the exclusion of varying grammatical voices. In this sense, it differs from other studies of constituent order that restrict discussion to a very narrowly defined construction type, such as the qualifications used by Siewierska (1988: 8) to identify a prototypical transitive clause: stylistically neutral, independent, indicative mood, containing full noun phrase (NP) participants, where the subject is definite, agentive and human, the object is a definite semantic patient, and the verb represents an action, not a state or an event. However, I do adopt a set of qualifications for sentences considered in this study. These conditions do not necessarily restrict the discussion to certain clause types, but they rather *facilitate* the recognition of true basic constituents. They also ensure that this study employs the same criteria across different languages. Without these, arguably any language may *appear* to be P1, as predicates may arguably occur in
first position in any language at least in some restricted context (see Chapter 1). Accordingly, unless otherwise specified, the term ‘predicate initiality’ or simply ‘P1’ will be used in this study to refer to the occurrence of the predicate before its core arguments under the conditions listed in (2.2), and further elaborated below.

(2.2) Conditions for the identification of unmarked constituent order.

a. Only transitive clauses are entirely conclusive.
b. Pronouns and demonstratives are not conclusive as arguments.
c. Indefinite arguments are not conclusive.
d. Negative and interrogative clauses are not conclusive.
e. Elements in the (left/right) periphery are not indicative of constituent order in the main clause.

First, the conclusive evidence used in this paper comes from transitive clauses, not from intransitive clauses, though the latter are addressed marginally, as well. Most often, though there are many exceptions, the constituent order of intransitive clauses follows the same pattern as that of transitive clauses, simply lacking the object. However, some languages feature certain types of intransitive clauses that show alternative constituent order. For example, many Romance languages such as Spanish (2.3a), as well as Russian (2.3b) show predicate-initial order in existential and presentative clauses (with verbs of existence or appearance, and change of state), though the regular constituent order is SP(O) (Vincent 1988: 61).

(2.3) Existential/presentative clauses in European languages

a. Spanish [spa, Indo-European, source: author’s data]
   *Hay dos opciones.*
   ‘There are two options.’

b. Russian [rus, Indo-European, source: author’s data]
   *Приближает-ся чёрная машина.*
   approaches-refl black car
   ‘A black car is approaching.’

In many South American languages, intransitive clauses are more generally P1, whereas transitive clauses do not show P1. This pattern is followed e.g. by the Arawakan language Paumari, as shown in (2.4) with the predicate in initial position in the intransitive clause (2.4a), but with the predicate following the subject in the transitive clause (2.4b).

(2.4) Paumari [pad, Arawakan, source: Chapman and Derbyshire 1991]
Transitive clauses need not contain two overt phrasal arguments; in fact, they usually do not. Most often, at least one of the core arguments is expressed by a pronoun or dropped, which is an economical way of referencing discourse-given participants. This phenomenon is the result of the natural development of discourse, in which new information is introduced incrementally, as opposed to all at once (cf. Givón 1975, Du Bois 1987, Chafe 1992). While pronouns are not taken as indicative of constituent order (see below), object pronouns do indicate transitivity. Transitivity is harder to determine if core arguments are completely dropped, but in some languages (e.g. Oceanic and Salish languages) overt transitive morphology provides a definite cue. Although both subjects and objects (for the present purposes just two differently marked core arguments) need to be shown to follow the predicate in a language for it to be identified as P1 in this study, they need not co-occur in the same clause. This diagnostic procedure is exemplified in (2.5) with a sentence from Hul’q’umi’num’ (=Halkomelem).

(2.5) Hul’q’umi’num’ transitive clause with overt transitive marking

Hul’q’uminun’ [hur, Salish, source: author’s data]

Ni’ =tst lum-nuhw tthu= spe’uth.

Aux 1P see-TR the= black.bear

‘We saw the black bear.’

In (2.5), transitivity is indicated most directly and evidently by the transitive suffix -nuhw on the predicate, but also by the presence of two core arguments, one represented by the full phrase tthu spe’uth (oblique arguments would be marked by the proclitic ‘u=’), and the other one by the enclitic =tst (adjuncts are never expressed by clitics in Hul’q’umi’num’).

From (2.5), we can deduce that objects follow the predicate, though other clauses need to be examined to determine the position of the subject.

Second, clitic elements are not considered as arguments in the identification of constituent order, because their placement so often differs from phrasal arguments, leading to ambiguous data. As clitics are, in contrast to affixes, often not marked as such in the literature, the commonly cliticized categories of personal and demonstrative pronouns are consistently excluded from the evidence for this study. This allows us to identify for example Fijian (2.6) as P1, even though the placement of pronouns might lead us to suspect a different conclusion, as is indeed argued for by Schütz (2014).

(2.6) Fijian transitive clause with subject pronominal agreement by means of proclitic

Fijian [fij, Austronesian]
The same situation with enclitics functioning as pronominal agreement has led to discussion about constituent order in Nicobarese languages, and many others. Although there may be sound arguments (e.g. based on intonation) for true argumenthood of pronouns in such languages, this study is only concerned with the placement of phonologically independent NP arguments.

Third, only definite arguments are considered conclusive for the purposes of this study, where definite is understood to mean that the referent is identifiable by the speaker and (presupposed by the speaker) by the listener (Lyons 1999: 2–7). This measure is taken because indefinite participants (especially patients) are often incorporated into the predicate, expressed as oblique phrases, or in some other way not valent as a core argument, but are still erroneously identified as a core argument. Definite agent and patient participants, on the other hand, usually function as core arguments. The two Hawaiian clauses in (2.7) illustrate this difference.

(2.7) Incorporated participant (a) vs. full argument (b) in Hawaiian [haw, Austronesian, source: author’s data]

a. [P[head][modifier]] S:
   \[Ke= \[inu\]_head \[wai\]_mod nei]_P \[=au]_S
   \[ART= \) drink \) water \) here =1s
   'I am drinking water.'

b. P S O:
   \[Ke= \[inu \) nei]_P \[=au]_S \[i= \) ka= wai]_O
   \[ART= \) drink \) in \) =1s \) OBJ= \) ART= \) water
   'I am drinking (the) water.'

In (2.7), only (2.7b) can have a definite reading, while the incorporated noun *wai* ‘water’ without an article in (2.7a) cannot be definite. In other words, a definite patient needs to be expressed by an object, while an indefinite patient may be incorporated. In Hawaiian, incorporated patients still follow the predicate head and do not create any confusion in terms of constituent order. But the Piman language O’odham [ood, Uto-Aztecan] places indefinite patients in front of the predicate head, while definite patients follow the predicate, leading to apparent ambiguity with respect to the relative ordering of
predicate and object (Payne 1987). One could argue that only the latter type represents a true object. Hence, accepting only definite arguments is a cautionary measure to avoid accepting misidentified objects. Definiteness is unmarked in many languages but luckily is obligatorily marked in the European languages that are commonly used as the matrix languages for grammatical descriptions (English, Spanish, French). Hence, translations in grammars usually mark definiteness.7

Fourth, non-default clauses like negative and interrogative ones often involve syntactic structure that obscures constituent order, and hence are used only as secondary evidence in this work. Negators represent various syntactic elements across different languages. In most European languages, the negator is an adverb, which does not affect the order of the basic syntactic constituents. In many other languages (e.g. Japanese or Turkish), the negator is an affix on the verb, which does not affect constituent order, either. However, in a number of languages, the negator is a verb (e.g. Finnish) and/or may itself function as the predicate (e.g. in Philippine languages), which may truly or only apparently alter the relative order of core arguments and the predicate. In addition to this, negative clauses often involve indefinite objects, which are commonly marked differently from definite objects in affirmative clauses (see above). In order to avoid such complications and possible misinterpretation, negative clauses are categorically excluded as primary evidence in this study.

Questions often appear to differ from declarative clauses in constituent order. In content questions, the question word, which may represent a core argument, is often fronted. This happens in German (2.8).

### (2.8) German question word in clause-initial position

- **German** [deu, Indo-European]
  - *o [Was] p [isst] s [der Büffel]?*
  - what eats the buffalo
  - ‘What does the buffalo eat?’

Even though German main clauses commonly follow the order SPO, the structure in (2.8) is only superficially a deviation from the basic constituent order. In German, the clause-initial position is reserved for an element in focus, be that the subject, the object, an adjunct, or even a verb. In (2.8), the focused element is the object, instantiated by the question word *was* ‘what’. The conjugated verb, forming the core of the predicate, follows in clause-second position, and any remaining arguments follow after that.

Other languages may give the appearance that the constituent order of questions is different from that of statements because the question word functions as a predicate but may for semantic reasons be misinterpreted as an argument. This is the case in Ilokano (2.9) and other Philippine languages.

### (2.9) Ilokano question word in clause-initial position

---

7This study necessarily relies on the quality of the translations in reference grammars, taking their value for definiteness as decisive.
Ilokano [ilo, Austronesian, source: author’s data]

\[ p[Ania] s[[kanen]_{head} [ti \ anuang]_{mod}]? \]
what \ will.be.eaten \ the \ buffalo

‘What does the buffalo eat?’

Despite the superficial similarity between (2.8) and (2.9), the former shows OPS order, while the latter shows P1. This illustrates how content questions may complicate the discussion, even if they are not truly exceptional or to be excluded by principle.

Polar questions may truly involve a change in constituent order in many European languages that have the syntactic process of inversion. Consider the French sentence (2.10), showing PSO order resulting from inversion.\(^8\)

(2.10) French [fra, Indo-European] polar question showing PSO order resulting from inversion

a. \textit{Avez-vous \ faim?} \\
\( p[\text{have}]s[2p] \ o[\text{hunger}] \) \\
‘Are you hungry?’

b. \textit{Vous \ avez \ faim.} \\
\( s[Vous] p[\text{avez}] o[\text{faim}] \) \\
‘You are hungry.’

Consequently, negation and questions are excluded as primary evidence for the identification of basic constituent order in this study.

Another process that appears to result in non-default constituent order is dislocation, by which an element (often the topic) appears in the left or right periphery of a sentence. Consider the following English sentence: \textit{This table, my dad made (it) in his youth}. While it appears to represent OSP order, with an object \textit{this table} in initial position, there is good reason to assume that the phrase \textit{this table} is not fully integrated into the clause: first, it is separated from the rest of the clause by a prosodic break (indicated by the comma), and second, it is optionally cross-referenced by a pronoun in the main clause. Thus, one may argue that the constituent order of this sentence is still SPO, though this is not directly obvious. While dislocation of core arguments is not particularly common in English—the grammaticality of the illustrative example above is even questionable to some speakers—this process is very common in other languages. In Asian languages, such as Chinese or Japanese, the dislocation of a topic is particularly common. In many European languages (and quite generally in many of the world’s languages), elements in focus often appear in sentence-initial position. Usually, the fronted elements are typically nominal phrases, but some languages allow verbal elements to be fronted, as well. In German, only non-finite verb forms (such as infinitives or participles) may be fronted in this fashion, namely when they are in contrastive focus ex:germanfocus, but other languages like Lithuanian

\(^8\)This example just serves to demonstrate a principle. Of course, the sentence is not conclusive about constituent order also because its arguments are instantiated as a pronoun and an indefinite noun.
(Vytautas et al. 2006: 695–7), or Dagik of the Niger-Congo family (Vanderelst 2016: 234), also allow finite verbs to be fronted.

(2.11) Focalization in German [deu, Indo-European, source: author’s data]

a. Sehen kann ich ihn (zwar), aber hören nicht.
   see.inf can I him (indeed) but hear.inf not
   'I can indeed see him, but I cannot hear him.'

b. Gegessen hat sie (zwar) schon, aber getrunken noch nicht.
   eaten has she (indeed) already but drunk yet not
   'She has already eaten but not yet drunk.'

c. *Lacht man (zwar) gerne, aber weint nicht.
   laughs one (indeed) gladly but cries not
   intended: 'One laughs gladly, but does not cry gladly.'

Such dislocated elements are often treated as falling outside of the clausal core and may thus not be relevant to the question of this study. Furthermore, the analysis of such complex sentences requires in-depth understanding of each language under investigation, which is unrealistic for a large typological study like this one that surveys many languages. Yet, linguistic descriptions usually identify dislocation (albeit often in fuzzy terms, such as “emphasis”), especially since the affected sentences are often accompanied by unusual stress or intonation patterns, but languages with productive dislocation of topics are potentially misleading. Since one common application of dislocation is the focalization of an argument, sentences with anything else than broad focus (i.e. sentences with the narrative function of event reporting or answers to the question “What happened?”) should be disregarded from this study. In practice, this standard is difficult to uphold because very few language descriptions even address such pragmatic concerns, and analyzing focus in a narrative text requires deep familiarity with the language. Hence, the only standard that can realistically be upheld in a large-scale typological study such as this one is that sentences which are explicitly identified by a language expert as marked due to the use of dislocation/topicalization/emphasis are excluded from the evidence considered in this study.

Other studies on constituent order, whether in a single language, in a language family, or in broader language typologies, take into account further criteria for the identification of a basic constituent order (cf. e.g. England 1996: 96–7). One common consideration is that of relative animacy between the arguments. That is, transitive scenarios in which there is a clear gradient in animacy between the involved participants (e.g. a person and an inanimate object) sometimes allow more syntactic variability or do otherwise follow a different constituent order from sentences in which the two participants are on par in terms of animacy (e.g. two persons). One reason for this difference is because the functional load on syntax to assign semantic roles is minimal where the intrinsic semantic nature of the participants already predicts the direction of action, while syntax is required to disambiguate the situation where either participant is equally capable of and likely to perform
the action. Another reason is that participants with higher animacy are often promoted syntactically, such that the subject position is reserved for them. Such as system is known as a direct-inverse alignment. In other cases, the relative ordering of the arguments is dependent on animacy, even though syntactic alignment is insensitive to it. While such considerations are certainly worthy of attention in the discussion on constituent order in general, they are not relevant in this present study, as neither the assignment of subject and object roles nor their relative ordering is at the focus of this study, but rather the order of the predicate and both of its arguments.

Following the guidelines presented in (2.2) leads to unambiguous results for constituent order in most languages. This constituent order would be considered as a language’s dominant or basic constituent order. However, there are some languages (usually but not exclusively highly synthetic ones) whose constituent order is variable even under these conditions, namely languages that have been described to have “free word order.” In such cases, some researchers have employed corpus-based statistics to identify the most common order. However, Dryer (1995) presents convincing arguments why the most common constituent order is not necessarily the basic, i.e. least marked, constituent order. Not everyone would agree with Dryer on this point, but the practical consideration that such statistical surveys were not performed with the conditions (2.2) in mind renders any statistics inconclusive for this study. Needless to say, most languages have not been the subject of any such statistical count whatsoever, and of course a methodologically coherent survey of constituent order spanning all the world’s languages that would take frequency in discourse as a criterion for determining basic constituent order is impossible for practical reasons.

So-called “free” or “flexible” constituent order can refer to different linguistic types: 1) languages that have more than one commonly used constituent order, and 2) languages that do not have any restrictions on constituent order, i.e. allow any order. The difference between these two kinds is that constituent order in the first type is understood to represent, or even be governed by, syntactic principles, whereas constituent order in the second type is independent of syntax but is rather representative of discourse pragmatics (cf. Thompson 1978, which typifies languages by grammatical, pragmatic, or both grammatical and pragmatic word orders). The reason for the syntactic variation in the first type could be because of the strong influence of a dominant language with a different syntactic profile, or it could be because pragmatic profiling by the means of syntactic dislocation is such a common stylistic device that it competes with the unmarked order, and in some cases the language is even undergoing a syntactic shift in which the marked order is becoming semantically bleached and is establishing itself as the new norm. It is worth noting that SPO is a possible or even common alternative in arguably all P1 languages, as the subject is readily topicalized, given its high discourse continuity and hence topicality. It has even been argued that P1 languages are generally more flexible than other types of languages, that the same rigidity in constituent order that many predicate-final languages exhibit is never observed in languages with a basic P1 order (Edward
Keenan, quoted in Payne 1995: 454). Languages of the second type may also show a statistical preference for some constituent order(s) over others, but this can usually be deduced from broader pragmatic principles, such as “new participants are introduced early in the sentence, and given participants are placed in the right edge of the sentence.” Thus, it is perfectly possible that a language of the second type is analyzed as following a P1 basic order, but a preference for P1 in this case is merely an emergent pattern from non-syntactic principles. Languages of this type often have no truly unmarked constituent order, as every order conveys specific presuppositions about the discourse status of the participants. Pragmatic and syntactic principles can be difficult to pull apart in some languages, as discourse status is often conveyed through syntactic means. However, since this study is concerned primarily with syntax, rather than with pragmatics, it excludes languages whose constituent order is explicitly described to be defined by pragmatic, rather than syntactic, factors, and languages that commonly employ all possible constituent order configurations. Although languages with purely pragmatic constituent order deserve particular attention, as they represent a possible historical pathway into the genesis of a P1 constituent order (as suggested by Givón 1977), they are beyond the scope of this study of syntactic structures, also for practical concerns, as these languages are numerous in North America and Australia. Yet, languages of the first type above, i.e. languages with two (or even three) alternative unmarked constituent orders, are included in this study, if at least one of these orders is P1, regardless of whether P1 is the most frequent order. Thus, instead of adopting a random cut-off point for methodologically diverse surveys of typologically diverse languages, this study includes any languages that have been described (or that I have otherwise observed) to commonly use P1 under the conditions in (2.2), and whose constituent order is described to be based on syntactic principles.

The criteria adopted here might give the impression that this study addresses but a narrow perspective on a very specific type of construction, excluding many other possible types. However, this impression is not justified. First, consider that probably every language shows sentences with the predicate in initial position. Hence, it is necessary to define precisely which structures we are comparing. Furthermore, and as explained in the preceding paragraphs, it is possible that even derived constructions such as interrogative or negative sentences, or such involving indefinite patients, or dislocation, can be shown to follow the same basic constituent order, just that this is harder to determine without an in-depth understanding of the syntactic processes and constructions of each language under investigation. Since such an investigation is hardly realistic for a large typological survey like this one, the conditions in (2.2) simply help to identify constructions where basic constituent order is most directly apparent. Besides that, it is possible that some constructions really do involve different constituent order than others. This typological study is not about assigning languages to mutually exclusive categories, but to recognize the distribution of a certain type. In other words, the languages included in this study as “P1 languages” need not be exclusively P1 but have/allow P1 order as a common and

---

9Cf. Payne 1995 for a discussion of the function of non-P1 clause types in basically P1 languages, a topic which shall not concern us here any further.
unmarked constituent order type, whether this would be considered their one and only true basic constituent order or not, and regardless of the question whether every language even has basic constituent order.
Chapter 3: Previous global surveys

Most typological work on constituent order so far has been concerned with the relative order of the predicate and the object, as they are often considered to form a constituent, the verb phrase. Following the lead of Greenberg (1963, 1966) and (Lehmann 1978 etc.), much work has been dedicated to finding a correlation between that and the relative ordering of heads and modifiers in other constructions, notably in noun phrases (order of head noun and adjective, head noun and possessor, and head noun and relative clause). Little attention has been paid to P1 as a distinct category. Yet, there are several publications dedicated to P1 in particular languages or language families, which shall not be discussed here but rather in Chapter 6. Payne’s (1990) *The pragmatics of word order: Typological dimensions of verb-initial languages* appears to be a general treatise of verb-initial languages but is actually a detailed description of the Peruvian language Yagua. However, Payne does treat P1 as a typological category on a broader scale in subsequent publications, most notably in the book chapter *Verb initial languages and information order*, where she also provides a coarse overview of the distribution of P1 across the world’s languages: “Austronesian, Salishan and Wakashan and perhaps other families of the northwestern United States and Canada, Otomanguean, Mayan, southern Arawakan, Celtic, Semitic, Nilo-Saharan, and isolates scattered here and there (e.g. Yagua, Taushiro, both located in the Peruvian lowlands)” (Payne 1995: 449).

While predicate initiality has not so much been the concern of typologists, it has been debated by syntacticians working in formal approaches. The most comprehensive works in that field are Carnie and Guilfoyle’s (2000) *The syntax of verb-initial languages* and Carnie et al.’s 2005 *Verb first: On the syntax of verb-initial languages*, both concerned primarily with the question of whether P1 syntax can generally be derived from some universal underlying structure. The conclusion in Carnie et al. 2005, which considers data from a wide variety of languages worldwide, is no. The reader is also referred to Clemens and Polinsky’s (2017) *Verb-initial word orders (Primarily in Austronesian and Mayan languages)*, which concludes that verb-initial languages do not form a uniform group and that VOS and VSO languages differ in terms of syntactic mapping.

Of special interest for the study of P1 languages is Downing and Noonan’s (1995) *Word order in discourse*, which discusses the pragmatic functions and effects of constituent order alternation in various languages, but with special attention to P1 languages.

In terms of quantification and distribution, P1 has been surveyed only in more general studies of basic constituent order. Perhaps the first study of this kind is Greenberg 1963,
which included a very limited sample of thirty languages. In this hardly representative set, Greenberg included six P1 languages, namely Berber, Hebrew, Māori, Maasai, Welsh, and Zapotec. (He also included Maya, but analyzed it as SVO, though most varieties of Maya are in fact P1, cf. §6.1.2) The most prominent survey of basic constituent order in the world’s languages is the World Atlas of Language Structures (WALS, wals.info). WALS encoded 1,444 language entries for the dominant constituent order of subject, object and verb. Their count is given in Table (3.1), and a mapping thereof is presented in Figure (3.1):

<table>
<thead>
<tr>
<th>Basic constituent order</th>
<th>Number of languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOV</td>
<td>565</td>
</tr>
<tr>
<td>SVO</td>
<td>488</td>
</tr>
<tr>
<td>VSO</td>
<td>95</td>
</tr>
<tr>
<td>VOS</td>
<td>25</td>
</tr>
<tr>
<td>OVS</td>
<td>11</td>
</tr>
<tr>
<td>OSV</td>
<td>4</td>
</tr>
<tr>
<td>SOV/SVO</td>
<td>29</td>
</tr>
<tr>
<td>VSO/VOS</td>
<td>14</td>
</tr>
<tr>
<td>SVO/VSO</td>
<td>13</td>
</tr>
<tr>
<td>SVO/VOS</td>
<td>8</td>
</tr>
<tr>
<td>SOV/OVS</td>
<td>3</td>
</tr>
<tr>
<td>No dominant order</td>
<td>189</td>
</tr>
</tbody>
</table>

Table 3.1: WALS Features 81A and 81B: Order of Subject, Object and Verb (1,444 entries)

WALS finds that 9% (134 out of 1,444) of the languages in the sample canonically place the verb at the beginning of the clause (with 6.6% VSO, 1.7% VOS and 1% either VSO or VOS). However, it is questionable whether this sample is representative of the world’s languages. The sample size is 1,444 out of a total of 7,688 registered languages on Glottolog, which represents a slender 18.7%. Also, the WALS sample is neither statistically random nor representative of the world’s phyla. As a case in point, which also served as a motivation to conduct the present research in the first place, I point out that the number of P1 languages of the Philippines alone surpasses the total count of P1 languages in the WALS sample, with 186 languages registered by Glottolog (cf. §6.2.1). Nonetheless, the WALS study has until now represented the most comprehensive survey of P1.

The emergent picture from WALS is that P1 is not evenly distributed across the world but occurs in certain linguistic areas: most languages with VSO or VOS basic constituent order cluster on Pacific Islands, the Pacific Northwest coast of America, Mesoamerica, and East Africa. Isolated languages are also found on the British Isles, Northwest Africa, and the northern half of South America. The unprecedented breadth of the WALS study notwithstanding, WALS has some limitations beside its sample size and possible data gaps: it does not consider alternative constituent orders, but only basic constituent order, and it is unclear whether the distribution of P1 across areas and language families is represented in proportion (or whether the sample is skewed toward certain areas or families). Still, for
lack of comparable alternatives, WALS serves as the standard reference for quantifications of constituent order (among many other features that are also encoded on WALS). As the hitherto most complete survey of P1 on a global scale, WALS also served as an immensely useful starting point for the present survey.
Chapter 4: The present survey

Unlike previous sample-based estimates on the prevalence of predicate initiality in the world's languages, this present study is the first to attempt a (reasonably) exhaustive count and discussion of the world's P1 languages. This approach has the methodological consequence that this study does not have a clearly-bound set of languages, a “sample” of the world’s languages. Rather, the “set” consists of all the world’s languages. Needless to say, I do not have single data points on every single language of the world, but based on a patchwork of previous typological, areal, and phylogenetic reference studies, I have been able to focus and zoom in on the few areas with alleged occurrence of P1, while excluding many areas and language families of the world, thus creating a big “remainder” category of languages, labeled “unknown/unlikely” that consists of languages on which I either have not found enough data or which fall in regions that have been described to be devoid of P1 in previous studies (cf. §§ 4.2.1–4.2.9). While it would of course be desirable to distinguish these data types, and to have data points for every single language, this would represent a huge effort that goes beyond the scope of the present study. As a result, the data-set consists mostly of positive data. Gaps in such a comprehensive atlas are ineluctable, partly due to lacunae in language documentation but also due to the sheer scope of the project, especially for a single researcher. Still, the present survey represents not only the most fine-grained overview of the distribution of P1 in the world, but it also offers the most detailed survey of P1 for many of the families covered.

A typological study—or perhaps any scientific endeavour—always faces the challenge of finding the right balance between faithful representation of single data points (constituent order in single languages in this case) and their distillation into a manageable set of meaningful categories. Leaning too much in the direction of faithfulness first requires exuberant hours of literary review, really in-depth knowledge of every single language, which is unrealistic for such a large study, but lack of a certain level of generalization also complicates comparison across language families. On the other hand, over-application of generalization may neglect relevant differences between languages, and the categories may not provide a good fit for the data.

Some linguistic structures are categorical without ambiguity or overlap, some are even binary. Constituent order, with which the current study is concerned, is not one of these structures. Languages generally have multiple ordering options, and oftentimes these can be perused for pragmatic effects, e.g. to contrast one participant with another. This study deals with languages where the predicate can be placed in clause-initial position without
any special pragmatic effect, in a pragmatically neutral statement. To this end, a set of parameters has been defined in (2.2) in Chapter 2 to identify such languages. En bref, a language is included as P1 in this survey if it can use P1 in a declarative transitive clause with the arguments expressed as full noun phrases and without particular focus on any participant. It is important to bear in mind that that definition is arbitrary to some degree: some may consider it too wide, thus including too many languages of different structures, while others consider it too narrow, thus not including relevant languages or only considering a very specific structure missing greater commonalities. Indeed, I have received both comments from scholars, which should suggest that this study fares somewhere in the middle.

Section §4.1 describes the process of data collection under the parameters defined in (2.2), while Section §4.2 gives an overview of the results obtained from that process.

### 4.1 Methodology and Sources

This goal of this survey is to produce a reasonably comprehensive atlas of predicate initiality in the world’s languages. Included are languages that are currently spoken or were still spoken in the 20th century (though cf. Chapters 6 and 7 for a discussion of historically attested languages as well). One would think that such a project requires surveying the entirety of the 7000+ languages spoken in our world, which would be a daunting if not impossible task, especially for a single researcher. However, this project is feasible thanks to (a) previous studies in this field, in particular the World Atlas of Language Structures (cf. Chapter 3), (b) the concentrated distribution of P1 in certain areas and families, (c) the recent publication of typological handbooks covering many major linguistic areas of the world, and finally (d) the availability of electronic searchable language resources.

The present survey relies on large-scale handbooks on the linguistic areas and major language families of the world as a point of departure. The series by Mouton de Gruyter turned out to be especially useful for the present study. It has become standard for such works to deal with typologies of constituent order only in the 21st century. Thus, a study of this scale would have been extremely difficult and labor-intensive, if not impossible, just a couple of decades ago.

Starting with major handbooks, I investigated each language that has been reported to exhibit predicate initiality, as well as its closest relatives and immediate neighbors. I was able to exclude entire regions, such as Siberia, from investigation thanks to such typological handbooks. While the identification of basic constituent order in a given language can be a less-than-evident and, in some cases, controversial task, P1 is typologically such a marked structure that it hardly goes unnoticed and unreported in these handbooks. Typological outliers are readily identified, making it unlikely that a P1 isolate in a typologically distinct area goes unnoticed. The same is generally true at

---

1 Excluding languages that became extinct during the course of the 20th century would lead to the omission of many languages of the American Pacific Northwest, which is one of the hotspots of P1.
the level of single languages: even if P1 is not identified as the basic constituent order, its occurrence will usually be mentioned in grammars if present. However, in areas of P1 dominance, it may occur that especially older sources do not bother to mention the globally unusual constituent order, as experts in the field would consider than an obvious given. I also complemented my literature research with direct consultation of typologists and fieldworkers working in various areas, also to obtain a more insightful characterization of the languages in question.

Luckily, electronic versions of many publications, recent and centuries-old, accelerate the literature review immeasurably. Although I began my literature review with physical books, working my way through shelves of languages in P1-dominant areas such as Polynesia, Mesoamerica, and Northern Africa, starting at the University of Hawai‘i at Mānoa, continuing at Harvard University, and finally at the University of Zürich, this process took a sudden turn with the onset of the Covid-19 pandemic in March of 2020. After that, the literature review shifted almost entirely to online resources. Working with electronic books has the tremendous advantage that they are searchable. Hoping that my experience in data mining will be of use to other researchers in the field of language typology, I shall describe the exact procedure that proved to be the most efficient for the present survey. The present survey relies on two pieces of data, explicit statements of language experts on constituent order, and example sentences that demonstrate P1 under the parameters chosen for this study (cf. (2.2) in Chapter 2). The rote course of actions to extract the former is presented as (4.1), and for the latter as (4.2).

(4.1) Procedure to efficiently extract descriptions on constituent order in searchable publications.

1. Review the table of contents (in older, and especially in French publications, often at the end) for sections on word/constituent order in particular, or general sentence-level syntax / clause structure.

2. If 1) yields no results, search the document for “order” (Spanish orden, Portuguese ordem, French ordre). Add a space in front of the English term to avoid results with “border”. In Romance languages the term for constituent order is usually followed by the preposition de (Portuguese dos/das), so it can be included in the search for more relevant search results. The German equivalent for “word order” is Satzstellung/Wortstellung, hence the relevant search term is “stellung”. The Chinese phrase is ��. The complete Malay phrase for “basic word order” is pola dasar kalimat/ayat, though a search for “pola” alone will usually yield results.

3. An additional search for the grammatical labels “VSO”, “VOS” (add the initial space especially in publications written in Spanish and Portuguese to avoid words with that ending), “VAO”, “SVO” etc., “subject”, “object”, “transitive” (add the initial space to exclude results for “intransitive”), “predicate”, “structure” (and in French also “énoncé”) often leads to relevant statements.
(4.2) Procedure to extract example sentences with full NP arguments.

1. In many cases, a grammar provides adequate example sentences demonstrating the basic constituent order of a language; however, the examples provided are often intransitive or include pronominal arguments, which are not necessarily indicative of a language’s transitive-clause syntax (as pronouns frequently cliticize, cf. Chapter 2). Subjects are more often topical and thus replaced by pronouns than objects, so it is generally easier to encounter an NP object than an NP subject. If the example sentences provided in the syntax section are inadequate, follow the subsequent steps.

2. In most publications, the translations of example sentences are flagged by single or double quotation marks. Thus, for publication written in the major European languages (but not, e.g., Russian, Chinese or Japanese), example sentences with a definite NP subject can be found searching for the string of quotation mark and definite article, e.g. “The ” in English, or “Le”, “La ” in French, including a space after the article.

3. If 2) does not yield results, search for the string of quotation mark plus personal pronoun (e.g. "my ") or demonstrative pronoun (e.g. "this ").

4. Many grammars contain narratives toward the end of the book. These can be browsed for example sentences, although most examples will contain pronominal arguments, and it has to be considered that many may not represent broad focus. Narratives usually also contain personal names, which can be searched for in the entire document (and in Spanish publications a search for the most common names Juan, José, María is often successful). Typical protagonists in mythological narratives, such as Raven, Coyote, or Hyena, are also worth searching for.

5. Finally, if none of the above yielded results, search for prototypical agents, such as "the man/men/woman/women/children/dog(s)", prototypical patients, such as typical food items, or for prototypical transitive verbs, such as "hit, ate, bit, cooked, killed" in the respective matrix language.

6. For all example sentences, read the context to ensure that the sentence does not represent a syntactically or pragmatically marked structure.

With the above-described procedures, the relevant information can usually be extracted from any grammar within twenty minutes, or sometimes as little as five minutes. This depends on the format of the publication, with complicating factors being outdated theoretical frameworks (the publications of the 1960s and 70s in a tagmemic approach appearing particularly opaque), lack of glossing (mostly pre-1990s), and neglect of syntax in grammars (not uncommon prior to the 1980s, when this changed in part thanks to the influence of Greenberg’s work). It is of course possible that the procedure applied here misses some relevant information, which could be gleaned by attentive reading rather than searching for keywords, but the time investment for the former would make a global
survey utterly impossible (at least for a dissertation). I have found that additional effort through more reading and browsing of examples is seldom successful. However, older publications often require such a less structured, more tedious process; although they may be digitized and searchable, the lack of standardized grammatical terminology, glosses and flagging of example sentences with quotation marks makes the above procedure largely (though not entirely) inapplicable.

Many languages do not have complete grammars or studies dedicated to clause syntax. However, publications on minority languages that are concerned with unrelated topics, such as phonology, often contain some syntactic description or useful example sentences. Dictionaries in particular often contain sketch grammars, even though the title of a publication would not indicate that. Therefore, in the absence of whole grammars or syntax papers, it is worth consulting other resources on a target language using the above procedure for a quick and efficient survey.

It should be borne in mind that this survey, as most typological surveys, relies on the analysis of other studies. Although this study goes further than just adopting a language’s alleged basic constituent order by considering alternative pragmatically unmarked ordering possibilities, it still relies on the documentation and assessment of other authors. Ideally, specific studies on constituent orders were consulted for a particular language. However, since such are not available for most languages, grammars, or just grammar sketches, with more or less elaborate sections on constituent order were used. Thus, all the languages in this survey are not based on the same quality and depth of research. For maximal comparability, a single research project would investigate all of the world’s languages under the same methodology. Beside the fact that such a project is utterly unachievable in scope, one could expect that the researchers’ level of expertise in any single language is superficial at best. Thus, working with more holistic primary publications has the advantage that the descriptions were authored by researchers who are experts on the languages in question, although their methodologies may be varied. However, author’s assessment of a language’s constituent order was not taken at face value but was tested against primary language data. Only if an assessment was confirmed by sentences fulfilling the criteria under (2.2) in Chapter 2, was a language included in the database as a “P1 language.” Due to scarcity of primary data for most languages, this standard could only be applied to part of the languages that I expect make use of P1. Many more languages are described as following P1, but this could not be confirmed with primary data. In cases where these alleged P1 languages are closely related to confirmed P1 languages, they were included in the database as “suspected” P1 languages, marked with the label [susp], as opposed to [conf] for confirmed P1 languages, and [neg] for languages that do not make use of P1 or only do so under very specific pragmatic conditions.

Where available, I consulted multiple sources on the same language. In most cases, the assessment on basic constituent order made in one publication was confirmed in others. However, in not too few cases, different authors have different analyses on this topic. This is understandable and expectable, as constituent order usually allows for a certain level of flexibility, as opposed to other syntagms such as phrase-internal word order. However,
disagreement between different authors does not pose a major complication for this study, as this work is not concerned with the identification of a single basic constituent order. If one author defines one possible constituent order as basic, while another author argues for preference of another order, that is probably because both orders are common. Where publications elaborate on their analysis and constituent order alternatives—which is not always the case—we can gain an understanding of the factors that lead to one order or the other, of relative frequencies, of pragmatic and structural markedness. However, where no such elaboration is given, and especially where only one publication on a language’s constituent order is available, we cannot be sure whether P1 is not also available as an alternative to the author’s chosen basic constituent order.

Another limitation of this study stems from the fact that constituent order analysis often relies on elicited data: because clauses with two overt NP arguments are usually rare in natural discourse, elicitation is commonly used as a more economical alternative to compiling extensive texts. This is a potential problem as constituent order is quite susceptible to influence from contact languages (cf. Chapter 2). However, circumventing this limitation would require access to extensive corpora of glossed texts, which is unfortunately only available for a small minority of the languages of the world. Thus, this limitation does not concern this study in particular, but research on constituent order in general.

A final complication of this study concerns only a minority of languages but shall nonetheless be stated here because it challenges the general understanding of constituent order in general: In some languages, the respective ordering of P and S, and that of P and O, does not match the ordering of all arguments (P, S, and O) in a single clause. An example of this is described in detail in Benito Pérez 2016 on Poqomam Maya: when only one argument is expressed by a noun phrase, it follows the predicate in the vast majority of cases; however, when two arguments are both expressed as noun phrases in the same clause, the subject occurs preferentially before the predicate. For the purposes of this survey, Poqomam Maya qualifies as a P1 language.

4.2 Numerical and areal overview of results

The results of this survey are presented from two different perspectives. First, this section presents a general overview of the results and presents them from an areal perspective. Chapter 6 then discusses P1 from a phylogenetic and historical perspective by investigating the syntactic history of the language phyla where P1 is attested.

This study identified 396 languages confirmed for P1 (“confirmed”), and another 415 languages with reasonable expectation of P1 (“suspected”), amounting to a combined 811 languages. Given this concrete number of P1 languages, we can venture a fairly accurate estimate of P1 in the languages of the world by dividing the number of languages included in this study by a total count of languages. As is well known, the views on the number of

\(^2\)Bear in mind that the actual number may be slightly lower, as not all of the 415 expected P1 languages will confirm the expectation.
languages in the world can diverge vastly, with commonly cited figures ranging from six to seven thousand. Since this study is based on Glottocodes, we can compare this study to the total number of languages recognized by Glottolog, which issues the Glottocodes.\(^3\) Again, it is worth mentioning here that Glottolog falls on the “split-friendly” end of the spectrum, counting 7688 entries (as of March 2022). This means that Glottolog generally counts more languages in a given family or area than most experts on those families and areas would, resulting to inflated numbers. One does not need to agree with this figure or consider all of those entries as separate languages to accept Glottolog’s aptitude for comparison with the present study. However, Glottolog’s total count includes 940 extinct languages, some of which have, like e.g. Latin, ceased to be spoken for centuries. Thus, Glottolog’s total count of spoken languages is 6748. As outlined in §4.1, the present study includes not only spoken languages, but also languages that have still been spoken in the 20th century. The reason for this decision is because excluding all the recently extinct languages, some of which are being revived, would neglect many well-documented languages, as well as a great deal of the diversity in the Americas, precisely in some P1 hotspots such as the Pacific Northwest. It would be quite time-consuming to assess how many of the 940 extinct languages in Glottolog have only ceased to be spoken in the 20th century. Considering that the language extinction rate accelerated significantly in the 20th century, and that the likelihood of languages disappearing without documentation has significantly decreased in the modern era, we can safely assume that a large part of those 940 languages, probably the majority, was still spoken in the 20th century. Since assessing that exact proportion would be worthy of a thesis in itself, I shall simply estimate here that two thirds of those 940 extinct languages only became extinct in the 20th century, which means that the actual total number of languages to be compared with the count of P1 in this study is 7375. Accordingly, we arrive at a percentage of P1 within the set of languages with a Glottocode spoken in the 20th century of 11%. This is slightly higher than the 8.7% of WALS (120 languages out of a sample of 1376 languages, cf. §3). Although the two figures are quite comparable, which speaks for the reproducibility of this study, the difference also requires some explanation. First, this discrepancy lies within the range of accuracy a study based on a language sample like WALS can realistically reach, but it is certainly also due to a difference in definition and criteria for inclusion: the WALS study is concerned with basic constituent order, while the present study includes pragmatically unmarked structures even if they are not identified as a language’s basic constituent order (cf. Chapter 2). As a result, some languages identified in WALS as P1 are not included in the present study, such as Wathawurrung, Wembawemba, Wolio or Koregaje (due to a lack of data), but more often, languages that are not recognized as P1 in WALS are included in this study, such as Amis, Paiwan, Breton or Modern Greek. Of course, many languages in this study are not included in the WALS sample, at all, such as the Enlhet-Enenlhet languages, Nadèb, Miluk, or the Western Berber languages.

Map (4.1) shows predicate initiality for confirmed P1 languages (dark green) and

\(^3\)Glottolog also includes Glottocodes for dialects and language families and sub-groups, but the only relevant Glottocodes in this study are those referring to Glottolog’s “language” level.
Numerical and areal overview of results

4 It would of course be desirable to have separate data entries for all languages in the world, but such a project would be more appropriate for a research team, rather than a single person. For the sake of transparency, it should also be noted that Map (4.1) and all other maps in this study include Glottolog’s extinct languages in the rest category of “unknown/unlikely.” The inclusion of these is merely a technical result of data sampling and mapping, but they are not included in the statistics of this study, and their inclusion does not have a visible effect on the maps.
Mamoré River in the Bolivian Amazon, in the Grand Chaco, the Southern Andaman Sea, in Madagascar, and on the British Isles. By contrast, P1 is virtually absent from all of Eurasia (except for the Celtic languages on the British Isles, Greek, and a few Austroasiatic languages), Australia, Southern and Western Africa, all of North America except the West coast, the Amazon (besides the small clusters mentioned above), and Patagonia. On Map (4.2) the major P1 hotspots are marked with polygons, and the minor P1 clusters with circles.

Figure 4.2: Major P1 hotspots (polygons) and smaller P1 cluster (circles)

This distribution is somewhat surprising, as one might expect some correlation between occurrence of P1 and areas of high linguistic diversity, in other words one would expect areas with more languages to also feature more P1. However, P1 is absent from some of the most linguistically diverse areas of the world: West Africa, South and Southeast Asia, and New Guinea.

The following subsections provide a closer look into the distribution of P1 by geographic areas.

4.2.1 Pacific Islands

The highest number of languages with P1 constituent order is in the Pacific region: at least 350 P1 languages have been identified here. Almost all of the P1 languages in this area belong to the vast Austronesian language family. However, most Austronesian languages are not P1: the numerous languages of Indonesia mostly follow SPO, while many languages around New Guinea follow SOP. P1 is dominant in Polynesia, the Philippines and Taiwan, so dominant that there are only few exceptions in these areas. P1
is also common, though less dominant in Sabah (bordering the Philippines), and Eastern Melanesia (bordering Polynesia). For a more detailed description of the distribution of P1 in the Austronesian family, see §6.2.1.

By contrast, the roughly 800 non-Austronesian languages of the New Guinea area, which do not form a genetic unit but represent a number of independent families, are predominantly predicate-final (Reesink and Dunn 2018: 950). According to Foley, the sole non-Austronesian language in this area with P1 is the Kuot, an isolate spoken on New Ireland (Foley 2000: 383, Foley 2018: 896).

### 4.2.2 Australia

The aboriginal languages of Australia are renowned for their very flexible constituent order (Dixon 2002: 78). Not only can constituents be arranged in any order, but the constituents themselves can be discontinuous, quite similar to the syntax of more widely known Latin. This was first described for Warlpiri by Ken Hale (1983), under the label non-configurationality. Despite this generally liberty in constituent order, some languages may still show a statistical preference for a particular order, although it is often unclear whether this preference is syntactically motivated or merely the result of more general pragmatic principles. In the majority of cases where one particular constituent order is described as preferred, that order is SOP (cf. WALS). However, a couple Pama-Nyungan languages of Victoria in Australia’s far south are described as having a basic POS order. The problem
with Victorian languages is that they were all strongly endangered or even moribund by the early 20th century, and they are all extinct now. The description of these languages is generally rather scarce, to the point that according to Blake and Reid 1998, there is only one reliable source on Victorian languages, namely Hercus 1986, which is based on research with the last speakers of Victorian languages in the 1960s. Hercus reports that Wemba-Wemba was the only language “sufficiently well preserved for an attempt at detailed grammatical study” (Hercus 1986: v), and he describes it as following a basic POS order (Hercus 1986: 51). And indeed, POS order is recurrent in the examples provided by Hercus, but they are almost all intransitive (stative or active) or with pronominal subjects (suffixes on the verb). The one example with a transitive nominal subject is indefinite and does thus not meet the criteria for the current study. (Another example with a definite nominal subject appears in the song of the owlet-nightjar, but here the constituent order is OPS.) The same pattern is apparent in Hercus’s Wergaia and Madhi-Madhi data, all treated as dialects of the same Western Victoria Kulin variety by Glottolog, in addition to several examples of SPO. Furthermore, WALS classifies Wathawurrung as a VOS language based on an impressionistic secondary appraisal by Blake and Reid 1998, although they indeed provide one two examples of post-predicate subject nominal phrases. In the absence of other sources on Wemba-Wemba or other Victorian languages, and without any indication in the literature that other Australian aboriginal languages have a preference for P1, Australia is considered free of P1 for the purposes of this study, although it seems quite possible that at least the extinct members Kulin subgroup had a true preference for POS.

4.2.3 North America

Though separated from the P1 languages of Polynesia by vast stretches of open ocean, two adjacent areas in the greater Pacific region are further major hotspots of predicate initiality: the Pacific Northwest, and Mesoamerica (which is treated separately in the next subsection). In the Pacific Northwest, the families with P1 are the Tsimshian, Wakashan, Salish, Chimakuan (of which only data on Quileute or available), Chinookan, Coosan, and Sahaptian families (excluding Nez Percé), as well as the isolates Alsea and Siuslaw, stretching from the coastal Alaska-British Columbia border in the north down to central Oregon in the south. Though the P1 languages in this area accumulate on the coast, they stretch as far inland as western Montana with the interior members of the Salish family, which is consistently P1.

To the north, the P1 area of the Pacific Northwest is bound by Tlingit, Eyak, and the Athabaskan languages, which are all predominantly predicate-final. To the south, the area gives way to languages with free constituent order in Central Oregon, and increasingly
Numerical and areal overview of results

Figure 4.4: Distribution of predicate initiality in North America

However, there is a smaller cluster of P1 in Central California comprising the Chumash languages around Santa Barbara, Yokutsan in the San Joaquin Valley, and the now extinct Salinan isolate indigenous to Monterey County.

As Table (4.1) demonstrates, the West Coast of North America is linguistically highly diverse with numerous independent phyla (although there have been attempts to assign these to larger groups, among which the Penutian and Hokan hypotheses are the best known, although both are contended). The interior and East Coast of America are far less diverse, and P1 is not a phenomenon there. The syntactic profile of the rest of the continent (see Table (4.2)) matches that of California (minus the P1 cluster in the center): constituent order tends to be quite free, but where there is a preferred order, that is commonly predicate-final.

For clarification it should be restated that outside of the Pacific Northwest, many other languages in North America in fact do allow P1 order, but only because their syntax allows for any type of constellation, with discourse pragmatics defining the order. Among the many languages with free constituent order, the Algonquian languages deserve particular attention: because pragmatic factors can lead to a relative frequency of P1, some authors have argued that P1 should be considered the basic constituent order in Algonquian (most notably Tomlin and Rhodes 1979 for Ottawa, and Dahlstrom 1995 for Fox and Algonquian in general, but cf. also Reinholzt and Wolfart 1996 on Plains Cree, Junker 2004 on East Cree, Drapeau 2014: 296–7 for Montaignais). The syntax of Algonquian languages in general is quite comparable: Algonquian languages track their referents overtly on both verbs and nouns, in a system that contrasts topical referents (proximate)
and others (obviate). Because of this explicit marking, the functional load on constituent order is minimal. Accordingly, Algonquian languages are most often described as having “free word order” or belonging to the “non-configurational type” (cf. e.g. Dove 1996 on Fox, Denzer-King 2010 for Siksiká/Blackfoot). Not only are almost all permutations of a verb and its arguments grammatical in this language family, but the arguments (and sometimes the verb itself) may also be syntactically discontinuous (Dahlstrom 1987). That is, “part of an argument may appear before the verb and the remainder after the verb.” (Dahlstrom 1995: 1) Of course constituent order is not just random, but it is employed to signal information status. Several publications on various Algonquian languages mention that newly introduced referents usually enter the narrative as clause-initial nominals. Generally, the clause-initial position is reserved for new or contrastive information, and old information follows. In other words, constituent order is not determined by syntactic relations, but rather by pragmatics. Ordering tendencies are nonetheless observed for single languages. Thus, Rhodes asserts that in Ojibwe dialects, the predicate is most often positioned first in the clause, though it also commonly occurs medially and, less commonly, clause-finally (Tomlin and Rhodes 1992). Similarly, Drapeau describes Innu (Montaignais) constituent order to be verb-initial, when all arguments of a clause are discourse-given, though newly introduced participants are fronted to the initial position, regardless of their syntactic function (Drapeau 2014: 296–7). Dahlstrom (1995) goes as far as to claim that all pre-predicate arguments fulfill the function of topic or focus, while the unmarked position for arguments is after the predicate, and although Dahlstrom’s study is based on Meskwaki (Fox), she ventures to suggest that his analysis “may be applied to Algonquian in general” (Dahlstrom 1995: 2). That is, according to Dahlstrom’s analysis, P1 should be considered the basic constituent order for Algonquian languages.

However, although Dahlstrom’s phrasing suggests that P1 is unmarked and pragmatically neutral in Algonquian, other studies (Dove 1996: 59) describe this order to imply that the arguments are discourse-given. Arguably, there is no truly pragmatically neutral (“basic”) constituent order in Algonquian at all, as the function of constituent order is precisely to impart pragmatic information. As outlined in Chapter 2, such languages with “free” or pragmatically defined order are not included in this study, to set them apart from languages like Salish or Mayan, in which P1 is syntactically defined and pragmatically neutral in that it does not imply anything about the discourse status of the arguments, and in which deviations from P1 are overtly marked syntactically, but also obviously so pragmatically. Furthermore, although that may simply be the result of a linguistic tradition, descriptions of constituent order in linguistic publications and didactic materials usually list all possible permutations without giving preference to a single one, very unlike such materials on Salish languages, for instance. In short, Algonquian (as well as Ktunaxa/Kutenai and many other languages of North America) is not included in this catalogue.

---

1 The two extra-Algonquian Algic languages Yurok and Wiyot follow the order SPO (cf. Robins 1958 for Yurok, and Reichard 1925 and Teeter 1964: 48 for Wiyot).

2 Quite the same argument was applied to Ktunaxa/Kutenai (Morgan 1992: 393–4).

3 Anderson (1996: 2) describes the analysis of a basic PSO constituent order for Fox as “misguided.”
because its constituent order is defined by pragmatics, rather than by syntax.
### Table 4.1: Non-P1 languages of the North American West Coast.

<table>
<thead>
<tr>
<th>Language family</th>
<th>Constituent order</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tlingit-Eyak-Athabaskan(^a) (Na-Dene)</td>
<td>SOP</td>
<td>Hargus 2013: 16, Leer 1991: 24, Dryer 1985</td>
</tr>
<tr>
<td>Haida</td>
<td>SOP</td>
<td>Enrico 2003: 45</td>
</tr>
<tr>
<td>Kalapuya</td>
<td>free (frequent SPO, employed especially in ambiguous contexts)</td>
<td>Rude 1987: 428–9, Banks 2007: 8</td>
</tr>
<tr>
<td>Takelma</td>
<td>SOP (used to disambiguate, also frequent OPS and occasional SPO, only very rare P1)</td>
<td>Kendall 1997: 12–4</td>
</tr>
<tr>
<td>Karuk</td>
<td>free</td>
<td>Bright 1957: 141</td>
</tr>
<tr>
<td>Chimariko</td>
<td>P-final</td>
<td>Dixon 1910: 335, Jany 2009: 141</td>
</tr>
<tr>
<td>Wintuan</td>
<td>free (SOP)</td>
<td>Pitkin 1963: 36, Shepherd 1997:35</td>
</tr>
<tr>
<td>Shastan</td>
<td>free</td>
<td>Silver 1966: 99:221</td>
</tr>
<tr>
<td>Palaihnihan</td>
<td>free(^b)</td>
<td>de Angulo and Freeland 1930: 8</td>
</tr>
<tr>
<td>Washo</td>
<td>SOP (consistently head-final)</td>
<td>Jacobsen 1964: 399–404</td>
</tr>
<tr>
<td>Yuki-Wappo</td>
<td>SOP</td>
<td>Kroeber 1911: 372, Li et al. 1977: 86</td>
</tr>
</tbody>
</table>

\(^a\) Most Athabaskan languages are actually spoken inland, not along the coast.

\(^b\) Nevin (1998: 47) suggests a basic VSO order, but this is not corroborated by any data.
Table 4.2: Non-P1 languages of interior and Eastern North America.

<table>
<thead>
<tr>
<th>Language family</th>
<th>Constituent order</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algic</td>
<td>free (Algonquian)</td>
<td>Dove 1996: 59</td>
</tr>
<tr>
<td></td>
<td>(Yurok, Wiyot SPO)</td>
<td>Robins 1958, Reichard 1925, Teeter 1964: 48</td>
</tr>
<tr>
<td>Zuni</td>
<td>SOP (strictly P-final)</td>
<td>Nichols 1997: 6</td>
</tr>
<tr>
<td>Keres</td>
<td>free</td>
<td>Lachler 2006: 302–10</td>
</tr>
<tr>
<td>Caddoan</td>
<td>OP (SOP / OPS)</td>
<td>Rood 1996: 559</td>
</tr>
<tr>
<td>Yuchi</td>
<td>SOP</td>
<td>Linn 2001: 449</td>
</tr>
<tr>
<td>Natchez</td>
<td>SOP</td>
<td>Kimball 2005: 402</td>
</tr>
<tr>
<td>Tonkawa</td>
<td>free</td>
<td>Hoijer 2018: 45</td>
</tr>
<tr>
<td>Tunica</td>
<td>SOP</td>
<td>Haas 1941: 98–9</td>
</tr>
<tr>
<td>Atakapa</td>
<td>P-final</td>
<td>Swanton 1929: 145</td>
</tr>
<tr>
<td>Chitimacha</td>
<td>SOP</td>
<td>Granberry 2004: 100, Hieber 2019: 319</td>
</tr>
</tbody>
</table>

*a* But see discussion in this section.
4.2.4 Mesoamerica

Mesoamerica represents one of the major P1 hotspots in the world. This area encompasses the southernmost Uto-Aztecan branch Nahua (formerly Aztec), and the Totonacan, Otomanguean, Mayan and Xinkan families, as well as at least a few Mixe-Zoquean languages. These languages are mostly clustered in Central-Southern Mexico and Guatemala (although the southernmost Nahua variety Nawat, also known as Pipil, is spoken in El Salvador). P1 in this part of the world stretches slightly beyond the commonly recognized confines of the Mesoamerican Linguistic Area (cf. Campbell et al. 1986) with the Uto-Aztecan Cora and Tepehuán branch in northern Mexico, and the Arawakan outlier Garifuna in Honduras.

P1 is so common in Mesoamerica that it can be considered the areal default. The three major language groups of the region, namely Nahua, Otomanguean and Mayan, are quite consistently P1, with a few exceptions addressed in the chapters on the respective families in Chapter 6. Of these three groups, the Otomanguean is by far the largest in terms of languages, with Glottolog listing 181 varieties. The Mayan family is considerably smaller, with 33 languages listed. The Nahua languages, with 31 listed in Glottolog, do not form an independent family, but are rather part of the vast Uto-Aztecan family, which stretches north as far as Idaho and Wyoming. As further described and discussed in §6.2.4, the northern members of this family do not exhibit P1 but are rather predominantly predicate-final. Totonacan, a smaller independent language family with six independent languages (13 Glottolog entries) native to northwest Veracruz, northern Puebla and southern Hidalgo in Mexico, is also P1-dominant. The southern border of the Mesoamerican P1 area form the five Xinkan languages in southern Guatemala, which are now unfortunately all
Numerical and areal overview of results

extinct. Breaking the consistency of P1 in Mexico are three independent linguistic phyla, Cuitlatec, Huave (listed with four entries in Glottolog) and Tequistlatecan (listed as three languages, cf. Turner 1966: 37, Suárez 1983: 95), which all exhibit SPO order (or more precisely Ergative-Verb-Absolutive for Huave). The medium-sized Mixe-Zoque family, though centrally located within the Mesoamerican P1 area, has been the subject of some contention (cf. §6.2.8), as only a small number of them have been described to follow P1 order, but most sources point toward a dominance of either SPO or SOP. Indeed, south of the Mayan and Xinkan families (and thus south of Mexico and Guatemala), P1 gives way to a block of predominantly predicate-final languages, starting in Honduras with Jicaquean (Holt 1999: 7) and Lencan (King 2017), continuing with Misumalpan in Nicaragua (Suárez 1983: 95 for Miskito, Norwood 1997: 110 for Mayangma/Sumu, Green 1999: 69 for Ulwa), and stretching on in Central America and on into the South American continent with the large Chibchan family (see §4.2.5).

4.2.5 South America

Compared to Mesoamerica, P1 order is far less common in South America. The few dozen languages that do follow P1 in South America mostly cluster along the eastern flank of the Andes, in Peru, Bolivia, and Paraguay. Most notable are the Campan branch of the Arawakan family in Peru, and the small Enlhet-Enenlhet family of Paraguay. It should be noted here that, although not fulfilling the criteria for P1 order in this study, many South American languages show P1 order in intransitive clauses, but not in transitive clauses. For instance, this pattern is the norm for Tupian and Chonan languages, among many others. Unlike Arawakan, which besides Campan features a few more isolated P1 languages, the other major family in the north of the continent, Cariban, exhibits virtually no P1 (with the exceptions of Ikpeng and Panare), but instead follows a “fairly rigid OV sequence” (Derbyshire 1999: 55). Likewise, the two major language families of Brazil, Tupian and Macro-Gê, exhibit virtually no P1 at all (with the odd exception of Tenetehara/Guajarára), but they mostly follow SOP order (Rodrigues 1999b: 114, Rodrigues 1999a: 187). However, Kariri, which used to be included into Macro-Gê, also uses P1. It is quite challenging to obtain a clear picture the Amazon region, first because many languages of the region are only sparsely documented, and second because the whole region represents a patchwork of many small families. Some of these smaller Amazonian families show a preference for P1, namely the Chapacuran (attested mainly through Wari’), as well as the isolates Yagua and Taushiro, both spoken in the Peruvian Amazon, and the isolates Itonama, Cayubaba, and Movima, which are spoken in Bolivia’s Beni Department. A more unified picture is presented along the west coast of the continent, including the Andes, with a solid belt of predicate-final languages, starting with Chibchan, Chocoan, and Barbacoan in the north, adjacent to the vast Quechuan family, then Chicham (Jivaroan), Aymaran, Araucanian, and finally Chonan and Kawesqaran (Alacalufan).

An overview of constituent order in South America is presented in (4.3)–(4.4).

10However, Kim 2008 describes San Francisco del Mar Huave to commonly employ POS beside SPO.
Numerical and areal overview of results

Most languages in the Tukanoan family of the Western Amazon place the predicate in clause-final position. Only one language, Koreguaje, has been described to regularly place the predicate before the core arguments, by Cook and Criswell (1993: 89). However, the authors do not present any conclusive data to support this statement. In the examples with overt nominal subjects, these precede the verb consistently. In fact, the verb often appears in clause-final position. I have not found any conclusive evidence of P1 for Koreguaje in other sources, either. Another language, Cubeo, has been described as allowing PSO constituent order besides the more common SPO and OPS orders by Maxwell and Morse (1999: 139). However, the only sentences cited to represent this order are with pronominal arguments and are thus not relevant for this current survey. Thus, in the absence of relevant examples, P1 cannot be confirmed for Tukanoan.

Martins and Martins (1999: 254) contrast Nadêb with its sisters Dâw and Hupda-Yuhup in that the former has a “strong tendency towards verb-initial order”, whereas the latter feature P-final order. However, they do not support this assessment of Nadêb with confirming data, and in fact other sources on Nadêb disagree, characterizing it as following OSP order with an SPO alternative order (Weir 1984: 24; 268; Epps and Obert 2021: 257). Without confirming P1 data from Nadêb, I do not include it as a P1 language in this study, although later publications may indeed confirm Silvana and Martins’s description.

There is general agreement in the literature (see sources above) about Borôro constituent order. The SOP order also characterizes the neighboring major family of Macro-Jê, with which Borôro was previously grouped. However, I found a conflicting description in Crowell 1979, which argues that the most common orders are SPO and POS. However, since P1 is not mentioned in any other source on the language (including Crowell 1979), I assume that Crowell’s (1979) observation was not confirmed in more recent studies.

Two Matacoan languages are described to make frequent use of P1 in the literature: Nivaclé (Fabre 2016: 376–7) and Iyojwa’ja Chorote (Carol 2014: 116). The latter is not supported by any data, but for Nivaclé, Fabre (2016: 376–7) cites supporting data and even claims that P1 is statistically predominant. However, Campbell et al. (forthcoming) point out that while the order PSO is common in subordinate clauses, it is possible but “extremely rare” in main clauses. Nivaclé is therefore not included as a P1 language in this study. The other languages of the family show predicate-initial syntax only in intransitive clauses (cf. Tovar 1981 and Åke Alvarsson and Claesson 2014 for Wichí). Overall, P1 is rather marginal in Matacoan.

Gualdieri (1998: 89) states that Mocoví makes use of POS and SOP as alternative constituent orders to the basic SPO order, but she adds that SOP is used only in elicited sentences, and she describes POS as a “marked” order (although she does not elaborate under what specific circumstances this order comes to use.)
Figure 4.6: Distribution of predicate initiality in South America
Numerical and areal overview of results

Table 4.3: Constituent order in South America (roughly organized from north to south, families with P1 in **bold**).

<table>
<thead>
<tr>
<th>Language family</th>
<th>Constituent order</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yaruro</td>
<td>SPO</td>
<td>Guerreiro de Pirela 2016: 136</td>
</tr>
<tr>
<td>Guahiban</td>
<td>SOP (but Macaguán SPO)</td>
<td>Queixalós 2000: 37, Kerr and Berg 2018: 14, Buenaventura 1993: 52</td>
</tr>
<tr>
<td>Cariban</td>
<td>generally OV (rigid), but...</td>
<td>Derbyshire 1999: 55</td>
</tr>
<tr>
<td></td>
<td>Ilpeng PSO</td>
<td>Pachéco 2001, Chagas 2013: 49</td>
</tr>
<tr>
<td></td>
<td>Panare OPS, also PSO</td>
<td>Payne and Payne 2013: 19–20</td>
</tr>
<tr>
<td>Arawakan</td>
<td>mixed</td>
<td>cf. §6.2.5</td>
</tr>
<tr>
<td>Timotean</td>
<td>SPO</td>
<td>Arrieta E. 1993: 102, Adelaar and Muysken 2004: 127</td>
</tr>
<tr>
<td>Yanomaman</td>
<td>SOP</td>
<td>Aikhenvald and Dixon 1999: 350</td>
</tr>
<tr>
<td>Tukanoan</td>
<td>OP (SOP/OPS)(^T)</td>
<td>Barnes 1999: 224</td>
</tr>
<tr>
<td>Nadahup</td>
<td>P-final(^T)</td>
<td>Martins and Martins 1999: 254</td>
</tr>
<tr>
<td>Boran</td>
<td>SOP (OSP and SPO also possible)</td>
<td>Seifart 2005: 51</td>
</tr>
<tr>
<td>Chicham (Jivaroan)</td>
<td>P-final</td>
<td>Wise 1999: 329</td>
</tr>
<tr>
<td>Pano</td>
<td>SOP</td>
<td>Loos 1999: 234</td>
</tr>
<tr>
<td>Arawan</td>
<td>mixed: Paumari: Arg-Pred-Pivot, Madi: Pivot-Arg-Pred</td>
<td>Dixon 1999: 305</td>
</tr>
<tr>
<td>Piraña</td>
<td>SOP</td>
<td>Aikhenvald and Dixon 1999: 355</td>
</tr>
<tr>
<td>Kariri</td>
<td>PSO</td>
<td>Rodrigues 1999a: 188</td>
</tr>
<tr>
<td>Cholón</td>
<td>SOP</td>
<td>Alexander-Bakkerus 2005: 147</td>
</tr>
<tr>
<td>Quechuan</td>
<td>SOP</td>
<td>Coler 2014: 574</td>
</tr>
<tr>
<td>Tupian</td>
<td>generally SOP, but some P1</td>
<td>Rodrigues and Cabral 2014: 9, cf. §6.2.11</td>
</tr>
</tbody>
</table>

41
Table 4.4: Constituent order in South America (roughly organized from north to south, families with P1 in **bold**) [continued].

<table>
<thead>
<tr>
<th>Language family</th>
<th>Constituent order</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macro-Jê</td>
<td>SOP</td>
<td>Rodrigues 1999a: 187</td>
</tr>
<tr>
<td>Takanan</td>
<td>SPO</td>
<td>Aikhenvald and Dixon 1999: 366</td>
</tr>
<tr>
<td>Harákmbut</td>
<td>P-final</td>
<td>Wise 1999: 329</td>
</tr>
<tr>
<td>Katukína / Kanamari</td>
<td>Erg-Pred-Abs</td>
<td>Queixalós 2010: 241, Ishy de Magalhães 2018: 217</td>
</tr>
<tr>
<td>Kwaza</td>
<td>SPO/SOP</td>
<td>van der Voort 2004: 91</td>
</tr>
<tr>
<td>Kanoê</td>
<td>SOP (SPO rarer)</td>
<td>Bacelar 2004: 229</td>
</tr>
<tr>
<td>Aymaran</td>
<td>SOP</td>
<td>Coler 2014: 574</td>
</tr>
<tr>
<td>Mosetén / Chimane</td>
<td>SPO</td>
<td>Sakel 2003: 295–6</td>
</tr>
<tr>
<td>Guató</td>
<td>PSO</td>
<td>Palácio 1984, Rodrigues 1999a: 188</td>
</tr>
<tr>
<td>Zamucoan</td>
<td>SPO</td>
<td>Ciucci 2013: 8, Bertinetto 2009: 30</td>
</tr>
<tr>
<td>Enlhet-Enenlhet</td>
<td>P1</td>
<td>Kalisch 2019</td>
</tr>
<tr>
<td>Charruan</td>
<td>SOP</td>
<td>Barros and Jaime 2013: 22</td>
</tr>
<tr>
<td>Araucanian</td>
<td>SPO</td>
<td>Adelaar 1996: 1333</td>
</tr>
<tr>
<td>Chonan</td>
<td>SOP</td>
<td>Garay 2010</td>
</tr>
<tr>
<td>Gününa-Küne (Puelche)</td>
<td>SPO</td>
<td>Orden 2011: 229, Orden 2018: 206</td>
</tr>
<tr>
<td>Yamana/Yagán</td>
<td>SPO/SOP</td>
<td>Adelaar and Muysken 2004: 579</td>
</tr>
</tbody>
</table>
4.2.6 Africa

In Africa, P1 is not uncommon in the north and east of the continent. The African P1 split into two groups. The northern P1 languages, distributed along the northern coast of Africa and across the Saharan desert (but a few also on the northern border of Cameroon and Nigeria, belong to different branches of the Afroasiatic family. The other P1 cluster in Africa is located in the general area of the East African Rift, from northern Tanzania in the south and reaching South Sudan and Ethiopia in the north. The languages in this cluster, in addition to many more languages in Central Africa and the Central Sahara, were grouped into a large Nilo-Saharan family by Greenberg, but the validity of this proposal, at least in its entirety, has been questioned and does not represent the current communis opinio. The present study approaches these languages in the basic families that are universally accepted, even though there may be a good case for a genetic relationship between some of them. The P1 languages in this area thus belong to the Nilotic, Surmic, Kadugli-Krongo, Kuliak groups. The southern end of this region forms the isolate Hadza (a click language). For more detail on P1 in these and the Afroasiatic languages, cf. Chapter 6.

Though included in Map (4.7), Madagascar is culturally and historically very distinct from the African mainland, as its linguistic origins lie in Borneo. Its P1 syntax is due to its Austronesian heritage (cf. §6.2.1).

P1 is unattested in the culturally highly diverse West Africa (Mande and Ijoid languages are strictly predicate-final, cf. Ebermann 2005: 123, Schreiber 2020: 180, Connell
Numerical and areal overview of results

et al. 2012), as well as in the Bantu-dominated southern half of Africa, where constituent order is predominantly SPO (or more specifically, S-Aux-O-P). Some rare exceptions to this areal generalization include the Senufo sub-group of Atlantic-Congo in West Africa, as well as the (non-Bantu) Khoe-Kwadi languages, which both show predicate-final syntax (Güldemann 2018: 150–3). (The other non-Bantu languages of Southern Africa, Kx’a and Tuu, follow the areal norm of SPO.) For an overview of constituent order in the languages of Africa, cf. Tables (4.5)–(4.6). As is evident in that graph, I used Bender 1996 as an extremely helpful starting point for a typological overview of the "Nilo-Saharan" languages; I checked his characterization of the languages against primary sources, where they were available, although I refrained from checking every single language in a group that was not described to feature any P1. My personal assessment differs from Bender’s in a few cases (e.g. Krongo) but generally confirmed it. A good overview of the typology of the languages that some group into the phylum Niger-Congo can be gained from Bendor-Samuel 1989 (especially Williamson’s chapter), and from Güldemann 2018.

The P1 areas in Northern Africa and the East African Rift are not as consistently P1 as e.g. the Pacific Northwest or Mesoamerica. Even the particular language families that include P1 are mixed with respect to constituent order.
Table 4.5: Constituent order in Africa (roughly organized from west to east, families with P1 in **bold**).

<table>
<thead>
<tr>
<th>Language family</th>
<th>Constituent order</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Afroasiatic</strong></td>
<td>mixed:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Semitic SPO/PSO</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Berber mostly P1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chadic SPO/PSO</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cushitic SOP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cf. §6.2.2</td>
<td></td>
</tr>
<tr>
<td>Kru</td>
<td>SPO (but S-Modal-OV)</td>
<td>Marchese 1989: 131</td>
</tr>
<tr>
<td>Bangime</td>
<td>SPO</td>
<td>Heath and Hantgan 2018: 15</td>
</tr>
<tr>
<td>Dogon</td>
<td>SOP (strictly)</td>
<td>Bendor-Samuel et al. 1989: 173</td>
</tr>
<tr>
<td>Songhay</td>
<td>mixed:</td>
<td>Bender 1996: 38</td>
</tr>
<tr>
<td></td>
<td>Northern Songhay SOP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Southern Songhay SOP</td>
<td></td>
</tr>
<tr>
<td>Saharan</td>
<td>SOP</td>
<td>Bender 1996: 39</td>
</tr>
<tr>
<td>Central Sudanic</td>
<td>SPO</td>
<td>Bender 1996: 40</td>
</tr>
<tr>
<td>Daju</td>
<td>most SPO, but Sila SOP</td>
<td>Tucker and Bryan 1966, Bender 1996: 39</td>
</tr>
<tr>
<td>Maban</td>
<td>SOP</td>
<td>Bender 1996: 39</td>
</tr>
<tr>
<td>Taman</td>
<td>SOP</td>
<td>Bender 1996: 39</td>
</tr>
<tr>
<td><strong>Nilotic</strong></td>
<td>mixed:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>West Nilotic SPO</td>
<td>Hieda 1991: 100</td>
</tr>
<tr>
<td></td>
<td>with exceptions</td>
<td>cf. §6.2.3</td>
</tr>
<tr>
<td>Katla-Tima</td>
<td>SPO</td>
<td>Hellwig 2019: 516–8, Nüsslle 2018: 281</td>
</tr>
<tr>
<td></td>
<td>but Tima free</td>
<td>Schneider-Blum and Hellwig 2018, Dimmendaal 2009: 342–3, Dimmendaal 2010: 234</td>
</tr>
<tr>
<td></td>
<td>cf. §6.1.7</td>
<td></td>
</tr>
<tr>
<td>Julud</td>
<td>SPO</td>
<td>Nüsslle 2018: 281</td>
</tr>
<tr>
<td>Temein</td>
<td>SPO</td>
<td>Bender 1996: 39</td>
</tr>
<tr>
<td>Nyimang</td>
<td>SOP</td>
<td>Bender 1996: 39</td>
</tr>
</tbody>
</table>
Table 4.6: Constituent order in Africa (roughly organized from west to east, families with P1 in **bold**) [continued].

<table>
<thead>
<tr>
<th>Language family</th>
<th>Constituent order</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nubian</td>
<td>SOP</td>
<td>Heine et al. 1966:39, Bender 1996: 39</td>
</tr>
<tr>
<td>Lafot (Tegem-Amira)</td>
<td>SPO / SOP</td>
<td>Güldemann 2018: 153</td>
</tr>
<tr>
<td>Eastern Jebel</td>
<td>SPO</td>
<td>Stirtz 2012: 282, Bender 1996: 39</td>
</tr>
<tr>
<td>Koman</td>
<td>SPO (or V2)</td>
<td>Bender 1994, Killian 2015: 218ff., Otero 2019: 93</td>
</tr>
<tr>
<td><strong>Surmic</strong></td>
<td>mixed:</td>
<td>Bender 1996: 39, cf. §6.1.4</td>
</tr>
<tr>
<td></td>
<td>Southwest Surmic</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SOP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Southeast Surmic</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SOP</td>
<td></td>
</tr>
<tr>
<td>Berta</td>
<td>SPO</td>
<td>Bender 1996: 39</td>
</tr>
<tr>
<td>Gumuz</td>
<td>SPO</td>
<td>Bender 1996: 39</td>
</tr>
<tr>
<td>Nara</td>
<td>SOP</td>
<td>Bender 1996: 39</td>
</tr>
<tr>
<td>Kunama</td>
<td>SOP</td>
<td>Bender 1996: 39</td>
</tr>
<tr>
<td><strong>Hadza</strong></td>
<td>PSO</td>
<td>Sands 2013: 265</td>
</tr>
<tr>
<td>Tuu</td>
<td>SPO</td>
<td>Güldemann 2014: 17</td>
</tr>
<tr>
<td>Kx’a</td>
<td>SPO</td>
<td>Güldemann 2014: 17</td>
</tr>
<tr>
<td>Khoe-Kwadi</td>
<td>SOP</td>
<td>Güldemann 2014: 17</td>
</tr>
</tbody>
</table>
4.2.7 Europe

Europe is far less diverse than the other continents. Almost all its languages belong to the vast Indo-European family, which stretches to South Asia in the east. The European branches of this family mostly follow an SPO syntax, although the Germanic languages (excepting English) follow a V2 order in main clauses, in which the conjugated verb follows the first constituent of the sentence in declarative main clauses (whereas the conjugated verb occupies the clause-initial position in questions, and the clause-final position in dependent clauses). However, the Celtic languages native to the British Isles are prominent in the literature on P1. The six Celtic languages follow P1 quite consistently. Less known is that P1 is also a common constituent order in Modern Greek, although its usage is less consistent than in Greek, with SPO as another common order (Horrocks 1983, Mackridge 1985: 234–7). Cf. §6.2.7 for a detailed discussion of constituent order in the Indo-European family. The Finno-Ugric languages in the east of the continent have also adopted a basic SPO order, though that can be reconstructed to an earlier SOP order (Laakso 2011: 190, Kiss 2013). The same is true for the westernmost Turkic representatives of Gagauz and Karaim (Menz 2011: 17). Finally, the Basque language isolate on the Iberian peninsula follows a predicate-final order (Haase 2011: 212).
4.2.8 Asia

P1 order is practically absent from all of continental Eurasia. While the languages of Europe generally follow SPO order, the languages of most of Asia follow a consistent SOP order. This starts with the Asian members of the Indo-European and Finno-Ugric families, whose European cousins follow SPO (Masica 1993: 332 for Indo-Iranian). The Turkic family follows SOP consistently, with the exception of two of its westernmost members, Garauz and Karaim on the west coast of the Black sea, which adopted the European SPO order (Menz 2011: 171). The Turkic family stretches across central and northern Asia into eastern Siberia (with Sakha/Yakut), the whole area of which represents a solid SOP region with its various independent language families (Anderson 2004: 60, Georg 2008: 163, Pakendorf 2010: 715). This SOP dominance even continues to Korea and Japan on the Pacific coast.

SOP is also invariable in the linguistically diverse Caucasus area (Daniel and Lander 2011: 143), including the Kartvelian, Northeast Caucasian, and Northwest Caucasian families, as well as the Iranian Ossetian and the Turkic Azeri, but excluding the Indo-European Armenian language, which does show SOP order, but only with indefinite objects, while the unmarked order including a definite object is SPO (Ajello 1997: 223–4, Dum-Tragut 2009: 555–62). Another solid predicate-final area is South Asia, with members of the Indo-Aryan (Indo-European), Dravidian, Tibeto-Burman, and Munda (Austroasiatic) groups all agreeing in his basic syntax (Masica 1976: 19ff.). This generalization also
Numerical and areal overview of results

includes the isolates Burushaski (Berger 1998: 177;199), Nihali (Nagaraja 2014: 108), and Kusunda (Watters 2005: 381). Many languages in this area show ergative alignment to some degree, hence the ordering of the arguments before the predicate is described in different ways in the literature, depending on whether the ergative or absolutive argument is understood as the subject. However, this shall not concern us further for the purposes of the present study.

The only area in Asia that diverges from the SOP dominance in Asia described above is China and Southeast Asia. In contrast to the rest of the vast Sino-Tibetan family, which follows SOP order, the Sinitic (Chinese) branch follows a basic SPO order, although they also allow a marked SOP structure (Chappell 2001: 2-3, Chappell and Li 2016: 5, Yue 2017: 130). Likewise, the the Hmong-Mien (Ratliff 2010: 227, Sposato 2014: 111) and Kra-Dai languages (Diller et al. 2008, Jenks and Pittayaporn 2014: 1), both independent families that originated in Southern China but have stretched into Southeast Asia, follow a basic SPO order. The Austroasiatic languages, whose roots like within Southeast Asia, also mostly follow SPO order, even in the ancient Khmer and Mon inscriptions (ref:xxxx). However, this family shows more variation: first, the westernmost branch of the Munda languages follows the South Asian norm of SOP (see above); second, P1 is attested in three of the twelve branches. While P1 in some languages can be ascribed to contact with the neighbouring Austronesian family (cf. §6.2.1), not all of it can. For a detailed discussion of P1 in the Austroasiatic family, see §6.2.6.

4.2.9 Sign languages

Sign languages constitute several distinct families, but they shall be treated together here because of their high degree if globalization. While there is a slight preference for SOP in spoken languages (followed closely by SPO), this preference is much stronger in sign languages, especially in the early stages of language emergence (Goldin-Meadow et al. 2008). Sign languages often follow the constituent order of the local spoken languages, but where there is a discrepancy, it is more often the case that the sign language follows SOP, although the spoken language follows SPO, rather than the other way around. Examples of SOP sign languages with SPO spoken counterparts are LIS (Italian Sign Language), DGS (German Sign Language), or Al-Sayyid Bedouin Sign Language. Other orders beside SPO and SOP are not attested at all (Fischer 2014: 6, de Vos and Plau 2015: 272–3).
Chapter 5: Constituent order change and its reconstruction

In the reconstruction of proto-languages, the primary focus is usually on phonology and lexicon, with morphology following. Syntax, especially clause-level syntax such as constituent order, is often left untouched in the reconstruction of proto-languages. This is due to several reasons: first, constituent order is less tangible in that it often has no phonological basis as do the other major aspects of linguistic structure (or if it does in the form of intonational patterns, this part of a language’s phonology is seldom described adequately); second, syntactic structure cannot be derived from word lists, which is all the data we have for many of the world’s languages; third, it has been noted that constituent order is particularly prone to language contact (Dryer 1992: 83). To illustrate the fluidity of constituent order, I shall mention here my personal fieldwork experience with a fluent Halkomelem (Salish) speaker in Vancouver: in elicitation (translation from English), he fairly consistently produced sentences with the subject in initial position (mirroring the English structure), while in free narration virtually all sentences started with the predicate. Robert Blust (p.c.) reports the same experience in his fieldwork with Thao (Austronesian) on Taiwan, where the contact language Chinese is also subject-initial. Likewise, Lyle Campbell (p.c.) mentions that this happens frequently in fieldwork elicitation with P1 languages of Mesoamerica, where Spanish SPO constituent order is very often given in elicited sentences, even when it does not at all show up in narratives and dialogue. Many more such cases are reported in the linguistic literature. Finally, there is no consensus on the methodology of syntactic reconstruction as there is for phonological reconstruction. In fact, the field is so underdeveloped that syntactic change has repeatedly been attributed flippantly to contact with hitherto unknown languages. Similar statements in the field of phonology would simply not be acceptable. In the words of Harris and Campbell, “syntactic borrowing is perhaps the most neglected and abused area of syntactic change. Excesses in the past are well known and require little comment (...)” (Harris and Campbell 1995: 120). Although this situation has been remedied somewhat by comparatively recent attempts to move toward a more formalized theory of syntactic change, e.g. Harris and Campbell 1995, and Barðdal and Eythórsson 2012, one major persisting problem in the establishment of a generally agreed upon methodological standard for syntactic reconstruction are the various conflicting theories of syntax itself.

Harris and Campbell make the important point that in any type of diachronic change,
not only need the initial and final structures be typologically plausible, but the mechanism deriving one from the other needs be accounted for, as well. For constituent order in particular, language contact happens to be a strong force behind many of the observed changes in constituent order. An illustrative case for this is Romansh, a phonologically and morphologically very conservative Romance language. In terms of constituent order, though, Romansh follows neighboring German almost perfectly, as exemplified in (5.1).1

(5.1) Adoption of German V2 syntax into Romansh
Romansh (Romontsch Sursilvan) [roh, Indo-European, Romance]
German [deu, Indo-European, Germanic],
German: Das glaub-e ich nicht.
Romansh: Quei crei-el jeu buc.
that believe-1s 1s not
‘I don’t believe that.’
Cf. Italian [ita, Indo-European, Romance]
Non lo cred-o.
not it believe-1s
‘I don’t believe that.’

Language contact is so influential on constituent order that Smith (1981: 52) even suggested that changes in constituent order are primarily due to contact. While such a statement is difficult to ascertain, it is well established that constituent order can also change as the result of language-internal processes.

A change in constituent order can also derive from language-internal processes of dislocation—both synchronically and diachronically. A dislocated argument appears in the left or right periphery of a clause, is sometimes overtly marked, and usually prosodically separated from the rest of the clause. Dislocation can occur for various reasons, which may be driven by prosodic or discourse-pragmatic motivations (cf. Harris and Campbell 1995: 238. One often occurring prosodic motivation is the cliticization of a function word, resulting from phonological reduction to due frequent usage, and subsequent “floating” of the clitic away from its syntactic and logical head (usually toward the beginning of the clause, cf. below). Another prosodic motivation is the demotion of heavy (i.e. long) arguments to the clause periphery, so as not to disrupt the clausal core. Pragmatic motivations involve marking a participant as the topic or focus of a sentence. Example (5.2) shows left (5.2a) and right (5.2b) dislocated topics in English.

(5.2) Left and right dislocation in English [eng, Indo-European, Germanic]

a. That new building on King Street, they finished building it.

b. They went to the store across the road, the construction workers.

1More precisely, Romansh is not influenced by Standard German, but by the local variety of Alemannic, which is apparent in many lexical calques.
Based on such sentences, one could argue that English has OSP and PS constituent orders in addition to commonly recognized SP(O). However, the constructions in (5.2) show two signs that they contain dislocated topics, which are not part of the core clause: first, the noun phrases in question are prosodically separated from the rest of the clause by intonation and/or a pause (indicated in (5.2) by the comma); and second, they are taken up by pronouns in the core part of the sentence. One may thus argue that dislocation in the above sentences does not change the constituent order, at all, since the core clause still follows the regular SPO pattern. But such subtle prosodic and syntactic markers as pauses and pronoun copies may disappear over time. Also, if frequently employed—perhaps for stylistic reasons, or perhaps in analogy to heavy arguments—such constructions can lose their markedness and become the new norm. In French, for example, right dislocation is so common that some have analyzed the resultant dislocated structure as the norm and consider the pronouns in the sentence core as mere agreement markers on the verb. A well-known example of this structure is the quote by author Raymond Queneau in (5.3).

(5.3) Right dislocation in French [fra, Indo-European, Romance]

Il l’-at-il jamais attrapé, le gendarme, son voleur ?

He him-has-he never caught the policeman his thief

‘Has he ever caught him, the policeman, his thief?’

If the pronoun il ‘he’ in (5.3) is re-analyzed as an agreement marker on the verb, French would in fact represent P1 order. Despite the frequency of such constructions, most native speakers would still consider the structure in (5.3) to be derived and marked, compared to the more canonical SPO. Yet, this case demonstrates how frequent use of such a construction could over time lead to its fossilization and entail a change of constituent order.

Harris and Campbell (1995: 151ff.) argue that dislocation (especially when used to mark a focus) is itself often the result of a historical process. They show several cases where dislocation derived historically from a biclausal structure involving relativization. This historical development is exemplified in (5.4) with Breton.

(5.4) Historical development of biclausal focus cleft to monoclausal left dislocation in Breton [bre, Indo-European, Celtic, source: Borsley and Stephens 1989]

a. Ar vugale eo a lenne al levriou.
   the children is that read the books
   ‘It is the children that read the books.’ (p. 419)

b. Ar vugale a lenne al levriou.
   the children that read the books
   ‘The children read books.’ (p. 408)

Both sentences (5.4a) and (5.4b) are grammatical in modern Breton, but (5.4a) shows the older biclausal pattern with the copula linking the two nominal elements of an equational clause: a regular noun phrase ar vugale ‘the children’ and a relative clause a lenne al
levriou ‘that read the books’. The clause type in (5.4b) simply elided the copula, resulting in a construction that appears as left dislocation, with the dislocated noun phrase ar vugale overtly separated from the rest of the clause by a. Although these sentences represent a contemporary alternation in modern Breton that reflects a productive process, it is well imaginable that from this stage, the structure (5.4b) would fossilize, and perhaps eventually drop a phonologically weak particle a. This development would derive a SPO order from the original PSO order of Celtic languages. Thus, while dislocation often represents a synchronically productive pattern with marked pragmatic function, the resultant structure may become so frequent as to establish itself as the new norm, leading to a change in constituent order.

Right dislocation is often applied to phonologically heavy arguments (see above), but it is also a common repair strategy employed when the speaker senses that the referents of the anaphoric elements in the sentence were unclear to the listener and need to be re-stated; hence, it is also called an afterthought. Left dislocation (henceforth fronting), on the other hand, usually represents a deliberate intention to manipulate the listener’s attention, e.g. in a contrastive context or when introducing or inquiring about new information, which also explains why question words often appear in clause-initial position. This is rooted in a general psychological principle of salience of initial elements (cf. Clarke et al. 2015). As referents, rather than actions, are more salient and commonly contrasted and tracked throughout narratives, these are the elements that are eligible for fronting. Predicates, on the other hand, are not usually fronted (but cf. Chapter 2). This unsuitability of the predicate for dislocation may be part of the reason why P1 constituent order is cross-linguistically comparatively rare (cf. below).

The re-ordering of clausal constituent for the reasons described above generally begins as an extemporaneous process applied to specific prosodic/pragmatic conditions. But this process can gain functional ground over time and slowly establish itself as the new norm. This expansion often happens through the mechanism of analogy, which will be illustrated on the basis of cliticization: frequently used function words, most typically pronouns, often cliticize (involving phonological reduction and loss of an intrinsic accent, instead leaning on the left or right of other phonologically independent words) or develop clitic variants of the full words, and these clitics tend to float in the sentence, not appearing in the same position where the original independent word stood. It is especially frequent for clitics to follow the initial word of the sentence (Wackernagel 1892). While the emergent order may originally be phonologically conditioned, with clitics in a different position from their full independent counterparts, that order may be re-analyzed to be syntactically conditioned, so that non-clitic elements are then placed in the same position as the clitics. By this process, Germanic languages developed a word order in which originally only short auxiliaries but eventually any conjugated verb occurs in the second position of the sentence. Any other verb form remained in clause-final position, the default for the predicate complex in Proto-Indo-European. P1 order is relatively unlikely to be the result of such a development because there is pressure for the parts of the predicate complex, or of any other syntactic constituent, to remain contiguous (Harris and Campbell 1995: 237).
Even when an auxiliary cliticizes, it is more likely to lean on the main verb than on a syntactically more distant constituent. By contrast, pronouns (unless possessive) usually form their own syntactic constituents (an argument) and are thus freer to move around.

Another grammatical feature that is commonly understood to be influential in constituent order change is case marking, or rather the loss thereof. While case marking itself has no direct effect on constituent order, its loss may require remodeling of the constituent order to avoid ambiguity. Apparently, this remodeling is especially frequent if the predicate is peripheral in the original constituent order (predicate-initial or predicate-final), i.e. where arguments are directly adjacent. It has been suggested that such adjacency leads to ambiguity for syntactic functions, which can be resolved by separating the core arguments by the predicate, leading to SPO constituent order (Vennemann 1973). However, Harris and Campbell (1995: 216ff.) demonstrate that such re-ordering is neither universal nor predictable. The most-cited example of such change is from the SOP syntax of Proto-Indo-European to SPO in Modern English. However, the causal connection between loss of case marking and evolution of predicate-medial constituent order in English is questionable. Old English (Anglo-Saxon) still retained (at least some) case marking but had already abandoned the old SOP structure in favor of a verb-second (V2) syntax as in Modern German (Allen 1977). As explained in the previous paragraph, this structure resulted from the cliticization of auxiliary verbs and is unrelated to case marking. It is possible, though, that the already existing option of SPO, indeed the most common order in V2 syntax, was solidified as case marking was lost in English. A link between case marking and constituent order change (toward a predicate-medial construction) has also been suggested for Arabic (§6.2.2) and Bornean languages (Clayre 1996). Generally speaking, it is more likely that the loss of case marking renders syntax less flexible rather than that it induces a change of constituent order itself.

The historical processes described above may explain the relative rarity of P1 order (cf. §4): frequent application of fronting can over time lead to a generalized argument-initial constituent order, thus departing from P1. Right dislocation could lead toward P1 order, but its occurrence seems to be more ad hoc and less deliberate, and probably less frequent, than fronting. Thus, the rarity of P1 order may be the result of a preference of arguments in initial position due to the salience of participants. As actions are less salient, predicates are not usually fronted, leaving few syntactic options for it to move into clause-initial position.

Broadening the scope from syntactic change in a single language to the reconstruction of syntax in a language family, we have to recall that syntactic reconstruction needs to be based on the known processes of syntactic change. As with the reconstruction of other linguistic forms, it cannot simply rely on the principle that majority wins. Certain types of changes occur more frequently than others, and some may affect several languages of a family independently, given a certain inclination, such as weakness in the phonological signal, structural opacity or inconsistency, or heavy burden on memory. Such independent developments in the same direction have been called "drift" in the literature (Sapir 1921, Vennemann 1975). An example of a drift is the constituent order within the noun phrase
of Romance languages (Harris and Campbell 1995: 355–6): while virtually all major modern Romance languages show noun-adjective order (excluding small sets of divergent adjectives), we know from original ancient sources that the preferred order in their proto-language Latin was adjective-noun. Luckily, the historical documentation removes all doubt in this case, but hardly any other language families have such evidence available.

Analogically, it is entirely possible that a language family or group with (virtually) exclusive SPO order (e.g. Malayic, cf. §6.2.1) actually descends from a proto-language with PSO order. Such a general drift is indeed argued for in Hopper (1987). The same situation could arise not only from a drift with many sources of innovation, but it could also follow from a single innovation in one particularly prestigious or otherwise influential language (e.g. Malay), which would then be borrowed into related neighboring languages. Especially the second scenario always needs to be considered, necessarily in terms of geography and political history, due to the predisposition of constituent order to borrowing (see above). Of course, the two processes can work together, such that an inclination or a drift in progress can be accelerated by borrowing from a particularly innovative language.

On the other hand, an opposite drift, in which a family with consistent P1 order (e.g. Mayan) is derived from a subject-initial constituent order, would be less likely because the historical pathways towards predicate initiality are fewer and more rarely followed, as discussed above. Still, such a possibility cannot be excluded a priori. It is possible, especially for languages with extensive head-marking, that referential noun phrases are regularly dropped from many clauses, especially the ones that correspond to the continued topic in a narrative (usually the subject). In such pro-drop languages, the referents in a discourse need to be re-activated sporadically, usually ad hoc by an afterthought, i.e. right dislocation, a process that results in P1 order. It is conceivable that such a structure could fossilize, perhaps due to the loss of head marking that originally tracked the referents. This restructuring may occur in several languages at the same time, or it may start in just a single language, from which it then spreads through language contact. However, it seems less likely for a sporadic ad hoc repair strategy to be fossilized than for a deliberate syntactic operation.

While independent, i.e. language-internal, development of P1 order is thus rather unlikely, P1 order is more commonly introduced to a linguistic system by means of borrowing from an unrelated language, usually to individual languages, but given the right political setting to an entire family of languages. Importantly, such a claim can only be made when the original source language can be identified (as for Nahua, cf. §6.2.4), and any vague statements about possible but unidentified substrate influence (as some people claim for Celtic) belong to the realm of legend.

Still, synchronic phrase-level syntax alone is not conclusive about syntax in the distant past. Compared to phrase-level syntax, morphology is less readily borrowed, though there is no linguistic sub-system that is entirely safe from borrowing (Harris and Campbell 1995: 149). One is tempted to use morphological patterns to reconstruct syntax, along Givón’s famous claim, “Today’s morphology is yesterday’s syntax.” (Givón 1971: 413).
This dictum refers to the idea that the order of morphemes in a word reflects the order of those morphemes before grammaticalization, when they were still unbound, independent words. However, this is problematic especially with respect to constituent order: the morphological elements that allegedly reflect earlier constituent order are agreement markers on the predicate, but these often represent earlier pronominal forms, which we know are not the best indication of constituent order, because they are often cliticized (cf. above and 2). To illustrate this problem, Comrie (1980) presents the case of Buryat Mongolian, in which bound subject agreement evolved from earlier subject enclitics, even though the basic constituent order is (and has been since Classical Mongolian) SOP. Thus, the fact that subject agreement is realized in many languages by verbal suffixes (Eskimo-Aleut, Finno-Ugric, Indo-European, Semitic, Trans New Guinea, Turkic, etc.) is not so much representative of a tendency for the subject to follow the verb (on the contrary!) than of the frequency of pronominal enclitics.

More useful evidence in the reconstruction of syntax—indeed, in linguistic reconstruction in general—comes from relics or archaisms in an otherwise regular system. Since languages do not deliberately introduce randomness into their system, irregularity is usually an indication of earlier patterns that became opaque due to phonological change or the spread of a competing innovative pattern. Meillet (1903) even went as far as to proclaim that in general we should reconstruct based on exceptions, not on rules. Certainly, relics are valuable in reconstruction, but those may not always be available. In the field of syntax, relics may be found in languages aberrant from their genetic setting (when contact from foreign languages can be excluded), or they may be found in nested structures. Especially in contact situations, higher syntax is more readily affected than lower-level structures, i.e. sentence structure is affected before clause, phrase, and word structure (Aikhenvald and Dixon 2002, 2007). This also means that constituent order change first affects main clauses and diffuses to subordinate clauses later (Givón 1977, Vennemann 1973, 1975). As a result, subordinate clauses are often more conservative than main clauses (Bybee 2002). Subordinate clauses are also generally less flexible than main clauses because they are often not accessible for pragmatic profiling such as topicalization. On the other hand, subordinate clauses do not always represent the same structure as main clauses, at all: they are often non-finite constructions consisting of gerunds and participles, thus representing noun phrase-internal syntax rather than clausal syntax. The Turkish example in (5.5) illustrates this quite well.

(5.5) Turkish [tur, Turkic, source: author’s data]

a. Murat git-iyor.
   M. go-IPV
   'Murat is going.'

b. Murat-ın git-me-sin-i ist-iyor-um.
   M.-GEN go-NMLZ-3S.POSS-ACC want-PFV-1S
   'I want Murat to go.'

It shows that although Murat is the subject of the subordinate clause in (5.5b), he is not
marked as a subject (as in (5.5a)), but rather as a genitive attribute to the nominalized verb, which functions as the object of the matrix clause. Partly for this reason, subordinate clauses are not considered as evidence of P1 in this study (cf. Chapter 2). Still, arguments about syntactic reconstruction on the basis of constituent order in subordinate clauses deserve careful attention, with the said caveat in mind. As for aberrant languages, it has been suggested that peripheral languages are more conservative than central languages. As majority languages are spoken by more people across a wider area, their percentage of L2 speakers is also generally higher. These L2 speakers may import their own speech habits into the majority language, by way of imperfect language acquisition of the target language (Nichols 1992, Croft 2000, Næss and Jenny 2011). Smaller peripheral languages are far less prone to this type of change. Furthermore, they are not so much the target of deliberate standardization efforts, which are usually influenced by globally dominant languages, such as English or Latin. Therefore, smaller marginal languages deserve particular attention in the reconstruction of such contact-prone structures as constituent order.

Just as for the determination of a single language’s basic constituent order, typological correlates, such as the orientation of adpositions or the ordering of demonstrative and head noun, are also used as evidence in the reconstruction of syntax within a whole family. While I cautioned against the application of this method in the former case (because of its risk of data skewing, cf. Chapter 2), it can be considered as evidence in the deep reconstruction of language families. As languages tend toward harmony and regularity in the ordering of head and modifier at various phrase levels, typological discrepancies may be indicative of language contact. And as higher levels of syntax, such as clausal constituent order, are more prone to external contact (as well as internally motivated change), the direction of a typological shift may become apparent. For example, a language with head-modifier order in clausal syntax (i.e. predicate-object order) but modifier-head order within noun phrases (e.g. adjective-noun or relative clause – head noun) likely derives from an earlier modifier-head structure, suggesting that the order predicate-object may be innovative or borrowed, as is indeed the case for Sinitic. Although this methodology has some explanatory power, it is by no means fool-proof, as the typological correlations are not perfect (probably to a considerable part precisely because of language contact). Still, it can be employed in conjunction with the other methods described above.

Based on these general observations of syntactic change, and especially the pathways toward and away from P1 order, the following Chapter 6 investigates the syntactic history of the language families with attested P1 order. Luckily, this investigation does not start from zero for most families but rather critically reviews previous claims in that respect.
Chapter 6: The history of P1 from a phylogenetic perspective

Chapter 4 provided a geographic overview of the global distribution of P1. This chapter looks at P1 in terms of language families. It investigates the families (including isolates) with attestations of predicate initiality, ordered by their prevalence of P1. In the following sections, the language families with P1 are discussed by prevalence of P1: families with a majority of languages using P1 (overview in Table 6.1, discussed in §6.1), with a minority of languages using P1 (overview in Table 6.2, discussed in §6.2), and language isolates and dialect networks with P1 (overview in Table 6.1, discussed in §6.3).

Table 6.1: P1-majority language families.

<table>
<thead>
<tr>
<th>Language family</th>
<th>Number of P1 languages [conf/susp]</th>
<th>Total number of entries on Glottolog⁵⁸</th>
<th>Percentage of P1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otomanguean</td>
<td>171</td>
<td>182</td>
<td>94%</td>
</tr>
<tr>
<td>Mayan</td>
<td>29</td>
<td>31</td>
<td>94%</td>
</tr>
<tr>
<td>Salish</td>
<td>25</td>
<td>25</td>
<td>100%</td>
</tr>
<tr>
<td>Surmic</td>
<td>7</td>
<td>11</td>
<td>64%</td>
</tr>
<tr>
<td>Totonacan</td>
<td>7</td>
<td>13</td>
<td>54%</td>
</tr>
<tr>
<td>Enlhet-Enelhlet</td>
<td>6</td>
<td>6</td>
<td>100%</td>
</tr>
<tr>
<td>Kadugli-Krongo</td>
<td>6</td>
<td>6</td>
<td>100%</td>
</tr>
<tr>
<td>Wakashan</td>
<td>6</td>
<td>6</td>
<td>100%</td>
</tr>
<tr>
<td>Xinkan</td>
<td>5</td>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>Chumashan</td>
<td>5</td>
<td>6</td>
<td>83%</td>
</tr>
<tr>
<td>Sahaptian</td>
<td>4</td>
<td>5</td>
<td>80%</td>
</tr>
<tr>
<td>Chinookan</td>
<td>3</td>
<td>3</td>
<td>100%</td>
</tr>
<tr>
<td>Kuliak</td>
<td>3</td>
<td>3</td>
<td>100%</td>
</tr>
<tr>
<td>Tsimshianic</td>
<td>3</td>
<td>3</td>
<td>100%</td>
</tr>
<tr>
<td>Coosan</td>
<td>2</td>
<td>2</td>
<td>100%</td>
</tr>
</tbody>
</table>

⁵⁸ Since this study is only concerned with modern languages, I have subtracted from the total on Glottolog the older language stages that I could identify in cursory fashion.
Table 6.2: P1-minority language families.

<table>
<thead>
<tr>
<th>Language family</th>
<th>Number of P1 languages [conf/susp]</th>
<th>Total number of entries on Glottolog</th>
<th>Percentage of P1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austronesian</td>
<td>346</td>
<td>1265</td>
<td>27%</td>
</tr>
<tr>
<td>Afroasiatic</td>
<td>53</td>
<td>354</td>
<td>15%</td>
</tr>
<tr>
<td>Niloic</td>
<td>25</td>
<td>55</td>
<td>45%</td>
</tr>
<tr>
<td>Uto-Aztecan</td>
<td>31</td>
<td>67</td>
<td>46%</td>
</tr>
<tr>
<td>Arawakan</td>
<td>19</td>
<td>77</td>
<td>25%</td>
</tr>
<tr>
<td>Austroasiatic</td>
<td>16</td>
<td>156</td>
<td>10%</td>
</tr>
<tr>
<td>Indo-European</td>
<td>7</td>
<td>520</td>
<td>1%</td>
</tr>
<tr>
<td>Mixe-Zoquean</td>
<td>3</td>
<td>19</td>
<td>16%</td>
</tr>
<tr>
<td>Chapacuran</td>
<td>2</td>
<td>12</td>
<td>25%</td>
</tr>
<tr>
<td>Cariban</td>
<td>2</td>
<td>42</td>
<td>5%</td>
</tr>
<tr>
<td>Tupian</td>
<td>1</td>
<td>71</td>
<td>1%</td>
</tr>
</tbody>
</table>

a Since this study is only concerned with modern languages, I have subtracted from the total on Glottolog the older language stages that I could identify in cursory fashion.

(6.1) Isolates (or small dialect networks) with P1

- Alsea-Yaquina
- Cayuvava / Cayubaba
- Guató
- Hadza
- Huave (4 entries in Glottolog)
- Itonama
- Movima
- Kariri
- Kuot
- Salinan
- Siuslaw
- Taushiro
- Yagua
- Yokuts (4 entries in Glottolog)

Note that this categorization is by no means reflective of speaker numbers or of the internal diversity of the language family. Nor should one derive syntactic reconstruction for the family solely on these numbers, because reconstruction is not simply a matter of "the majority wins." For the reconstruction of syntax, as for any type of reconstruction,
the internal branching of the family must necessarily be taken into account. The problem with family trees, though, is that there is often agreement about lower-level branches of a family but disagreement about their exact relationship. This is true even for as well-studied language families as Indo-European, where a closer relationship between Italic and Celtic is under debate, or the external relationship of the isolate branch Albanian, just to name a couple. Especially for families with unresolved internal structure, the numeric prevalence of P1 becomes more relevant. When P1 is represented in the majority of languages in a family, and importantly across several sub-branches, the reconstruction of P1 appears obvious. But as P1 is typologically unusual, its presence in several branches can be enough to warrant its reconstruction, even if those branches represent only a minority of languages in the entire family. However, because constituent order is so prone to contact (cf. Chapter 5), we cannot be certain of the reconstruction if there are contemporarily or historically culturally dominant languages of the same type in the area. Generally, reconstruction is more feasible in larger, geographically dispersed families than in smaller families with their languages in constant exchange, because of the possibility of lateral borrowing in the latter case. Isolates are particularly challenging, albeit not impossible, to reconstruct. Arguments for a differing constituent order in the past of a language isolate can come from structural discrepancies such as a typologically unusual profile, e.g. P1 coupled with postpositions. Or they can come from obvious signs of contact in other parts of a language’s grammar, e.g. in the lexicon. In this chapter, only the isolates with such noteworthy arguments are presented, as the default assumption, in the absence of evidence to the contrary, is the continuation of structure.

Hence, the families presented above are studied in more depth in the following subsections, investigating the time depth of P1 in each family, and paying particular attention to cases where P1 appears to have arisen in recoverable history. Previous literature on this topic is critically examined both in view of the present statistics of this study and in consideration of the synchronic patterns of syntactic change discussed in Chapter 5.

6.1 Language families with P1 in a majority of languages

6.1.1 Otomanguean

*Ethnologue* counts 177 varieties of Otomanguean, though the Catalogue of Endangered Languages recognizes only 63 distinct languages. Practically all of these are predicate-initial, mostly PSO as e.g. Zapotec, though some have been described as lacking a dominant order between subject and object.

Numerically speaking, most Otomanguean languages belong to one of two major sub-groups, Mixtecan and Zapotecan (though these do by no means represent the entire diversity across the eight branches). Mixtecan consistently follows PSO as its basic constituent order without exception, according to Hollenbach (1991). And for Zapotec, Rosenbaum (1974: 11) states that "the unmarked surface order of main constituents is VSO."
This generalization is iterated by Operstein and Sonnenschein (2015: 11), who assert that the order POS is ungrammatical. Marlett (1985: 103) Marlett qualifies this generalization only slightly by adding that Rincón Zapotec deviates from the general PSO order by using POS instead. Another major sub-group, Chinantec, is described by Merrifield (1994: 210) to be consistently P1, PSO for the most part, with the exception of Group IV (Lalana, Lealao, Tepinapa), which follow POS order.


\textit{ndyēhē tásù ndīvī}

\textit{cont.see hawk chicken}

'The hawk sees the chicken.'

(6.3) Isthmus Zapotec [zai, Zapotec, source: Bueno Holle 2019: 114]

\textit{bi-chuugu' rigola que pera que}

\textit{compl-cut man dist pear dist}

'The man cut the pears.'

(6.4) Ozumacin Chinantec [chz, Chinantec, source: Rupp 2012: 258]

\textit{gajngiñh dsa (kih) øyḥ}

\textit{killed person obj.anim 3s.poss.companion}

'The person killed her companion.'
The above three sub-groups account for 75% of Otomanguean languages. Other smaller groups are also described as generally following PSO order, such as Popoloca (Veereman-Leichsenring 1994: 241). I have checked the resources for every single language in this family, and I could confirm P1 with data for 61 languages. The majority of the remaining will use P1, as well, even as the basic constituent order, but I could not confirm this due to a lack of data. Only very few languages are explicitly described as following a non-P1 basic constituent order, including Ixcatécan, a couple of varieties of Otomi (Querétaro and Ixtenco) and San Francisco Matlatzinca. Some languages explicitly exclude any P1 order. Thus, Adamou (2018) describes Ixtatec to have a rigid SPO order. Palancar (2009: 559) describes Querétaro Otomi to follow a SPO order, being aware that this must have evolved from an older POS, which he explicitly states to be present only in old texts. (However, the predicate still most often precedes the subject in intransitive clauses.) As for Matlatzincan, both of its languages (Atzingo Matlatzinca = Ocuilteco = Tlahuica, and San Francisco Matlatzinca) are described as following a basic SPO order. Both varieties exhibit P1 as an alternative, but only Ociuteco is included as P1 in this survey, because for it PSO is cited as a “common syntactic construction” without further elaboration to the effect that it is a marked structure (Muntzel 1986: 136), while the possible order POS in San Francisco Matlatzinca is associated with special prominence on the subject (Gómez 2015: 20). Whether this truly reflects a difference between the two varieties or just a different research approach is a question that could only be answered by a dedicated study into the topic. Finally, one notable exception is Chichimeca, which follows a basic SOP order, which is at odds with the otherwise universal PO ordering of Otomanguean (Lastra de Suárez 1984: 35, 2014: 45, Lizárraga Navarro 2018: 74).

In all Otomanguean varieties, constituents can be fronted to preverbal position for focus, and although fronted arguments are overtly marked, subject fronting—probably under the influence of Spanish—is so common that SPO often prevails in actual texts, in the case of Popoloca in several, perhaps most, dialects. Nonetheless, P1 (mostly PSO) is still a commonly employed, pragmatically and structurally unmarked constituent order in the vast majority of Otomanguean languages. Therefore, Henry and Hollenbach’s (1988: 5) statement that PSO constituent order “appears to be reconstructible as the basic order in the Otomanguean stock” is uncontroversial.

6.1.2 Mayan

Kaufman (1986) claims that all Mayan languages are predicate-initial, with the sole exception of Ch’orti’. Several other authors have classified Mayan languages by constituent order, which mostly boils down to the respective ordering of post-predicate absolutive and ergative arguments (cf. e.g. England 1991).

(6.5) Q’anjob’al [kjb, Q’anjobalan, source: Toledo 2008: 49]

\begin{verbatim}
Max y-il ix Malin heb’ xal ix compl 3s.erg-see class M. PL class woman
\end{verbatim}

‘Malin saw the old women.’
Language families with P1 in a majority of languages

Figure 6.2: P1 in the Mayan language family

(6.6) Mam [mam, Mamean, source: England 1983: 193]
ma kub’ ky-tzyu-7n xinaq cheej
REC DIR 3P.ERG-grab-DS man horse
‘The men grabbed the horse.’

(6.7) K’iche’ [quc, Quichean, source: Ixcoy 1997: 345]
x-r-atin-isaa-j ri a Koot ri ixoq Ixwatzik
COMPL-3P.ERG-bathe-CAUS-TR DEM CL K. DEM CL I.
‘Ixch’umil bathed Kot.’

In some languages, that order is fixed and disambiguates the arguments, which are otherwise unmarked for case; but in many languages, the order is variable. Some authors point out the marginal role of clauses with two NP arguments and others caution about a difference between elicited and narrative data, of which the following excerpt by Barrett on Sipaka is an example:

“SVO word order is often given in elicitation data. This is probably caused by the combination of translating from Spanish (which has SVO word order) and producing elicited sentences, which almost always introduce new discourse topics. In actual natural language, SVO order is much less frequent [than] VSO. For example, in a narrative text containing 116 sentences with verbs
and at least one expressed noun phrase, SVO occurred in only 3 sentences.
Each of these sentences is an example of topicalization.” (Barrett 1999: 219)

Since authors use different data as the basis of their analysis, which is also shaped by
different theoretical approaches, a true comparison of the actual languages would require
a unified approach. Thus, the published data are as much a representation of linguistic
differences as of different linguistic traditions. An interesting contribution to this topic is
Benito Pérez’s (2016) detailed study of constituent order in Poqom: he observed that single
arguments of a transitive clause (whether ergative or absolutive) have a strong tendency
to occur after the predicate; however, in the few cases where both arguments are overt
as NPs, there is a strong tendency for the subject to precede the predicate, i.e. for the
two arguments to be separated by the predicate, thus showing a co-argument dependency
in constituent order. Yet, Benito adds that in the cases of SPO order, the initial subject
represents a topicalized participant.

Additionally, there is evidence that constituent order is changing or has already
changed with the growing dominance of Spanish: Schumann Gálvez reported in the year
1997 that elderly Mopán speakers regularly follow POS order, whereas younger speakers
follow the Spanish order SPO (Schumann Gálvez 1997: 183).

The discussion on constituent order is further complicated by ordering restrictions
imposed by animacy and definiteness. Both of these features are relevant in Ch’ol Mayan.
Vázquez Álvarez (2011) demonstrates how an animacy hierarchy forces one participant to
topic position, as in (6.8), where the snake (lukum) could not be placed in the prototypical
subject position after the object, because it ranks lower in animacy than the woman (xixik).

(6.8) Ch’ol [ctu, Cholan, source: Vázquez Álvarez 2011: 22]

a. *tye i-k’ux-u x-ixik lukum
   PFV 3S.ERG-bite-TV CLASS-woman snake
   Intended meaning: ‘The snake bit the woman.’

b. jin lukum tye i-k’ux-u x-ixik
   TOP snake PFV 3S.ERG-bite-TV CLASS-woman
   ‘It was the snake that bit the woman.’

Such restrictions, however, do not distract from the fact that P1 is a common and
completely ordinary structure in Ch’ol, according to Vázquez even the least marked
structure.

Contradicting Vázquez Álvarez, Coon (2010) argues that P1 in Ch’ol is not base-
generated. Her argument is based on the observation that objects cannot be overtly
marked as definite in a POS clause, but they are required to be so in a PSO clause. Since
this study is only concerned with surface structure, we can nonetheless recognize Ch’ol as
P1 for our purposes. The fact that definiteness marking is not applicable to a POS clause is
of no concern to this study, as these nominals can still receive a definite reading, i.e. they
are not generic or incorporated into the verb in any sense.
In a study dedicated to basic constituent order in Mayan languages, England (1991) confirms Kaufman’s characterization of constituent order in Mayan languages: she lists 27 Mayan languages as P1 (both PSO and POS), and only one as SPO – Ch’orti’, which belongs to the Cholan subgroup like Chontal. England further notes that SPO order in Ch’orti’ is “presumably innovative and recent” (England 1991: 455).

In WALS, however, Lacandón, Tz’utujil and Chontal are listed alongside Ch’orti’ as exceptions to the predominance of P1. The first two are listed to lack a dominant constituent order, while Chontal is listed with a SPO basic constituent order. Whatever their basic constituent order, all of these languages make common use of P1, exemplified in (6.9)–(6.11).

(6.9) Lacandón [lac, Yucatecan, source: Bergqvist 2008: 67]
\[ t-u-kuch-o7b’ \quad \text{in-b’a7tak} \quad \text{tzimin} \]
\[ \text{compl-3s.erg-carry-3p.abs 1s.gen-things horse} \]
’The horse carried my things.’

(6.10) Tz’utujil [tzj, Quichean, source: Dayley 1985: 305]
\[ X-keetij \quad \text{tzyaq} \quad \text{ch’ooyaa7} \]
\[ 3.abs-3p.erg-eat clothes rats \]
’Rats ate (the) clothes.’

(6.11) Chontal [chf, Cholan, source: Knowles 1984: 316]
\[ ðu \quad \text{hæ’e’-e’} \quad \text{ʔah yan} \quad ðu \quad \text{h-ih-č’upin} \]
\[ 3.erg hit-ipv M \quad \text{John 3.erg fig-woman} \]
’John hits his wife.’

In contrast to these languages, Ch’orti’ does allow POS, but it is “corriente poco usual” according to Pérez Martínez (1994: 118), while PSO is completely ungrammatical.

P1 is so widely spread across different subgroups of Mayan that Norman and Campbell (1978: 144) conclude convincingly that “it is virtually certain that pM [proto-Mayan] was verb-initial.” This conclusion is further corroborated by the fact that the earliest attestations of Classic Maya in hieroglyphic writing also show P1. The only debatable point about Proto-Mayan constituent order concerns the relative ordering of post-predicate arguments: while Norman and Campbell (1978) argue (on the basis of Tzeltal and Huastec grammars) that the order depended on the relative animacy of the arguments, England (1991) particularly reconstructs POS as the Proto-Mayan constituent order. She considers fixed PSO constituent order a local diffusion with origins in the Mamean branch.

### 6.1.3 Salish

The Salish language family consists of 23 languages, all of which follow an unmarked P1 syntax: “[i]n all Salish languages, the predicate is most often clause-initial, followed by nominal expressions and prepositional phrases coding participants in the event” (Kroeber
Language families with P1 in a majority of languages

1999: 37). This generalization is further confirmed in a comment by Henry Davis on WALS: “To be precise, all of them [Salish languages] are V-initial, with flexible post-predicative order (VOS/VSO). In most languages/dialects (all of Central Salish, Lower Lillooet, Thompson, Okanagan), the ‘unmarked’ order (in sentences presented out of the blue containing two animate arguments) is VSO; in some Interior languages (Upper Lillooet, Moses Columbia, Coeur D’Alene) it is VOS” (Dryer and Haspelmath 2013). This present study confirms this with data for 15 languages from both Coast Salish and Interior Salish, as well as the geographically isolated Nuxalk (Bella Coola). Some examples of P1 in different Salish branches are presented in (6.12)–(6.14):

(6.12) Island Halkomelem [hur, Coast Salish, source: Gerdts and Hukari 2009: 4]
\[yəqʷ-ət-əs tə̌̌ mastimaxʷ tə̌̌ ey̓əl̓eqt sc̓ešt (...)]
burn-TR-3.ERG DET person DM long.PL stick
‘The people burned long sticks (...)’

(6.13) Okanagan [oka, Interior Salish, source: Hébert 1982: 16]
\[q̓ʷq̓ʷʕə-s-t-ı́s th-tw’it ʕi qqwálx]
chew-1PV-t-3.ERG the boy the fish
‘The boy is chewing the fish.’

Although Salish languages may differ in their preferred ordering of arguments after the predicate (cf. Czaykowska-Higgins and Kinkade 1998: 37), many languages allow either order, and there is a stronger tendency for constructions with two overt arguments to be avoided, which even amounts to a hard constraint in some languages, e.g. in Lushootseed (Hess 1973: 89).

Salish languages are spoken across a large area, stretching from the Pacific coast all the way to Montana in the interior. The coastal languages have not had regular contact with the interior for a long time. Furthermore, Nuxalk (Bella Coola) is separated from other Salish people by Wakashan communities. Given this distribution of languages and the fact that all attested Salish languages (spoken or extinct) follow this pattern, it is safe to conclude that P1 was already a feature of Proto-Salish.

6.1.4 Surmic

The Surmic family is native to the border area between South Sudan and Ethiopia. It is included as a primary branch of the controversial Eastern Sudanic family by its advocates. In Unseth’s (1988) subclassification of Surmic, Majang is the only representative of a
Northern branch, while the Southern branch splits into Southwestern and Southeastern branches, each with a handful of languages. Out of these, about half, that is Majang and all Southwestern languages, are strictly predicate-initial.

   \[Bokotu\text{-}ko Bokaye d\text{é}pe.\]
   \kill+PST B. lion\]
   'Bokaye killed a lion.'

(6.16) Didinga [did, Southwest Surmic, source: Lohitare et al. 2012: 52]
   \[Ur\text{u}k \text{é}\text{-}t\text{-}i \text{oli} \text{cinni}\]
   \kill man-NOM bull his\]
   'The man killed his bull.'

(6.17) Murle [mur, Southwest Surmic, source: Arensen 1982: 52]
   \[A\text{cin} kire\text{eri} \text{o}\text{ro}z\]
   \sees jackal dog\]
   'The jackal sees the dog.'

Core arguments are ordered by topicality after the predicate, where the more topical argument occurs first. Case marking distinguishes unmarked absolutive from tonally marked nominative (Tucker and Bryan 1966: 289, Dimmendaal 1983: 251).

The Southeastern languages are characterized as SPO by Unseth (1988), but as following a more flexible verb-second word order by Dimmendaal (1998).

This distribution of constituent orders led Unseth (1988) to the conclusion that Proto-Surmic was P1. However, this hypothesis is called into question by Dimmendaal (1998), who argues that other characteristics of the Surmic languages, namely dependency-marking and post-verbal question words, are typologically unusual for predicate-initial languages, as formulated by Greenberg’s universals (Greenberg 1966). Dimmendaal suggests instead that the flexible verb-second structure of Southeastern Surmic is older and that some Surmic languages then adopted P1 from the neighboring Nilotic languages. As both of these hypotheses are quite probable, this topic deserves more attention.

6.1.5 Totonacan

The syntax of the relatively small Totonacan family spoken in the Eastern Sierra Madre of Central Mexico, allows for much freedom in the order of constituents, unlike the neighbouring Otomanguean, where constituent order is more rigidly PSO. The constituent order of elicited clauses often follows that of the elicited input (i.e. SPO in Spanish elicitation), though consultants accept any other order when the semantic roles are predictable from the context (MacKay 1999: 4 for Misantla Totonac, Kung 2007: 535 for Huehuetla Tepehua). P1 is one commonly employed order, and this is cited as the most basic order for several varieties, such as for Upper Necaxa Totonac (Beck 2004: 92).
However, as documentation of these languages is not exactly abundant, I could only confirm P1 for seven of the thirteen varieties distinguished in Glottolog.

(6.18) Tepehua of Tlachichilco [tpt, Tepehua, source: Watters 1985: 104]

\[ Sa:-l \quad Kwan \Jose \\
hit-PFV Juan \Jose \\
'Juan hit Jose' or 'Jose hit Juan.' \]

(6.19) Highland Totonac [tos, Totonac, source: Juárez Esteban 2016: 72]

\[ uksilh-lh=ɪ juse lusiya \\
see.3s>3s José Lucía \\
'Lucía saw José' or 'José saw Lucía.' \]

(6.20) Upper Necaxa Totonac [tku, Totonac, source: Beck 2004: 92]

\[ le:n-t \quad tsamá: puská:t tsamá: mufnī \\
take-PFV that woman that monkey \\
'The monkey carried off the woman.' \]

As there is no case marking in Totonacan languages, there is considerable ambiguity in the syntactic relations of two NP arguments, when context does not offer clarification (MacKay 1999: 4). Out of context, the first argument is generally interpreted as the subject,
Language families with P1 in a majority of languages

and SPO is often used to disambiguate. Therefore, Kung (2007: 536) considers SPO to be the basic constituent order for Huehuetla Tepehua.

Among the descriptions of Totonacan constituent order, I found only one source that described P1 as a marked order, namely Reid 1968 for Xicotepec De Juárez Totonac: Reid describes SPO to be the preferred constituent order, with PSO indicating special emphasis on the predicate (Reid 1968: 31–2).

Despite the general flexibility in constituent order, the Totonacan family is generally characterized as a P1 family, and thus the languages for which I could not confirm with any data are likely to also follow this basic constituent order, as well. In fact, Totonacan has been cited as a possible origin of P1 in Nahuatl, with which it has been in contact for many centuries (cf. §6.2.4).

6.1.6 Enlhet-Enenlhet

Figure 6.6: P1 in the Enlhet-Enenlhet language family

Not all languages of this small family of the Paraguayan Chaco have readily accessible data, but Kalisch (2019) presents an in-depth analysis of constituent order in Enlhet (Norte), the largest member of the family. According to Kalisch, all six languages behave
similarly in terms of constituent-order phenomena, although there may be differences in certain details (Kalisch 2019: 176).

P1 is the norm in Enlhet, though clauses with preposed nominals are not uncommon, namely when a new participant is introduced or when the preposed constituent is otherwise in focus (Kalisch 2019: 154–6). The pre-predicate focus position is not reserved for arguments, though, but can be filled by other constituents. The respective order of the core arguments after the predicate is not syntactically determined, but rather, the most topical argument appears last (Kalisch 2013: 269). Since this is most often the subject, this leads to an SOP ordering. Though the arguments are not case-marked, their syntactic relations are usually specified by verbal agreement markers. Elliott (2021) even argues that any NP participants do not form part of the clausal core but are to be understood as adjuncts instead (cf. Chapter 2 for a discussion of this type of analysis). However, if both subject and object have matching number and gender, ambiguity can only be resolved by context, as in (6.21)–(6.22).

(6.21) Enlhet Norte [enl, Enlhet-Enenlhet, source: Kalisch 2019: 144]
\[
\text{ang-ya’pa-s-kas-kek} \quad \text{lhaak} \quad \text{sa’kok meeme} \\
\text{f-bathe-CAUS:l-CAUS-PRIM REC.PST girl} \quad 1s.mother \\
‘Mother bathed the girl’ or ‘The girl bathed Mother.’
\]

(6.22) Enxet Sur [enx, Enlhet-Enenlhet, source: Elliott 2021: 284]
\[
\text{ap-tekpog} \quad \text{Carlos Juan} \\
\text{m-hit-DECL C. J.} \\
‘Carlos hit Juan’ or ‘Juan hit Carlos.’
\]

Beside Enlhet Norte and Enxet Sur, I was able to confirm P1 with data only for Sanapaná (6.23), but I found no counterexamples to a general P1 in the entire family.

(6.23) Sanapaná [spn, Enlhet-Enenlhet, source: Gysel 2017: 27]
\[
\text{apok-apas-ke=ɬta} \quad \text{ap-ketka} \quad \text{en-japoŋ neteŋ} \\
\text{2/3.m-send-r=REM.PST 2/3.m-child pl-father top} \\
‘Our father above sent his son.’
\]

Since there are no counterexamples to P1 Enlhet-Enenlhet languages, and since there are no other P1 language families in the general area, one can assume that this order can be reconstructed to the proto-stage of this small language family, although a more recent lateral diffusion cannot be excluded as a possibility, given the geographic proximity of its languages.

### 6.1.7 Kadugli-Krongo (Kadu)

These six languages of the Nuba mountains are described as consistently P1 in the literature, although I could only confirm this assertion with data from Keiga and Krongo, as there are big documentation gaps in this family. Kadugli-Krongo stands out among
the geographic grouping of the Kordofanian languages, as the other languages in the area exhibit mostly SPO (Jilud, Katla, Talodi, Heiban, Lafofa), though SOP also occurs (Rashad, Lafofa, and somewhat farther in Hill Nubian and Nyimang). The closest neighbors of the Kadugli are the Daju, who follow SPO syntax; the Krongo are neighbors to the Talodi, who also follow SPO. The nearest P1 languages to the Kadugli-Krongo family are found along the South Sudan-Ethiopia border, namely the Surmic languages (§6.1.4). A relationship between Kadugli-Krongo and Surmic is indeed proposed, as both are grouped in the contended Nilo-Saharan phylum, but even among advocates of this speculative family, Kadugli-Krongo is not always included (cf. Ehret 1995), and when it is (e.g. Bender 1989, 1991), it is not considered a sister to other P1 languages.

Since P1 in Kadugli-Krongo is consistent and cannot be attributed to language contact, it is most likely reconstructible to the proto-language. Whether P1 in the proto-language was inherited from an older stage that links it to other languages in the area is a question that can only be enlightened once more extensive documentation and research into Eastern Africa is accomplished.

### 6.1.8 Wakashan

All six languages of this small family native to the Pacific coast of British Columbia, Canada, use P1 regularly.

(6.24) **Heiltsuk** [hei, Northern Wakashan, source: Waldie 2002: 87]
Language families with P1 in a majority of languages

Figure 6.8: P1 in the Wakashan language family

\[ \text{dáduqvlá wísm-á-xí w’áč’-íá-xí} \]
\[ \text{watch man-D1-D2 dog-D1-D2} \]
\'The man watches the dog.\'

(6.25) Kwak’wala [kwk, Northern Wakashan, source: Littell 2016: 45]
\[ \text{hamap=ux-da gwi7gasux x=ux-da xatamx} \]
\[ \text{eat=MED=DET RED-pig ACC=MED=DET carrots} \]
\'The pigs are eating the carrots.\'

(6.26) Nuu-chah-nulth [nuk, Southern Wakashan, source: Davidson 2002: 120]
\[ \text{haʔuk-šiƛ’-aƛ nəixtin quIšin-m̓ i:t} \]
\[ \text{eat-PRF=TEMP salmon.eggs raven-son.of} \]
\'Raven began to eat the salmon eggs.\'

(6.27) Makah [myh, Southern Wakashan, source: Davidson and Werle 2010: 142]
\[ \text{wiʔibaqstiqitdu duwiqsubitdis.} \]
\[ \text{wiʔiba-aqstiq=it=du duwiqsu=bit=dis} \]
\[ \text{angry-inside.MOM=PINV=1P father=PST=1P.PS} \]
\'Our father got angry at us.\'

At least in Southern Wakashan, the respective order of core arguments after the predicate is not strict (Davidson 2002: 109).
Reconstruction of P1 to Proto-Wakashan appears obvious. However, the Wakashan family is small and geographically fairly contiguous, with a history of frequent contact, thus leaving open the possibility of syntactic borrowing inside the family or from its P1 neighbors, especially the large Salish family (§6.1.3). Although Wakashan languages have a prominent and fairly long history of linguistic documentation, thanks to Franz Boas, Edward Sapir, and Morris Swadesh, the reconstruction of their history appears to be in its initial stages. Fortescue (2009) argues mostly based on the presence of lexical suffixes that the prevalent PSO order in Wakashan languages nowadays may have been pre-dated by an earlier SOP stage. He suggests that this shift may have been triggered by contact with Salish languages. Whether or not one finds Fortescue’s arguments convincing, proto-Wakashan is still reconstructed as predicate-initial even in his model, as the proposed shift would have taken place in pre-Wakashan.

6.1.9 Xinkan

All four members (5 entries in Glottolog) of this small language family located in southern Guatemala ceased to be spoken during the 20th century. Luckily, they were documented mostly by Terrence Kaufman and Lyle Campbell during the 1970s so that we can today gain a good understanding of their syntax. Rogers 2010 offers insight into grammatical aspects of these languages, including constituent order. According to Rogers, the individual languages are syntactically so similar that they can be described as one.
“This means that all of the Xinkan languages have very nearly identical syntactic structures and that there has been little historical change from proto-Xinkan syntax to the daughter languages” (Rogers 2010: 337). Predicates almost always appear in clause-initial position, though pragmatic factors can give rise to any other conceivable constituent order (Rogers 2010: 351).

(6.28) Jumaytepeque [xinc1235, Xinkan, source: Rogers 2010: 243]
\[ut'u-yi' \quad a \quad ur'ul \quad a \quad miya\]
lay.PERF-3S.PERF.TR the egg the chicken
'The chicken laid the egg.'

(6.29) Guazacapán [xinc1246, Xinkan, source: Rogers 2010: 39]
\[opo-y' \quad palh \quad uuts'i \quad ayaalha\]
break.PERF-3S.PERF.TR now nixtamal woman
'The women broke the nixtamal.'

Campbell (1978) suggested that the Xinkan languages acquired POS constituent order through contact with Mayan, given the latter’s areal historical dominance, but since all Xinkan varieties agree in this order, it could just as well have been a feature of the proto-language.

### 6.1.10 Chumash

![Figure 6.10: P1 in the Chumash language family](image)

Klar (1977: 133) asserts that this family native to the area around Santa Barbara, California, follows a POS constituent order, as a whole. This order is exemplified below.
Language families with P1 in a majority of languages

(6.30) Barbareño [boi, Chumash, source: Applegate 2017:88]
\[
\text{s-'oyon-us} \quad \text{hi} \quad \text{s-kok'o} \quad \text{hi} \quad l-'ihi'y
\]
3-help-3.OBJ LNK 3-father LNK ART-man
'The man helps his father.'

(6.31) Ineseño [inz, Chumash, source: Applegate 1972]
\[
\text{s-ulü-aq-pey-us-wun} \quad \text{ha-weselu ha-miy}
\]
3-against-quick-stick.to-3.OBJ-3p.OBJ ART-calf ART-wolf
'The wolf chases the calves.'

(6.32) Obispeño [obi, Chumash, source: Klar 1977: 133]
\[
\text{ča} \quad \text{yaktakaka yatqawi}
\]
chases the.quail the.dog
'The dog chases the quail.'

The only exception to this generalization is Cruzeño, which exhibited mainly SPO order “at the end of its existence.” According to Klar, however, internal evidence of this language points toward an earlier P1 initial order. It is evident to her that Proto-Chumash order was POS, as virtually all languages agree, including the otherwise divergent Obispeño (6.32).

California is generally characterized by free constituent order, albeit an increasingly predicate-final constituent order toward the south. Yet, the Chumashan languages are not the only syntactically divergent languages in the area; in fact, they form a small P1 cluster in central California together with the neighboring Salinan and Yokuts languages (§6.3.5 and §6.3.6, resp.). Since no relationship between these languages can be established, it is most likely that P1 spread through contact in this area, and it is quite possible that the Chumash were the origin of this small Sprachbund, as they were blessed with an abundance of maritime resources, which they traded for goods from the interior. The Chumash lived in large towns, and even hosted regular public ceremonies during the solstices, which other tribes would attend. Thus, the Chumash formed a cultural centre that certainly influenced the nations around them (Gamble 2008).

6.1.11 Sahaptian

This family, commonly included in a broader Plateau (Penutian) phylum (along with Molala and Klamath, which has a free constituent order with a tendency toward SOP), comprises the two distinct languages of Sahaptin in inland Washington and Oregon, as well as the Nez Percé (\textit{Niimiipuutímt}) language of Idaho. Though Sahaptin can be further split into local varieties (\textit{Glottolog} distinguishes 4 varieties), these are mutually intelligible and are not always differentiated in linguistic description. The best described variety, which forms the base of the discussion below, is the Yakama variety of Sahaptin, known by native speakers as \textit{Ichishkiin Sinwit}.

Sahaptin commonly employs P1 constituent order, with the subject and object following in either order, though “[w]ord order is subject to pattern preferences, to stylistic
choices, to tendencies, not to invariable rules” (Jacobs 1931: 269). Nonetheless, Jacobs considers PSO the most usual constituent order.

(6.33) Northwest Sahaptin [yak, Sahaptin, source: Jacobs 1929: 208]

i-nánan-a áyat xʷayamá  
3.NOM-carry-pst woman eagle  
‘The eagle carried away his woman.’


i-tkáta-xan-a wawatáwyá ts’ts’k-na  
3.NOM-eat-HAB-pst antelope grass-OBJ  
‘Antelope kept eating the grass.’

Similarly, P1 clauses are readily found in Nez Percé, and according to Aoki (1979: 1), PSO can be understood as the basic constituent order for this language.

(6.35) Nez Percé [nez, source: Phinney 1934: 350]

hi-tulúu-m-e háacwal wá’wa  
3.NOM-cast.into.water-CISLOC-pst boy hook  
‘The boy cast the hook into the water.’

However, according to Rude (1992, 2009), constituent order in both Sahaptin and Nez Percé is not related to syntactic role, but rather to discourse prominence. He concludes that in Nez Percé, "post-verbal nouns tend to be more topical than pre-verbal ones"
(Rude 1992: 206). In a survey of (rare instances of) Northwest Sahaptin transitive clauses with both overt subjects and objects, 70 out of 149 had the predicate preceding the two arguments, which occurred equally commonly in either order (Rude 2009: 16). Though the single most common constituent order in Rude’s count was SPO, the predicate actually preceded the subject in 86 out of 149 clauses, and it preceded the object in 120 out of 147 clauses. Predicate-final order is possible but quite uncommon, with only 13 instances. Thus, while constituent order may be fairly free in Sahaptin, P1 is definitely very common. At least based on Rude’s publications, it seems that the same tendency toward P1 is not present in Nez Percé: pre- and postverbal subjects and objects occur in almost exactly equal proportions (Rude 1988: 227). Hence, though P1 is possible in Nez Percé, there does not seem to be a preference for it, and therefore it is not included in the present survey. It is of course possible that Sahaptin and Nez Percé are quite similar in constituent order, but based on the available studies, only Sahaptin appears to show a preference for P1.

Sahaptin forms the southern border of the Pacific Northwest P1 area: to its north are the P1 Salish languages (§6.1.3), down the Columbia River toward the Pacific coast it borders the P1 Chinookan family (§6.1.12), but to the south it borders the predicate-final Shoshone of the Uto-Aztecan family, a tribe with which they were chronically at war.

### 6.1.12 Chinookan

This small family of three languages along the lower Columbia River of Washington and Oregon lost its last native speakers at the beginning of the 21st century in the Upper Chinook language. The other languages Lower Chinook and Kathlamet became extinct in

---

1Bear in mind that SPO is a very common alternative in many P1 languages.
the 19th and early 20th century, respectively. Luckily, all these languages were documented early on by Franz Boas, allowing us insight into their syntax. Still, I have been able to confirm P1 only for Lower Chinook, for lack of available data on the other varieties.

(6.36) Clatsop-Shoalwater Chinook [cch, Chinookan, source: Boas 1911: 647]

\[
tgi-gE\, n\, \text{\text{"n\, xaut\, t\,\, k\,\, an\, \,\, t\,\,\, l\,\,\, Em\,\, \,w\,\, \,\,\, E\,\, \,ma}}
\]

3p>3s-watch m-soul pl-ghost

'The ghosts watch the soul.'

Boas (1911: 646–7) argues that nominal arguments do not form part of the core sentence but are to be understood as appositions instead. However, he provides no arguments for this analysis and since the present study is only concerned with the ordering of nominal arguments and their predicate (cf. Chapter 2), we can still recognize Chinookan as P1. The common usage of P1 sets Chinookan firmly in the Pacific Northwest Linguistic Area, alongside such consistently P1 languages as Salish, Wakashan and Chimakuan. Because of intense contact with Salish languages (cf. Silverstein 1974), it is quite possible that P1 in the small Chinookan family was borrowed from the areally dominant and consistently P1 Salish.

Note that the pidgin Chinook Jargon does not follow regular P1 but SPO instead, despite its usage by speakers of various P1 languages (Thomason 1983: 844).

6.1.13 Kuliak

Ik, Soo, and Nyang’i, the three languages of this small family in the northeastern corner of Uganda bordering Kenya, agree on a basic PSO order, and this can readily be confirmed by published linguistic literature such as (6.37)–(6.39).

(6.37) Ik [ikx, Kuliak, source: Schrock 2017: 569]

\[
\text{Áts’\, á\, n\, \text{\text{"k\,\, k\,\, k\,\, k}}}
\]

gnaw.3s dog.nom bone-acc

'The dog gnaws the bone.'

(6.38) Soo [teu, Kuliak, source: Carlin 1993: 113]

\[
t-\, \text{\text{\text{"g\,\, a\, i\,\, a\, r\,\, a\, n\,\, k\,\, u\,\, d\,\, t}}}
\]

pfv-shoot soldiers dog-pl

'The soldiers shot the dogs.'

(6.39) Nyang’i [nyp, Kuliak, source: Beer 2017: 175]

a. \[
\text{kenwaki} \, \text{soat-\, \text{\text{\text{"t\,\, n\,\, a\, o}}}}
\]

kick bee-lv 1s.poss

'The bees are killing me!'

b. \[
\text{ajisa} \, \text{mersi} \, \text{\text{\text{"get-\, \text{\text{\text{"ik}}}}}}
\]

eat leopard goat-pl

'A leopard is eating the goats!'
As (6.37) shows, objects are overtly marked in Ik by an accusative case. (This marking was lost in Soo.) Subjects can be placed before the verb, but they are then exceptionally marked by the accusative case (Serzisko 1989: 388). This pattern is obligatorily used in the subjunctive, relative clauses, and cleft constructions.

The Kuliak languages are commonly grouped within Nilo-Saharan, but even for the ones who agree with this speculative group, its precise external relationship is subject to debate: Bender (1989) raised doubts that Kuliak is related to the areally dominant (but likewise contested group) Eastern Sudanic group, and it is since rarely included. Bender also notes that Kuliak languages have been strongly influenced and are being replaced by Nilotic languages. It is thus possible that P1 in this small family derives from contact with the P1 Nilotic group, especially since SPO ordering is regularly used in some types of embedded clauses.
Language families with P1 in a majority of languages

Figure 6.13: P1 in the Kuliak language family
6.1.14 Tsimshianic

The three Tsimshianic languages, Gitxsan, Nisga’a, and Sm’algyax, form the northern end of the Pacific Northwest P1 area at the border between British Columbia and southernmost Alaska. Tlingit, Eyak, and the Athabaskan languages to the north are predicate-final. Tsimshian languages follow P1 rather strictly, with the ergative and absolutive arguments following in this order.

(6.40) Sm’algyax (Coast Tsimshian) [tsi, Tsimshian, source: Dunn 1995: 60]

\[
yag\text{wa}=t \quad ni[t]\text{s-da} \quad ts’uut’s-a \quad laalt
\]

PRS=ERG see=ERG  bird=ABS worm

‘The bird sees the worm’
Language families with P1 in a majority of languages

(6.41) Gitxsan [git, Tsimshian, source: Hunt 1993: 23]

\[\text{stil-ə=s} \quad t=\text{John} \quad t=\text{Peter}\]

accompany-\text{erg}=\text{CASE CONN-J. CONN-P.}

‘John accompanied Peter.’


\[\text{yukw-t} \quad \text{giba-[t]-s} \quad [t] \quad \text{Mary} \quad t \quad \text{Lucy}\]

prg-3 wait.for-[3]-\text{erg} DET M. \quad DET L.

‘Mary is waiting for Lucy’

Hunt (1993: 56) describes any other order as ungrammatical for Gitxsan, although preverbal NPs do occur in questions and focus constructions (Hunt 1993: 57–8, Tarpent 1989: 26). Noun phrases can also function as predicates and are in this function placed in sentence-initial position. Oblique arguments and adverbials cannot break the Predicate-Ergative-Absolutive sequence, but always follow it.

Reconstruction of P1 to the Proto-Tsimshianic level suggests itself, although contact from the southern P1 neighbor Wakashan cannot be excluded, as all Tsimshianic languages are in close proximity.

6.1.15 Coosan

This small family of the southern Oregon coast consists of two now extinct languages, Miluk and Hanis, which, although spoken in a small contiguous area, are not particularly closely related to each other: Pierce 1965 found a lexical similarity of only 26% between them. Doty 2012 goes as far as to suggest that the similar features are due to contact, and that the two languages do not share common ancestry. He argues that Miluk shows a peculiar affinity to the Salish family, although this proposal is not commonly accepted.

Whatever the history of these two languages, they fit well into the Pacific Northwest Linguistic Area in terms of constituent order. Doty claims that in Miluk there is a strong tendency for the predicate to precede the two core arguments, where they are overtly expressed. The respective ordering of the two arguments after the predicate are described to be “based on nuanced discourse factors which are not yet clear” (Doty 2012: 38). In a statistical analysis of Hanis narrative speech, Dryer (1997) also comes to the conclusion that the constituent order is VS/VO, i.e. P1. This impression is furthermore confirmed by Kroober (2013: 106).

(6.43) Hanis [czz, Coosan, source: Kroober 2013: 106]

\[\text{kiyilú-w-it} \quad hə=\text{wixyi’s} \quad lə=\text{x=húʔmis}\]

\text{see-TR} \quad \text{ART=food} \quad \text{ART=erg=woman}

‘The woman saw the food.’

(6.44) Miluk [iml, Coosan, source: Doty 2012: 38]
Constituent order in the Coosan languages is not strict, however, and it is possible that the preference for P1 is just statistically emergent for non-syntactic reasons. WALS treats Hanis as lacking a predominant constituent order, and Mithun (1987) finds that the order is variable, or more precisely pragmatically defined, with newsworthy participants introduced earlier, also before the predicate. By contrast, Thompson and Kinkade (1990) list Hanis among the languages of Central Oregon with a preferred VOS word order, along with the Chinookan languages, Alsea, and Siuslaw.
Language families with P1 in a majority of languages

Figure 6.15: P1 in the Coosan language family
6.2 Language families with P1 in a minority of languages

6.2.1 Austronesian

Figure 6.16: P1 in the Austronesian language family

The Austronesian family is not only the second-largest language family in the world, but also the one with the most P1 languages by far, even though P1 languages only make up a minority of the entire family.

Blust’s (2013) extensive reference work on the Austronesian languages includes a useful overview on basic constituent order across different areas of Austronesia. His findings are replicated here as (6.3):

It shows that predicate initiality is very consistent in three main areas of the Austronesian world: Taiwan and the Philippines, Madagascar, and Polynesia. P1 in these areas is exemplified in (6.45)–(6.48) with Bunun (Taiwan), Hiligaynon (Philippines), Malagasy (Madagascar), and Hawaiian (Polynesia):

(6.45) Bunun [bnn, Bunun, source: Huang 2018: 50]
\[
\text{damu-un}=\text{ tamu’ungan}=\text{ tia a ma’utung}=a \\
\text{catch-PF}=\text{GEN police}=\text{DET.N.SBJ NOM thief}=\text{DET.NOM}
\]
‘The police caught that thief.’

(6.46) Hiligaynon [hil, Philippine, source: author’s data]
\[
\text{Gin-káon ní= nánay ang}=\text{ páhò }=\text{mo.} \\
\text{PF.PFV-eat GEN= mother SBJ= mango }=\text{2S.GEN}
\]
‘My mother ate your mango.’

\[
\text{Fafán’ ny zanak-ày ny éfitrâno} \\
\text{swiped ART child-1PE.GEN ART room}
\]
‘The room is swiped by our children.’ = ‘Our children swipe the room.’
Table 6.3: Basic constituent order in Austronesian languages according to Blust (adapted from Blust 2013: 461–2).

<table>
<thead>
<tr>
<th>Area</th>
<th>Verb-initial</th>
<th>Verb-medial</th>
<th>Verb-final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taiwan</td>
<td>most languages</td>
<td>one or two</td>
<td>none</td>
</tr>
<tr>
<td>Philippines</td>
<td>all or nearly all</td>
<td>none?</td>
<td>none</td>
</tr>
<tr>
<td>Borneo</td>
<td>some in north</td>
<td>most elsewhere</td>
<td>none</td>
</tr>
<tr>
<td>Madagascar</td>
<td>all dialects</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Mainland SE Asia</td>
<td>none</td>
<td>all</td>
<td>none</td>
</tr>
<tr>
<td>Sumatra</td>
<td>a few in north</td>
<td>most elsewhere</td>
<td>none</td>
</tr>
<tr>
<td>Java-Bali-Lombok</td>
<td>none</td>
<td>all</td>
<td>none</td>
</tr>
<tr>
<td>Sulawesi</td>
<td>a few in north</td>
<td>most elsewhere</td>
<td>none</td>
</tr>
<tr>
<td>Lesser Sundas</td>
<td>a few</td>
<td>nearly all</td>
<td>none</td>
</tr>
<tr>
<td>Moluccas</td>
<td>few or none</td>
<td>all or nearly all</td>
<td>none</td>
</tr>
<tr>
<td>New Guinea</td>
<td>none</td>
<td>many</td>
<td>many</td>
</tr>
<tr>
<td>Bismarcks</td>
<td>none</td>
<td>all</td>
<td>none</td>
</tr>
<tr>
<td>Solomons</td>
<td>a few</td>
<td>most</td>
<td>a few</td>
</tr>
<tr>
<td>Vanuatu</td>
<td>none</td>
<td>all</td>
<td>none</td>
</tr>
<tr>
<td>New Caledonia</td>
<td>none</td>
<td>all</td>
<td>none</td>
</tr>
<tr>
<td>Micronesia</td>
<td>one or two</td>
<td>most</td>
<td>none</td>
</tr>
<tr>
<td>Fiji-Polynesia</td>
<td>all</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

(6.48) Hawaiian [haw, Polynesian, source: Elbert and Pukui 1979: 53]

_Ua= make ke= keiki iā= Pele._

_Pfv= die Det= child nsbj= P._

‘The child was killed by Pele.’

Note in the examples above that many Austronesian languages make frequent use of ergative alignment, where the subject (or pivot) position is filled by the semantic patient. Although the sentences above may appear like passive sentences, they are neither structurally nor semantically marked but instead represent regular transitive clauses.²

It is not unusual for P1 Austronesian languages to place a topic phrase in front of the predicate, especially when it is contrastive. These are often overtly marked by a particle, which takes the form of _ga_ in Atayal or _ay_ in Tagalog, though they do not always require that. In direct translations from the major contact languages English and Chinese, as well as in formal settings such as in presentations or in broadcasting, this construction is often overly employed, giving the impression that this is the norm. However, P1 order is predominant in conversation, and it is also the educational standard for all the languages of Taiwan and the Philippines, Madagascar, and Polynesia. Furthermore, focused arguments are fronted in what resembles a cleft construction (6.49).

(6.49) Hiligaynon [hil, Philippine, source: author’s data]³

³However, the Hawaiian example (2.7) includes a particular verb type known as “loʻa verbs,” which are intrinsically non-accusative.
Páhò  ang= gin-káon  ni= nánay.
mango  SBJ= PF.PFV-eat  GEN= mother

‘Mango is what Mother ate.’

However, this construction is not necessarily aberrant from the general P1 structure, because it can be argued that the fronted argument represents the predicate, as any content word can function in predicate or argument position.

Even though there is not yet full clarity concerning the number of primary branches in the Austronesian family, there is general agreement that the indigenous languages of Taiwan ("Formosan languages") represent several (nine or ten in Blust’s approach, Blust p.c.) primary branches that are not found outside of Taiwan, while all languages outside of Taiwan form one subgroup, namely Malayo-Polynesian. Thus, it is not difficult to recognize Taiwan as the homeland of the Austronesian family. Combined, Formosan languages exhibit several phonological archaisms that are not found in Malayo-Polynesian. In terms of morphosyntax, Formosan languages are fairly uniform, showing a multiple voice system, which is otherwise found only in the Philippines and in Madagascar. Given the phylogenetic and geographic distribution of this typologically unique Austronesian alignment, its reconstruction to the highest level is self-evident. Interestingly, there is an almost perfect correlation of Austronesian-type alignment with P1. All languages of the Philippines, which are very conservative in their morphosyntax (as well as phonology), are P1 without a single exception to my knowledge; and the same picture prevails in Taiwan, though the marginalization of the indigenous languages by Chinese there has left a syntactic trace on some languages: in three out of 17 Formosan languages that survived into the 20th century, one can observe some deviation from the general dominance of P1: 1) Thao predominantly shows SPO word order, especially in sentences elicited from Chinese, but also commonly in narratives (cf. Li 2011), though P1 appears to be an alternative unmarked order; 2) Saisiyat follows SPO even more regularly, though P1 is mentioned by Ye (1991, 2018) as an alternative constituent order, but I have only encountered intransitive sentences that fit that template; however, P1 is predominant in older texts (Ogawa and Asai 1935: 36), demonstrating that SPO developed from an alternative marked construction to the standard, probably under Chinese influence (Ye 1991); 3) Pazeh/Kaxabu most commonly shows SPO order, though P1 is also used, apparently without any special pragmatic effect. These three Formosan languages are all native to the densely populated northwest of Taiwan, and they are drastically outnumbered by Chinese even in their ancestral land. They also count among the most severely endangered Formosan languages, with Pazeh/Kaxabu and Thao being moribund/extinct, and Saisiyat at a speaker number of around 4000. Two other Formosan languages, Basay and Luilang (Ketangalan) became extinct early in the 20th century, but I have been unable to find proper documentary sources on them. The remaining twelve Formosan languages, representing seven of the ten/eleven primary branches of the Austronesian family (in Blust’s classification), have a strong preference for P1. Considering the internal structure of the Austronesian family and this strong bias toward P1 in Taiwan, Blust (2013: 417) draws the uncontroversial conclusion that Proto-

88
Austronesian “almost certainly” was P1.

The predominance of P1 is even stronger in the Philippines, whose languages arguably form one major sub-branch of the Malayo-Polynesian branch (cf. Zorc 1978, 1986, Blust 2005, 2019), or several primary sub-branches of Malayo-Polynesian (cf. Reid 1982, Ross 2005). Glottolog recognizes 170 languages spoken within the political borders of the Philippines, and I have found no counterexample to the general P1 order (besides Sangil, for which see below). P1 is so much so the standard in Philippine languages that many descriptions do not bother mentioning this basic constituent order, although it is directly evident in example sentences. Even the Spanish-based creole varieties known by the names Chavacano and Ternateño place the predicate in clause-initial position, and as such they are unique in arguably being the only P1 creoles in the world.3 Interestingly, the Sangiric languages of northern Sulawesi, which Blust includes in his Philippine group but—except for Sangil on the Sarangani Islands—lie just south of the Philippines within the geographical borders of Indonesia, follow the areal pattern of Indonesia with SPO.4 This correlation of political borders and syntactic structures may be representative of the Malay sphere of influence: the national language of Indonesia, common medium of instruction and widely understood throughout the country, follows a strict SPO order (see below); by contrast, the Philippines has a laxer language policy, and the commonly taught but not universally understood Tagalog follows P1 syntax.

Beyond the conservative languages in the adjacent regions of Taiwan and the Philippines—as well as the outlier Chamorro (see below), which represents a primary branch of Malayo-Polynesian—P1 was mostly replaced by SPO order in Insular Southeast Asia. This is especially true of the languages of Indonesia. However, there are exceptions even within those areas, most notably the north coast of Sabah on the island of Borneo, and Sulawesi, which are both in immediate proximity to the Philippines. A notable P1 outlier within Indonesia are the Batak languages of Northern Sumatra, which are close to the Malayan epicenter of the Malacca strait (see below). Farther west yet, Malagasy, the national language of Madagascar off the southeast coast of Africa, has consistent P1 syntax. The origins of Malagasy can be traced to the Barito River in Southern Borneo, but this connection to the heartland of Indonesia, where SPO predominates now, dates back hundreds of years. Besides these few exceptions, practically all languages of Indonesia have not only abandoned P1 syntax, but also simplified the original Austronesian multiple voice system to a symmetrical voice system contrasting only agent and patient voices (similar to a European-type active-passive system). This is illustrated in (6.50)–(6.51), which also shows that the case marking found in the conservative Austronesian languages of Taiwan and the Philippine was likewise lost.

3Another Pacific creole linked to Austronesian, Pidgin Fijian, appears to have followed a P1 syntax in the recorded past, but later shifted to SPO. Virtually all creole languages in the world follow SPO order, even the ones with a lexifier of conflicting constituent order, such as Pidgin Hawaiian (Parkvall and Bakker 2013: 41) and Chinook Jargon (Thomason 1983: 844).

4There are claims for some Sangiric languages to follow an alternative P1 order (Kembuan et al. 1989: 111;131 for Talaul; Utsumi 2012: 107 for Bantik). However, these claims are not substantiated in the example sentences and texts, at least not for transitive sentences with nominal arguments.
Language families with P1 in a minority of languages

(6.50) Malay [zsm, Malayic, source: author’s data]
      man that AV-wash buffalo=3s.gen
      'That man washes his buffalo.'
   b. *Kerbau=nya di-cuci (oleh) lalaki itu.*
      buffalo=3s.gen PV-wash (by) man that
      'His buffalo is/was washed by that man.'

(6.51) Penan [pne, North Bornean, source: Mortensen 2021: 130]
   a. *Lemulun ineh maman kerbau=neh*
      man that AV.feed buffalo=3s.gen
      'That man is feeding his buffalo.'
   b. *Kerbau=neh p<in>an lemulun ineh*
      buffalo=3s.gen <PV.PFV>feed man that
      'That man fed his buffalo.'

Given that this broad-scale syntactic innovation coincides geographically with the Malayan sphere of influence, in particular the Sri Vijayan and Javanese empires (whose histories are interwoven), it is quite likely that Malay as the dominant trading language—whose influence, albeit in a weaker form, reached as far as northern Luzon—is the origin of this development. This assimilative process may still be continuing to this day with the growing dominance of Indonesia’s national language Bahasa Indonesia.

Since Malay is attested through manuscripts reaching back centuries, the syntactic shift from an older P1 to the modern SPO order is historically documented, and it turns out to be not all that old. Compared to modern Malay varieties, earlier Malay sources from as recent as the early 19th century show that Malay used to be much more flexible in its constituent order: although the agent voice strictly demanded the now common SPO, the at the time much more commonly employed patient voice allowed free ordering of the constituents, including common P1 (Hopper 1987, Cumming 1995). Hopper specifies that the P1 clauses serve the function of “narrating events and stating facts which are crucial to understanding the development of the discourse”, i.e. P1 clauses relate the foreground, as opposed to subject-initial clauses that supply background information (Hopper 1987: 461–2). Furthermore, Hopper argues that P1 clauses in early Modern Malay rank higher on the transitivity hierarchy (cf. Hopper and Thompson 1980). It is only later in the 19th and 20th century that a syntactic shift is apparent: the agent voice encroached onto the functional space of the patient voice, thus establishing itself as the unmarked norm bleached of pragmatics, and the now marginalized patient voice assimilated its constituent order to the dominant agent voice, whereby P1 was eliminated in Malay. In fact, it seems that Indonesian Malay may have led this syntactic shift rather

---

5Similar statements have been made for other P1 languages, e.g. Biblical Hebrew (Givón 1977) or Yagua (Payne 1993).
than Peninsular Malay, which Hopper attributes to its heightened influence from European languages, namely from Dutch. It is possible that this shift was further driven by the high number of second-language Indonesian speakers in multi-ethnic Indonesia.

Farther east, around the island of New Guinea, P1 gave way to predicate-final constituent order in many cases (though predicate-medial order is also not uncommon). This syntax is very unusual in the Austronesian family and not attested in other areas; it is undoubtedly attributable to influence from Papuan languages, which are predominantly predicate-final (Reesink and Dunn 2018: 950). The history of Micronesia is far more difficult to reconstruct than that of Indonesia, due to the lack of historical documentation, but the region is mostly dominated by SPO order, besides the noteworthy exceptions of Chamorro, Yapese and Palauan on the far western edge of Micronesia. Chamorro derives from a very early migration from the Philippines around 1500 BC and forms an independent branch within Malayo-Polynesian (Blust 2000); its syntax and morphology is quite comparable to Philippine languages, and it has no affinity to other Micronesian languages. Likewise, Palauan forms an independent offshoot from Malayo-Polynesian; its syntax is POS, but its voice system is symmetrical, unlike Chamorro or the languages of the Philippines. Finally, Yapese is nested within Oceanic, but has no close affinity to the other Micronesian languages; its syntax is PSO, but it appears to lack a voice system altogether.

(6.52) Palauan [pau, Malayo-Polynesian, source: Nuger 2010: 23]

\[ Te=ki\text{lang} \ a \ \text{rokui} \ \text{el} \ \text{ringngo} \ a \ \text{rengelek} \ \text{er} \ a \ \text{elii} \]

3P=ate DET all LNK apples DET children POSS DET yesterday

'The children ate all the apples yesterday.'

(6.53) Yapese [yap, Oceanic, source: Jensen 1977: 268]

\[ Kea \ \text{guy} \ Tamag \ Tinag \]

\[ \text{pfv.3s} \ \text{see} \ Tamag \ Tinag \]

'Tamag saw Tinag.'

Besides Taiwan and the Philippines, and Malagasy, Polynesia represents the third center of P1 in the Austronesian world. Virtually all languages located within the area of the Polynesian triangle—even slightly wider still, all languages within the Central Pacific linkage, including Fijian varieties (cf. (2.6) in Chapter 2)—are clearly P1. A marginal exception to this generalization is Rotuman, located north of Fiji and thus right on the border of the Polynesian triangle: it follows SPO order, although the possibility of PSO order in subordinate clauses still documents its earlier syntax. Polynesian languages outside this core area, the so-called Polynesian outliers in Melanesia, are overwhelmingly SPO, though some of them, such as Tikopia (6.54), and West Uvean (6.55), have retained P1, and others have the option of a marked construction P e SO, where the transitive subject is marked by the preposition e, not unlike Western Polynesian languages like Samoan.

(6.54) Tikopia [tkp, Polynesian, source: Early 1981: 121]
Language families with P1 in a minority of languages

\[
\begin{align*}
ka \ sau \ rei & \ ko \ Tafito \ te \ mei \\
pst & \ take \ then \ foc \ T. \ \ art \ breadfruit \\
\text{‘Tafito grabbed the breadfruit.’}
\end{align*}
\]

\[\text{(6.55) Fagauvea = West Uvean [uve, Polynesian, source: Djoupa 2013: 227]}\]

\[
\begin{align*}
\text{No} \ & \ teleki-na \ de \ fanaunga \ gi \ dingani \ de \ lu\ matua \\
\text{prosp} \ & \ carry-tr \ spec \ children \ \ obl \ school \ \ spec \ two \ old \\
\text{‘The parents will carry the children to school.’}
\end{align*}
\]

Polynesian languages do not have a complex (or even symmetrical) voice system; they typically mark non-subject constituents (whether direct objects or adjuncts) with an obligatory preposition \(i/ki\). Subjects can be marked with a preposition \(ko\) (ʻo in Tahitian or Hawaiian), but their post-predicate position alone identifies their syntactic role.

Finally, P1 is also widely spread in New Caledonia and the Solomon Islands, especially in the Northwest Solomons, where it is indeed dominant (contrary to Blust’s overview of Austronesian constituent order in (6.3) above).

\[\text{(6.56) Hoava [hoa, Northwest Solomonic, source: Davis 2003: 185]}\]

\[
\begin{align*}
\text{Vagi-a} & \ sa \ kaleqe \ sa \ leboto \\
take-tr.3s & \ art.sg \ old.woman \ art.sg \ machete \\
\text{‘The old woman took the machete.’}
\end{align*}
\]

\[\text{(6.57) Roviana [rug, Southeast Solomonic, source: Corston-Oliver 2002: 491]}\]

\[
\begin{align*}
\text{Seke-i-a} & \ e \ Zima \ se \ Maepeza \\
hit-tr-3s & \ pers \ Z. \ \ abs \ M. \\
\text{‘Zima hit Maepeza.’}
\end{align*}
\]


\[
\begin{align*}
\epsilon & \ \text{āli-} \ \text{ā-li} \ \ mw\ \ ω\ \ p\ \ \text{āpulip} \\
3s & \ see-tr \ art.ntr-def \ house \ subj \ art.nfem-def \ man \\
\text{‘The man saw the house.’}
\end{align*}
\]


\[
\begin{align*}
I & \ kaza-ni \ Maxat \ o \ pwirip \\
3s.sbj & \ pinch-tr \ M. \ \ sbj \ budgerigar \\
\text{‘The budgerigar pinched Maxat.’}
\end{align*}
\]

According to Evans and Palmer (2011: 489–90), almost all of the 38 languages on Bougainville are P1, with only the three Mono-Uruvan diverging as a result of contact with the Papuan languages of the area.

We can derive from the current distribution of P1 that Proto-Malayo-Polynesian had retained the P1 order of Proto-Austronesian, as P1 is amply reflected in its extremely far daughters in the Philippines, Madagascar and Polynesia. However, it remains a matter of
Language families with P1 in a minority of languages

debate whether Proto-Oceanic, which gave rise to all the Austronesian languages of the Pacific, was P1. Even though P1 is consistent in the Polynesian subgroup, other subgroups are less consistent. However, given the wide distribution of P1 in the Pacific even outside of Polynesian—notably Yapese, Fijian, the Northwest Solomonic languages and several New Caledonian languages, which all represent distinct subgroups of Oceanic that are moreover widely separated geographically—diffusion of P1 through intense language contact appears unlikely. Hence, it seems quite reasonable to assume inheritance of P1 in several branches of Oceanic from their common ancestral language, rather than repeated independent development of a typologically as uncommon a trait as P1. Reconstruction of P1 to the Proto-Oceanic level is supported by Lynch et al. (2002: 86). Arguments for the reconstruction of SPO often refer to the fact that the subject clitic precedes the verb in most Oceanic languages, but as explained in Chapter 5, reconstruction of clausal syntax should not draw from the positioning of clitics and affixation. A further piece of evidence for the reconstruction of P1 in the Melanesian Oceanic languages comes from Kuot, a non-Austronesian language on New Britain, which appears to have adopted P1 from its Austronesian neighbors in the distant past, prior to the latter developing predicate-medial structure themselves (cf. §6.3.4).

In conclusion, P1 within the Austronesian language family is a retention from its earliest stages. Yet, the majority of modern languages has developed SPO constituent order.

### 6.2.2 Afroasiatic

This large family of Northern Africa and the Horn of Africa has four extant branches: Semitic, Berber, Chadic, and Cushitic. Among these, all but Cushitic feature P1 in at least some languages. While P1 is represented only in a minority of Semitic (in various Arabic varieties) and Chadic (in the Biu-Mandara subgroup), it is dominant only in the Berber branch. Hence, predicate initiality is definitely not prevalent nowadays across the whole family. However, the picture was likely very different in the past. Afroasiatic has the longest recorded history of any language family, allowing us some direct insight into ancient languages of this family. Quite unfortunately, though, some branches of this family (especially Chadic) are so sparsely documented even nowadays that some researchers cast doubt on any reconstruction to the highest level of the proto-language. Still, given the long-standing written tradition covering several branches of the Afroasiatic family, the reconstruction of its syntax is arguably as accessible as that of any language family, despite its impressive time depth of probably more than 10,000 years (Ehret 2002).

One major branch of Afroasiatic is Semitic. Although Modern Hebrew and many

---

6The Omotic group of southwestern Ethiopia is traditionally understood as part of Afroasiatic, either as a sub-branch of Cushitic (Greenberg 1963), or as a separate branch (Fleming 1969, Bender 1971). It is unclear whether Omotic even represents a genetically coherent group. More recently, however, doubts have been raised about its inclusion in the Afroasiatic family altogether (e.g. Theil 2007). Glottolog does not treat Omotic as a part of Afroasiatic. Regardless of its classification, though, Omotic follows a predicate-final syntax like the neighboring Cushitic.
dialects of modern Arabic are SPO, Proto-Semitic was undoubtedly P1, because all ancient languages belonging to the Semitic group—Hebrew, Arabic, the Ancient South Arabian (Sayhadic), Aramaic, Ugaritic, Phoenician, Ge’ez, Amorite—exhibited PSO constituent order (cf. Huehnergard 2004, Huffmon 1965, Multhoff 2019, Birnstiel 2019). Examples (6.60)–(6.62) demonstrate P1 order in Biblical Hebrew, Ugaritic, and Geez, respectively.

(6.60) Biblical Hebrew [hbo, Semitic, source: Goldfajn 1998: 92]

\[
\text{wayyibr\` } e:\text{\`oh\' } h\text{-\`ad\`am b\`asalm\`o}
\]

created God ART-man in.his.image

‘God created man in his image.’

(6.61) Ugaritic [uga, Semitic, source: Tropper and Vita 2019: 500]

\[
t-\text{\`su } i-l\text{-m } \text{ra}\text{-\`st-hm}
\]

3P.M-lift god-PL head.PL-3P.M

‘The gods lifted their heads.’


\[
s\text{\`arh-\` } a\text{-\`ngu\` } \text{bet}
\]

build-3S.M.PST king house

‘The/a king built the/a house.’
Of the major ancient Semitic literary languages, only Akkadian shows predicate-final order, and this can be explained through contact with the non-Semitic Sumerian, from which it adopted the cuneiform writing system, and which it gradually displaced during the Middle Bronze Age (Adelaar 2009). The closest relative of Akkadian, Eblaite, had quite flexible constituent order, also allowing for SOP and SPO, besides PSO (Streck 2011: 349).

Although it is often claimed that Modern Spoken Arabic has abandoned its original P1 structure, a closer look at dialectal variation reveals that P1 is still in use (cf. Brustad 2000, Chapter 10). Especially the Modern South Arabian languages (Jibbali, Soqotri, Hobyót, Bathari, Harsusi and Mehri) still retain PSO as an unmarked constituent order in transitive clauses (Simeone-Simelle 1997: 1101). A few P1 examples from different modern Arabic varieties are presented in (6.63)–(6.65).

(6.63) South Levantine Arabic [ajp, Semitic, source: Mohammad 2000: 7]
\begin{verbatim}
gābal ʔehmad muna
\end{verbatim}
met.3s.m Ahmed Muna
‘Ahmed met Muna.’

(6.64) Sanaani Arabic [ayn, Semitic, source: Abbas 2018: 58]
\begin{verbatim}
ʔkal-an sabir al-burtkān-ah
eat.PFV-3s.f Abeer DEF-orange-F
\end{verbatim}
‘Abeer ate the orange.’

\begin{verbatim}
daraba l-ʔawlaad-u l-bint-a
hit.3.m they-boys-3p.m.nom the-girl-acc
\end{verbatim}
‘The boys hit the girl.’

Modern Standard Arabic allows for both PSO and SPO orders. The former structure was called verbal sentence by medieval grammarians, while the latter was called nominal sentence. Traditionally, the former was unmarked, while the latter reflected topic prominence. But in contemporary Standard Arabic, both can be pragmatically neutral without focus on a particular argument, although SPO order is chosen if the subject is in focus, in which case the subject also receives a pitch peak (AlShammiry 1999: 6). PSO always signals broad focus. The two orders are also asymmetrical with respect to the morphology involved: SPO always requires the verb to show full agreement with the subject in gender and number, while in PSO the verb agrees only in gender, but exhibits the unmarked singular number. This asymmetry is illustrated in (6.66). The same pattern also holds for Modern South Arabian (Watson 2012: 252).

Clause-initial predicates are more commonly encountered in intransitive clauses across modern Arabic dialects. This is especially the case when the verb is presentational in semantics, i.e. when it introduces an (often indefinite) subject by encoding location, motion, or appearance (cf. e.g. Ahmed 1992: 35 for Egyptian Arabic, or Jullien de Pommerol 1999: 199 for Chadian Arabic). This pattern resembles the Romance languages, exemplified in (2.3a) in Chapter 2, but it shall not be discussed here any further because we are only concerned with transitive clauses.

95
Language families with P1 in a minority of languages


a. raʔ-a al-mudars-ūn aT-Tālib
see.pst-3s.m the-teacher-m.pl the-student.sg

b. *raʔ-u al-mudars-ūn aT-Tālib
see.pst-3p.m the-teacher-m.pl the-student.sg

c. al-mudars-ūn raʔ-u aT-Tālib
the-teacher-m.pl see.pst-3p.m the-student.sg

d. *al-mudars-ūn raʔ-a aT-Tālib
the-teacher-m.pl see.pst-3s.m the-student.sg

All: ‘The teachers saw the student.’

Thus, the PSO construction appears to show attrition, as its verb shows only partial agreement. On the other hand, one could argue that the PSO construction is unmarked with respect to SPO because the subject does not need to be explicitly marked on the verb in the former, whereas (redundant) full reference is required in the latter (Brustad 2000: 319). Whichever synchronic analysis one assumes for contemporary Arabic, PSO can be identified as the older construction, and the change from PSO order to SPO order in Arabic was noted as early as the 14th century by Ibn Khaldûn, who argued that fixed constituent order was the result of attrition in the case marking (Owens 1988: 270). He attributed this change to foreign contact.

Some spoken dialects of modern Arabic are explicitly contrasted to Standard Arabic, in that they do not have PSO as an option (except perhaps in set phrases such as sayings). Example of such dialects are Gulf Arabic (Al-Bahri 2014: 126) and Sudanese Arabic (Roset 2018: 274). Egyptian Arabic allows PSO, but only when heavy emphasis is placed on the verb (Edwards 1988: 183). One encounters statements in the literature that modern spoken dialects of Arabic generally differ from Standard Arabic in constituent order, in that they do not make use P1. This is often related to the loss of case marking, which is understood to lead to a reliance on constituent order for the identification of syntactic roles. However, many other dialects (especially South Arabian ones) retain both options, despite a lack of case marking. Brustad (2000: 318) even makes the bold claim that PSO is common in all varieties of Arabic, though she concedes that this would need to be investigated in dedicated studies. Furthermore, her understanding of constituent order may be broader than the criteria applied in this study. Where PSO is available, the restriction on number agreement described above may or may not hold; it does not in Moroccan Arabic (6.67), i.e. the verb fully agrees with the subject in person, gender, and number in both SPO and PSO (Benmamoun 2000).

(6.67) Moroccan Arabic [ary, Semitic, source: Benmamoun 2000: 10]

a. kla-w lə-wlad
eat.pst-3p the-children

b. lə-wlad kla-w
the-children eat.pst-3p
c. "kla lɔ-wlad
   eat.pst.3s the-children
   All: ‘The children ate.’

Generally speaking, most modern Semitic languages have moved toward SPO as the unmarked constituent order. However, predicate-final order is followed by all Southern Semitic (or Ethiosemitic languages), including for instance Amharic and Tigrinya, as the result of contact with Cushitic. Also, some of the easternmost Arabic dialects have shifted toward predicate-final syntax as the result of contact with Persian and Uzbek. The same process has occurred in Jewish varieties of Northeastern Neo-Aramaic, and Bohtan Neo-Aramaic under the influence of Kurdish (Khan 2018, Fox 2002: 107). The evident picture is that over the course of history, many modern Semitic languages departed from the original P1 syntax.

Egyptian, which forms another major branch of Afroasiatic, was PSO, and strictly so in its early stages (Hapembath 2014: 8).

(6.68) Ancient Egyptian [egy, Egyptian, source: Gardiner 1957: 330]
   jst gm-n hm-f r-pr pn m dbt
   pcl find-prf majesty-3s.m mouth-house this in brick
   ‘His Incarnation found this sanctuary in a brick.’

Loprieno (1995) notes that Egyptian, not unlike Semitic discussed above, showed a diachronic tendency to replace the original PSO with a SPO order (along with a change from fusional to analytic morphology). In its latest form of Coptic, the standard constituent order had shifted to SPO, as evident in the Sahidic translation of the Bible around 300 AD (Lambdin 1982: 122–3). The Berber branch of Afroasiatic has been treated as a unit in some linguistic literature, with statements about “the Berber language” in general. For instance, Loprieno (1995: 4) remarks that “[t]he unmarked order of the [Berber] sentence, which can be modified in presence of pragmatic stress, is VSO.” In terms of constituent order, the Berber languages are indeed fairly uniform in following PSO.

(6.69) Ghomara [gho, Berber, source: Mourigh 2015: 396]
   y-umor ḥmed leftus i urgaz-ahen atġam
   3s.m-send.pst Ahmed money to man constr=sg.ana yesterday
   ‘Ahmed sent the money to that man yesterday.’

(6.70) Tamasheq [taq, Berber, source: Heath 2005: 18]
   ṣnḥay-Ξen méd-dΞen élú
   see.perf.pos-3p.m.sbj man-m.pl elephant
   ‘The/some men saw the/an elephant.’

   t-issirit tambitṭiḥ irbatṭini-s daffar waṭṣuṣ
   she-washed the.woman clothes-her behind the.tent
   ‘The woman washed her clothes behind the tent.’
However, not all languages agree with this generalization; one notable exception is Siwi, for which Souag (2013: 216) asserts that SV order “strongly predominates in Siwi, to a degree unexpected in a Berber language.” Souag adds that the reverse order VS is possible, but the only examples he provides are of intransitive clauses. Furthermore, even the languages that agree in following a basic PSO constituent order differ in pragmatic and stylistic strategies, and in the statistical picture that emerges as a result of those (cf. Mettouchi and Fleisch 2010). A recurrent topic in the literature is the opposition of PSO and SPO orders, which generally both occur with high frequency. The general consensus is that SPO arises through a fronting process, which can have various motivations (Mettouchi and Fleisch 2010). Generally, the change of constituent order is accompanied by a morphological distinction: post-predicate subjects are marked in the construct case (état d’annexion), while pre-predicate subjects are in the unmarked absolute case (état libre), which is also the citation form. It can be argued that from a morphological point of view, the post-predicate subject is visibly integrated into the clause, while the pre-predicate noun is rather in the left periphery outside of the core clause. Yet, it is possible that in some languages the fronting process may be purely historical, that the resulting SPO has since developed into a pragmatically unmarked construction on a par with the PSO order. However, despite a few exceptions and some lacunae in documentation, Loprieno’s broad characterization of the Berber branch is confirmed by sources on various languages.

I have found no sources for a proposed reconstructed syntax of Proto-Berber, but given the prevalent PSO order in its daughter languages, one can safely reconstruct that same order for the earliest stages.

By contrast to Semitic, Egyptian and Berber, the two remaining Afroasiatic branches of Cushitic, Chadic (or three if Omotic is included) are not generally described as P1. Cushitic consistently follows SOP order (Appleyard 2012: 210), and Chadic has a strong preference for SPO constituent order (Williams 1989, Loprieno 1995). However, one subgroup of Chadic, Biu-Mandara, features a few P1 languages with a basic PSO order. I have been able to confirm P1 with data for Hdi, Lamang, Parkwa/Podoko, Wandala, and Sakun/Sukur. Glavda is also claimed to follow the same constituent order in Buba and Owens (2007: 670), but I have not been able to confirm this assessment with data.

(6.72) Hdi [xed, Chadic, source: Frajzyngier 2001: 249]
\[
\text{zá-ghú-zá} \quad \text{ɗwák} \quad \text{tá} \quad \text{mndrá-ní}
\]
\[
eat-ISO-eat \quad \text{termites} \quad \text{obj} \quad \text{bottom-3s}
\]

‘The termites have eaten its bottom.’

(6.73) Lamang [hia, Chadic, source: Wolff 1983: 228]
\[
yáyá Yágh \quad \text{ tà } \quad \text{Bòkò}
\]
\[
begat \quad \text{squirrel} \quad \text{obj} \quad \text{hyena}
\]

‘Squirrel begat Hyena.’

(6.74) Sakun = Sukur [syk, Chadic, source: Thomas 2014: 149]
Language families with P1 in a minority of languages

'da pəká kywi=j də́gővu já-rá
FUT gather meat=DET hyena come-CENTRIFUGAL
'Hyena will gather this meat away.'

All of these are located along the northern border between Nigeria and Cameroon, but they do not form a subgroup within North Biu-Mandara to the exclusion of the many SPO languages in the same area, such as Gidar, Bana, Jina, Huba, Marghi or Matal, to name just a few. The syntactic past of the Chadic branch is difficult to assess without more complete documentation, but Frajzyngier notes the following, without reaching a conclusion about reconstruction: “The grammar of Wandala may well hold the key to the understanding of historical changes in Chadic syntax. Most Chadic are subject-initial. Some languages in the Central branch are predicate-initial in all tenses and aspects, and some languages are verb-initial in the perfective aspect and verb-medial in the imperfective aspect. Since the nominal argument in Wandala follows the verb and the pronominal subject precedes the verb, this language may represent and intermediate stage between subject-initial and verb-initial languages” (Frajzyngier 2012: 9). While Frajzyngier may be on the right track, he does not suggest a direction of change. However, a change from PSO to SPO is typologically much more common, as demonstrated within the Semitic branch discussed above, and if we conclude that Proto-Afroasiatic was indeed P1 (see below), then we are tempted to explain P1 in the Biu-Mandara branch as a continuation of the inherited pattern rather than as a typologically rare innovation, entailing that Proto-Chadic would have been P1, as well. Such a proposal is highly controversial, though, because the P1 languages are a tiny minority within the huge Chadic branch with 204 entries in Glottolog, and even within the Biu-Mandara sub-branch with 81 entries. However, no alternative explanation is immediately evident: although these languages are all spoken in a relatively confined area and could in theory have adopted P1 from a single outside source, no such source is synchronically evident, as they are surrounded by other Chadic languages, by the likewise SPO Gumuz, and the SOP Saharan languages. The historical reconstruction of the Chadic languages certainly requires more work, which presumes more extensive documentation.

Despite the long-lasting literary tradition of Afroasiatic languages, the reconstruction of the proto-language is still very contentious, and it seems that syntax has not even been included in this discussion. However, considering that for at least three primary branches of this family—Semitic, Egyptian, and Berber—P1 can doubtlessly be traced back far into their history, and that P1 also appears in a fourth branch, Chadic, which is furthermore in the periphery of the Afroasiatic territory and not in touch with other P1 languages, and lastly considering that verb initiality is a cross-linguistically rare phenomenon restricted to only 11% of the world’s languages (cf. §4.2), all signs point toward P1 in the Afroasiatic proto-language, as well. Thus, P1 in the various Afroasiatic languages seems to be retention from the earliest proto-language. However, most modern languages have departed from this syntax.
6.2.3 Nilotic

Nilotic displays a mosaic of constituent orders (Dimmendaal 2011: 112). The language family is divided into three branches: Western, Eastern, and Southern. Especially the Western branch exhibits diversity, with SPO order occurring in the Southern Luo group, OPS in Northern Luo, SOP/OPS in Burun, and V2 in the Dinka-Nuer cluster (arguably a separate branch of Nilotic). Southern Nilotic has been described as heterogeneous, with the several varieties of Kalenjin all following PSO, but Datooga following SPO (cf. Rottland 1982, Nyombe 1996). However, Kießling (2007) disagrees with that characterization of Datooga and convincingly demonstrates that its basic constituent order is verb-initial. In contrast to that, Griscom (2019) most recently characterizes Datooga as SPO. It is possible that the disagreement between Kiessling and Griscom are due to dialectal differences, or

---

9 Hieda (1991) describes Nuer as a P1 language, but primary language sources describe the constituent order to be subject-initial (Faust and Grossman 2015: 21). An aspect marker may precede the subject, and that aspect marker may bound to a pronoun, but with an independent noun as a subject, the phrase-initial aspect marker shows inflection whatsoever, and hence its analysis as an auxiliary is problematic.
that they simply disagree on a theoretical level. In either case, Datooga is at least partially P1, and since no syntactic data are available on Omotik, the last, moribund member of Southern Nilotic, this branch as a whole can be described to make common use of P1.

(6.75) Kupsabiny (Sebei) [kpz, Southern Nilotic, source: Kawachi 2016: 57]
\[
\text{ka-kúút-ut-è yoomét kártáásit meesà}
\]
\[
PST.3\text{-blow-thither-from wind.NOM paper table}
\]
‘The wind blew the paper off the table thither.’

(6.76) Datooga (Barabayiiga-Gisamjanga) [tcc, Southern Nilotic, source: Kießling 2007: 179]
\[
\text{qá-bár máydá dêedə}
\]
\[
s3\text{-beat calf.NOM cow.ABS}
\]
‘The calf beat the cow.’

Eastern Nilotic is generally described to be consistently P1, with the sole only exception of Bari, which uses SPO order (Bender 1996: 39).

(6.77) Maasai (Maa) [mas, Eastern Nilotic, source: Tucker and Mpaayei 1955: 128]
\[
\text{E-tana-payie o-sikiria en-tasat}
\]
\[
3s\text{-drag-away.PST m-donkey f-old.woman}
\]
‘The donkey dragged the old woman along with it.’

\[
\text{ɛ̀-à-r̀lɔ̀t-a’ a-pɛ̀sɛ a-mòti̻}
\]
\[
3\text{-pst-wash-ITIVE girl pot}
\]
‘The girl has rinsed the pot.’

Overall, the most widely attested constituent order is P1: it is predominant in two (Eastern, Southern) of the three sub-branches. As for the conflicting Western Nilotic branch, Hieda (1991) proposes that SPO and V2 can be traced back to a common structure TOP-V-FOC, which in turn could be derived from an earlier P1 structure by means of fossilization of leftward topicalization (cf. Dik 1980). This theory is supported by the fact that P1 is more widely distributed in subordinate clauses of a number of languages that follow SPO order in main clauses: Dinka, Shilluk, Jur, and Anuak (Hieda 1991: 104). Hence, the reconstruction of P1 to the Proto-Nilotic level is quite likely, unless it could be demonstrated that P1 spread at a later stage through a supra-regional dominant P1 language, for which I have found no indication. In fact, Nilotic has been cited as the source of P1 borrowed into neighboring language groups, e.g. Surmic (§6.1.4).

6.2.4 Uto-Aztecan

Predicate initiality is found in a number of Uto-Aztecan languages, though only in the Southern branch of the family, and in particular in the Nahua branch (former Aztecan),
Language families with P1 in a minority of languages

Figure 6.19: P1 in the Uto-Aztecan language family

exemplified below in (6.79)–(6.83). Nahuatl is the best-known language of this family and one of the largest native languages of the Americas in terms of number of speakers, with over a million native speakers (and many monolingual speakers). Nahua varieties are found over a large territory from Central Mexico down to El Salvador, and in earlier times even on the Pacific coast of Nicaragua and Costa Rica. Accordingly, there is considerable variation among the closely related Nahua languages and dialects, and the degree of intelligibility varies widely among varieties of Nahuatl/Nahuat. Across the whole Nahua complex, P1 is very common and in many cases considered the basic constituent order.

(6.79) Eastern Huasteca Nahuatl [nhe, Nahuan, source: de la Cruz Cruz 2010: 20]

\[ ki\-ma\-kah\-ki \quad koyo\te \ nopa \ te\-tl \]
3.OBJ-hand-leave-PFV coyote DEM rock-ABS
'
The coyote threw the rock,'

(6.80) Michoacán Nahuatl [ncl, Nahuan, source: Robinson and Sischo 1969: 60]

\[ ki\-ko\-ci\-lti\-k \quad maria \ in \ \hat{s}o\-lul \]
3S.OBJ-sleep-CAUS-PRET M. the baby
'Mary put the baby to sleep,'

(6.81) Tlaxcala-Puebla Central Nahuatl [nhn, Nahuan, source: Foster 1996: 53]
Language families with P1 in a minority of languages

Western Durango Nahuatl [azn, Nahuan, source: Canger 2001: 148]

\[ \text{ki-pe:w-a-lta-a-ʔh} \quad \text{sowa-me?} \quad \text{in-teki-w} \]
\[ \text{it-begin-caus-caus-caus-PL} \quad \text{woman-pl} \quad \text{their-work-sg} \]

‘The women begin their work.’

(6.82)

Pipil (Nawat) [ppl, Nahuan, source: Campbell 1985: 103]

\[ \text{ki-ta:lih ne i-chaketah ne taka-tsin} \]
\[ \text{it-place the his-jacket the man-dim} \]

‘The little man put down his jacket.’

However, there is not universal agreement on the basic constituent order of Nahua; even within the same variety this has been the subject of heated discussion, with some authors arguing for a basic PSO and others for a basic SPO order. While dialectal variation should be considered a relevant factor, the dominance of one or the other type of constituent order also varies from speaker to speaker, as it reflects level of education, level of bilingualism with Spanish, and finally narrative setting, which can be somewhat wooden during fieldwork sessions. Of course, ascribing SPO order solely to Spanish influence is not correct, either, as SPO is a pragmatically marked alternative to perhaps all P1 languages (cf. Chapter 2). Rather, increased influence from Spanish makes SPO a more common and less marked strategy. Hill and Hill (2004) provide an insightful discussion into the socio-linguistic factors of constituent order in Nahuatl. Based on the actual variation within the language and between speakers and narrative events, but also applying different methodology and theoretical backgrounds, researchers reach different conclusions about basic constituent order.

Even though I have searched for resources on all 31 Nahua varieties listed in Ethnologue, I have found sources with accompanying to confirm the unmarked use of P1 for only seven varieties. Another six varieties have been claimed to follow a basic P1 order in the literature, but I have not been able to confirm that with actual data. For the majority of the remaining varieties, I have not found any data or description of constituent order, at all. To the extent of my research, only four varieties are explicitly described as following a basic SPO order without common usage of P1. These are the dialects of Tetelcingo (Tuggy 1979: 9), Northern Puebla (Brockway 1979: 146), Michoacán (Sischo 1979), and Southeastern Puebla (MacSwan 1998). Hence, the distribution of P1 within Nahua as presented in this study is to be taken with a grain of salt. Nahuatl is not only one of the largest languages in the Americas but also one of the longest attested, with written sources dating back to the early 16th century. Classical Aztec was clearly P1 (Steele 1976).

(6.84) Classical Nahuatl [nci, Nahuan, source: Launey 1979: 38]
Besides Nahua in its various dialects, P1 is also attested in Tepehuán, which is not closely related to Nahua, but is among the next southernmost languages in the Uto-Aztecan family after Nahua.\textsuperscript{10} However, Huichol and Cora, which are also spoken fairly far south (as far south as Jalisco), do not make regular unmarked use of P1, but rather follow a basic predicate-final order, though in a highly flexible manner (cf. Gómez 1999: 56 as well as Iturrioz Leza and López 2006: 114 for Huichol, and Casad 2001: 118–9 for Cora).

All Uto-Aztecan languages further to the north, reaching as far as Oregon and Idaho with Northern Paiute and Shoshone, follow a basic SOP order, though pragmatic and stylistic factors allow for such flexible re-ordering that many languages are described to have free constituent order (Campbell et al. 1986: 547). Even movement of the verb to clause-initial position is possible under these circumstances (cf. Hill 2005: 432–4 for Cupeño). Thus, the general picture is that most Uto-Aztecan languages are predicate-final, although the single largest and southernmost group, Nahua, is predicate-initial, along with Tepehuán, which is not closely related genealogically but in geographic proximity. Accordingly, Langacker (1977) reconstructs Proto-Uto-Aztecan as SOP.

There is general agreement in the field that the Uto-Aztecan family originated north and spread south, across all of what is now Mexico and into Central America as far south as Costa Rica. More precisely, the Urheimat may not have been in the northernmost reaches of the current Uto-Aztecan territory, but rather in the U.S. Southwest, as reconstruction of botanical vocabulary suggests that Proto-Uto-Aztecan was spoken in “a mixed woodland/grassland setting, in proximity to montane forests”, such as Central-Southern Arizona along the Sierra Madre (Fowler 1983: 231, Shaul 2014). Furthermore, the linguistic diversity is greater in the north than in the south: although Uto-Aztecan as a whole is commonly split into a Northern and a Southern branch, which are quite neatly split between the two countries of the United States of America and Mexico, only the Southern branch is universally accepted, whereas Northern Uto-Aztecan is lexically so diverse that it does not emerge as a unit in lexically based quantitative studies (Miller 1984, Haugen et al. 2020). Since P1 is an areal feature of Mesoamerica, this constituent order is commonly held to be an innovation of Nahua as it moved south into the Mesoamerica Linguistic Area, where they encountered speakers of P1 languages (Campbell 2016: 212).\textsuperscript{11}

In the Uto-Aztecan languages with standard P1, it is quite evident that this structure represents a relatively recent innovation, as those languages exhibit a typologically

---

\textsuperscript{10}Tohono O’odham (= Papago), the northernmost member of the Southern Uto-Aztecan languages, has also been described to follow a basic PSO order by some authors (e.g. Saxton 1982: 108), even though every ordering of constituents is grammatical (Zepeda 1983: 130, Payne 1987, Wing 2020: 46). According to Payne (1987), SPO is the most common order, followed by POS (23%) and PSO (14%).

\textsuperscript{11}The assumption of a northern origin of Uto-Aztecan was challenged by Hill (2001), who on the basis of a reconstructed suite of nine words for the cultivation of maize argued for a Mesoamerican origin of Uto-Aztecan; however, Hill’s approach is generally not accepted in the field.
unusual combination of traits, such as P1 constituent order but postpositions. (Predicate-initial languages almost exclusively make use of prepositions, rather than postpositions.) As P1 is a Mesoamerican areal feature, characteristic of the neighboring Totonacan, Otomanguean and Mayan families, areal diffusion into the neighboring Nahua suggests itself.

### 6.2.5 Arawakan

This major family of the South American continent shows astounding variation in constituent order. Predicate-initial, -medial, and -final orders (even the rare OSP order) are each found in multiple languages. P1 is found in several main branches (cf. Noble 1965), but SPO is much more common. A typology of constituent order is given in Aikhenvald 1999 (p. 98) and Campbell 2012 (pp. 273–4). The commonly cited P1 languages in this family are the Campan languages and Yanesha’ (Amuesha). The attentive reader may note that Campbell 2012 (p. 273) also lists Palikur as a VSO language, but both Green and Green (1972) and Launey (2001: 13, 2003: 57–60) describe it as following a SVO order, not mentioning P1 order.

![Figure 6.20: P1 in the Arawakan language family](image)

---

12 The attentive reader may note that Campbell 2012 (p. 273) also lists Palikur as a VSO language, but both Green and Green (1972) and Launey (2001: 13, 2003: 57–60) describe it as following a SVO order, not mentioning P1 order.
Language families with P1 in a minority of languages

demonstrates that the position of the subject in a transitive clause is regularly before the verb (Michael 2008: 345).

i-cheK-i-ro shirampari inchato
'The man cuts the tree.'

(6.86) Poyenisati [cot, Campan, source: Castillo Ramirez 2017: 35]
i-kapiok-ak-i-ro paho ro i-sheka aka-niki
'Pablo gathers his food here.'

Various other Campan varieties (e.g. Ashéninka Perené, Mihas 2010: 13) are described in (in some cases extensive) grammars to commonly use PSO order, but I have not been able to corroborate this claim by the standard of the present study. However, several authors argue with Wise’s analysis of PSO as the basic constituent order, pointing out the frequency of SPO. It should be noted that Campan varieties are closely related, with experts commonly recognizing two or three distinct languages (cf. Pedrós 2018). It is thus possible that the apparent differences in constituent order merely reflect differences in data collection and analysis among different authors, rather than actual differences in linguistic practice. Yanesha’ (Amuesha), the closest relative of the Campan group, is also confirmed for P1.

(6.87) Yanesha’ (Amuesha) [ame, Southern Arawakan, source: Wise 1992: 119]
rr-en-an atollop mam
eat-prg-obj chicken manioc
'The chicken is eating the manioc.'

Wise appears to suggest that the PSO order in Campa and Amuesha is an innovation: "while Piro is SOV it is not rigidly so and may be moving towards VSO or SVO order; the Campa languages and Amuesha are all VSO languages, but the latter has moved further in that direction than the Campa subgroup. Variations in constituent order are related to topicalization and focus of attention, and possibly to other aspects of discourse organization" (Wise 1986: 636–7).

Beyond these languages, Aikhenvald lists another nine Arawakan languages with P1. Crucially, these languages are spread across both the southern branch, to which the Campan languages belong, as well as the northern branch, making P1 a quite widely spread phenomenon in Arawakan. Further P1 languages in Southern Arawakan are Bauré, Terêna, which I have been able to confirm, and Ignaciano Moxo, for which I did not find sufficient data.14 The additionally cited P1 languages in Northern Arawakan are Garifuna,

---

13Michael also adds that although native consultants had no difficulty interpreting clauses with two free arguments, he did not find a single instance of such a clause in his corpus (Michael 2008: 345).
14Its sister language Mojeño Trinitario is described to follow SPO order (Rose 2014: 90).
Wayuu’ (Guajiro), Resígaro, Lokono, and Añun (Paraujano). I have been able to confirm P1 for Garífuna and Wayuu’. I cannot confirm P1 for Resígaro, which Seifart (2012: 371) claims to have an SOP basic constituent order, though Huber and Reed (1992: xx) cite both SOP and SPO as possible constituent orders, with a tendency toward the latter. Lokono does exhibit predicates in initial position, but only in intransitive clauses, while the agent precedes the predicate in transitive clauses (Patte 2002: 109). Thus, the basic constituent order of Lokono is SPO, without any indication of P1 in Patte’s publications (Patte 2002, 2008). Finally, the data for Añun were too scarce to allow assessment.

In addition to Campan and the other above cited languages, which are described to follow P1 as their basic constituent order, a number of languages are described to allow P1 as an alternative order. Thus, Mosonyi (2000: 645) specifically mentions PSO as an option in Curripaco and Piapoco, though their basic constituent order is SPO. P1 is confirmed for Piapoco, though in the order POS, which is described by Klumpp as a construction with “backed subject” and likened to right dislocation, although there does not seem to be any pragmatic effect associated with this construction (Klumpp 1990: 39). I found no data to corroborate P1 for Curripaco. It is quite possible that more Arawakan languages allow P1 in a similar fashion as Piapoco, though a more complete coverage of this would require extensive research into each language, which is beyond the scope of the present study. For Yine (Piro), which is often cited as an SOP language, Hanson refuses to determine a basic constituent order, arguing that the order is dependent on pragmatic, rather than syntactic factors: focused arguments precede the predicate whereas non-focused arguments follow it (Hanson 2010: 292). Hanson also provides examples of P1, with the respective ordering of the arguments being free, i.e. possibly ambiguous (6.88).

(6.88) Yine [pib, Southern Arawakan, source: Hanson 2010: 296]

r-hiyla-ta-na-na sɨwa-ne mhenokli-ne
3-kill-V.CL-DUR-3P anteater-PL jaguar-PL

‘The anteaters killed the jaguars’ or ‘The jaguars killed the anteaters.’

Although SOP is comparatively rare in the Arawakan family, Derbyshire (1986) argues that the proto-language followed this order. His argument, following Greenberg’s tradition, is based on the ordering of heads and dependents, namely Noun-Postposition, Possessor-Possessed, and Head Noun-Adjective. The first two of these represent head-final orders, which are often cited as more prevalent in predicate-final languages; while the last one contradicts this pattern and is discarded as statistically less telling (Derbyshire 1986: 558–9). Derbyshire’s arguments for his reconstruction are reprinted below:

“There are two general characteristics that suggest the change [in word order] is in the direction of a move away from an earlier SOV order. The first is seen in the constituent order currently found in the noun and adpositional phrases of nearly all these [8 Brazilian Arawakan] languages: POSTP, GEN-N, N-ADJ.

---

15Hanson gives examples of PSO and POS constituent order, but the examples do not conform to the parameters used for this study, i.e. the arguments are represented by pronouns or are indefinite.
The first two are clearly associated with SOV, and the N-ADJ order is almost as common with OV as with VO (...). Change of basic patterns in phrases is less rapid than in main clauses, and this supports the hypothesis that the change has been from SOV to SVO, rather than the other way around. The one VOS language T[eřëna], has clearly moved the farthest, and this is reflected in the phrase constituent orders, which are generally consistent with a VO language, although there are still vestiges of an earlier pattern associated with OV (at least one postpositional form, and one of the orders found in noun phrases, ADJ-N (...)).” (Derbyshire 1986: 558–9)

Note that Derbyshire based his argument on only eight languages (with Terêna being the only P1 language in his sample). By contrast, Campbell 1997 recognized the prevalence of SPO order in the family and argues that “[t]he [Arawakan] proto language probably had SVO order. SVO word order is found today in most of the family, with frequent VS for intransitive verbs.” (Campbell 1997: 179)

Considering the spread of predicate initiality over different branches, it would have necessarily developed independently in a number of cases, if the earliest stage had followed a SPO order, but this is statistically unlikely. Modern Arawakan languages are generally fairly free in constituent ordering. This applies to both the P1 languages, as well as to languages that have been described as following an SPO basic constituent order (cf. e.g. dos Santos 2006: 217–23 on Wapixana). Mosonyi (2000: 645) specifically mentions PSO as an option in Curripaco and Piapoco, though their basic constituent order is SPO. It is likely that this syntactic flexibility also applied to the proto-language. More likely than historical derivation from a rigid SPO syntax to P1 in multiple languages is a slight shift in tendency within a pragmatically sensitive syntax. More research on the history of the Arawakan languages and language contact in that area would be necessary to get a better understanding of the emergence of verb initiality in the Arawakan family.

### 6.2.6 Austroasiatic

The Austroasiatic language family fits well into the Southeast-Asian area, where languages generally display SPO order. This is true for Kradaï, Hmong-Mien, and also for Sinitic bordering to the north. The oft-quoted exception to this is the predicate-final Munda languages, whose aberrant clausal syntax is generally ascribed to influence from Dravidian languages, which were historically dominant in that region, and more recently the Indo-Aryan languages, both of which follow predicate-final syntax (Anderson 2003). However, P1 is scattered across a few languages in the Austroasiatic family. These languages belong to several subgroups: Khasic, Paluanic, Nicobarese, and Aslian. Especially the southernmost branches Nicobarese and Aslian show P1 in several languages. Examples from Semelai (Aslian) and Mūöt / Nankwari (Nicobarese) are presented in (6.89)–(6.90).

---

16However, some languages are described as being very rigid in their constituent order, e.g. the P1 Campan language Wayuu (Guajiro) (Zubirí Olza and Jusayú 2012: 21–2).
Language families with P1 in a minority of languages

Figure 6.21: P1 in the Austroasiatic language family

(6.89) Semelai [sza, Aslian, source: Kruspe 2004: 255]

a. $ki=buk0$ $la=kniek$ $hn=pintu0$
   $3s=open$ $nom=husband$ $acc=door$

b. $ki=buk0$ $hn=pintu0$ $la=kniek$
   $3s=open$ $acc=door$ $nom=husband$

Both: ‘The husband opened the door.’


$kal0? n0t c0-n$ $kamal0?$
steal pig my-nom thief

‘The thief stole my pig.’

The Nicobaric branch is smaller than the Aslian branch, with five languages distinguished by Glottolog (six if Shompen is included, whose genetic affiliation is contested), compared to fifteen languages in the Aslian branch. I can confirm P1 for all Nicobaric varieties that have available descriptive records; however, there are only two of these, namely Car and Central Nicobarese (Nancowry). By contrast, Kruspe et al.
(2015: 436) claim that P1 is mostly attested in the southern sub-branch of Aslian (such as Semelai above), though several languages of the North Aslian group (Chewong, Maniq, Jedek, Mah Meri) are also described to make common use of PSO ordering (though I could not always confirm those claims with supporting data in the literature). Other Aslian languages generally follow the SPO order typical for Austroasiatic languages, though the constituent order of these languages is described as quite flexible, and it is common to place the predicate in initial position in intransitive clauses. In Jahai, an alternative PSO order is employed to place focus on the subject, which is also additionally marked by an enclitic ka=. This is noteworthy, insofar as focused elements are generally moved in peripheral position of the sentence, most commonly in initial position, in languages around the world. Jahai thus shows a pattern quite contrary to the crosslinguistic norm.

Most recently and presently still ongoing, a research project at the University of Zürich led by Mathias Jenny has been dedicated to the distribution and history of P1 in Austroasiatic, and thus the discussion of Austroasiatic historical syntax deserves some detail here. Jenny et al. (2015: 59) confirm that P1 is the exception to the very strong SPO pattern in the Austroasiatic family: “The most frequently encountered constituent order in transitive expressions in [Austroasiatic] languages is AVP, that is, the agent precedes the main verb, the patient follows it. This is also the normal constituent order in the Old Mon and Old Khmer inscriptions, and is found throughout the eastern (non-Munda, non-Nicobarese) branches of [Austroasiatic].” Their in-depth survey confirms that P1 in main clauses only occurs in a small minority of Austroasiatic languages. However, they argue that P1 is more widespread if one digs beneath the surface of main clauses. For example, P1 is common in the Palaungic branch, but restricted to dependent clauses in such languages as Shwe or Rumai. (The only language in the Palaungic branch that exhibits P1 as an alternative to verb-medial order in main clauses is Wa, though different varieties show different constituent orders, cf. Yamada 2020: 147ff.). By contrast, the Nicobarese language Car generally follows PSO order in main clauses, but in dependent clauses, SPO is regular. As in other areas of the world, P1 order is also more common in intransitive clauses.

In the southern coastal Austroasiatic branches (Nicobarese, Aslian), P1 order could possibly be a result of contact with Austronesian languages, with which they are and have for a long time been in contact, which is also quite evident in lexical borrowing. The consistent SPO order of modern Malay and most other languages of the Indonesian archipelago is not all that old, as the transition from P1 syntax to a verb-medial structure is historically documented in Malay (cf. §6.2.1). It is possible, then, that P1 was adopted by Austroasiatic speakers at an earlier point of time, when the dominant Austronesian languages in the area were still P1, or otherwise it could have been borrowed more recently from a language in the area that to this day makes use of P1 order, such as Batak or Nias.

While this contact scenario can provide an account for P1 in Nicobarese and Aslian, it hardly explains the presence of P1 in the other Austroasiatic branches Khasic and Palaungic, which are far inland, and have not in any records been exposed to P1 languages (but only the verb-final Indo-Aryan and Sino-Tibetan languages of the region, as well as historically the verb-medial Ahom of the Kradai family). “A handful of languages are
exceptional in the [Austroasiatic] family in having verb initial basic word order. These include Pnar, Amwi and War (but not the closely related Khasi), Wa (but not other Palaungic languages), and the Nicobarese languages” (Jenny et al. 2015: 59). Examples (6.91)–(6.93) show basic P1 constituent order in northern Austroasiatic languages.

(6.91) Pnar [pbv, Khasic, source: Ring 2015: 1196]
\[tʃim \, u=bru \, \text{ka}=\text{wat} \]
\[\text{take} \, \text{M}=\text{person} \, \text{f}=\text{sword} \]
'The man took the sword.'

(6.92) War [aml, Khasic, source: Koshy and Wahlang 2011: 158]
\[liʔ-iːt \, u=\text{jən} \, ha-ga-\text{meri} \]
\[\text{pst-love} \, \text{3S.M-John} \, \text{acc-3S.F-Mary} \]
'John loved Mary.'

(6.93) South Wa [prk, Palaungic, source: Yamada 2020: 137]
\[\text{suat} \, ?\text{ai ghok lik} \, \text{kah} \, \text{gən} \]
\[\text{kill} \, \text{Ai} \, \text{neck} \, \text{pig} \, \text{by} \, \text{knife} \]
'Ai stabbed the neck of the pig with a knife.'

If P1 was indeed adopted by some southern Austroasiatic languages, this still would not explain the occurrence of that feature in the Northern branches of Khasian (Pnar, Amwi, War) and Palaungic (Wa). Thus, Jenny et al. (forthcoming) suggest that P1 can be reconstructed to the Austroasiatic proto-language. Though the contemporary evidence for this is rather marginal, precisely that is part of their argument: peripheral languages have a tendency to remain conservative because they are less affected by change resulting from higher internal diversification in large population groups, as well as influence from the native languages of L2 speakers (Næss and Jenny 2011). And within the syntax of any given language, more highly nested constructions are generally more conservative than highest-level syntax like constituent order (Bybee 2002, Aikhenvald and Dixon 2002, 2007). Hence, the wider application of P1 in subordinate clauses than in main clauses, observed especially within the Palaungic branch of Austroasiatic, may be an indication that P1 could have once been the preferred order in the affected languages, and perhaps in the whole branch. Jenny et al. recognize further hints of earlier P1 in verb-noun compounds of otherwise verb-final Munda languages (6.94).

(6.94) Verb-noun compounds in Munda languages
a. Sora [srb, Munda, source: Anderson 2007]
\[\text{ɲam-kid-t-am} \]
\[\text{seize-tiger-N.PST-2} \]
'tiger will seize you'

b. Kharia [khr, Munda, source: Peterson 2011]
All evidence, stretching from P1 basic constituent order in Nicobarese, Aslian, and (marginally) in Khassic and Palaungic, to P1 in subordinate clauses in (various) Palaungic languages, to verb-noun compounds in Munda, covers a third of the main branches of the Austroasiatic family, and cannot be easily dismissed. However, the claim to reconstruct P1 to the proto-level on this basis is daring, and certainly unconventional. One problem with this account is that while the independent clause structure is used as evidence to reconstruct P1 for Palaungic, the same method excludes Nicobarese from the list of evidence, because although it features P1 in main clauses, its dependent clauses are verb-medial (see above). Jenny et al. concede that Nicobarese may have undergone change to a verb-medial stage, only to re-borrow P1 from Austronesian at a later stage. As mentioned before, the Aslian branch is another likely candidate for syntactic change under Austronesian influence. This would reduce the evidence gathered by Jenny’s team to the marginal evidence of P1 in the three western-most branches of the family: Munda, Khassic and Palaungic, the last two of which may form a common branch within the family and thus share a common history to the exclusion of other languages within the family. With the information at hand, P1 in the Austroasiatic could quite possibly be an innovation that arose in the west of the Austroasiatic group. If these branches do not have a common genetic relationship to the exclusion of other families, it is conceivable that P1 (or just structures reminiscent of it) spread at a later stage through contact between distinct but related languages. The hypothesis of an ancient origin of P1 in Austroasiatic would receive more weight if additional evidence could be drawn from further branches, especially from the east, and in languages that do not show evident signs of Austronesian contact.

### 6.2.7 Indo-European

Predicate-initial syntax is well-known for the Celtic branch of the Indo-European family, within which P1 is consistent. Examples (6.95)–(6.99) illustrate this pattern in five of the six “modern” Celtic languages, representative of both primary branches, Brythonic and Goidelic. (Cornish became extinct at the end of the 18th century.)

(6.95)  Irish [gle, Goidelic, source: Bobaljik and Carnie 1996: 225]

\[
\text{Chonaic Seán an \ madra.} \\
\text{Chonaic Seán an \ madra.} \\
\text{saw J. the dog} \\
\text{I saw the dog.}
\]

(6.96)  Scottish Gaelic [gla, Goidelic, source: MacAulay 1992b: 167]

\[
\text{Thug Iain leabhar do Anna.} \\
\text{Thug Iain leabhar do Anna.} \\
\text{gave I. book to A.} \\
\text{Iain gave a book to Anna.}
\]
Figure 6.22: P1 in the Indo-European language family


\begin{align*}
&\text{Hug} \quad \text{yn saggýrt e laue urree.} \\
&\text{put.pst the priest 3S.M.PS hand on.her}
\end{align*}

‘The priest lay his hand on her.’


\begin{align*}
&\text{Gwelodd y bachgen y ferch.} \\
&saw.3s \text{ the boy the girl}
\end{align*}

‘The boy saw the girl.’

(6.99) Breton [bre, Brythonic, source: Borsley 1990: 82]

\begin{align*}
&\text{Sevel ar mogerioù o deus graet ar vasonerien.} \\
&\text{raise the walls have-3p done the masons}
\end{align*}

‘The masons have raised the walls.’

The only language that does not follow a basic PSO syntax is Breton. For a long time, Breton was consistently described as being P1 just like the other Celtic languages.
However, starting with Varin 1979 (cf. also Stump 1984), a number of researchers have argued that the basic constituent order is SPO. This was the onset of a debate that persists into the present. Varin’s argument is based on the observation that many earlier Breton writers make frequent use of SPO. She argues that the PSO construction represents a stylized unnatural (“chemical”) form of language, which is being pushed by intellectuals. Her argument has been rebutted by Timm (1989: 369) and Borsley and Stephens (1989), who both claim that clauses with SPO order are instances of fronting (cf. Chapter 2). However, it appears that this pragmatic effect has been bleached by frequent use, possibly under the influence of French (Heine 2006). Timm (1989: 371) provides statistics to show that SPO and PSO are both commonly employed orders, with approximately equal distribution, which is quite expected for a P1 language. However, Kennard (2018) points out that the finite element is never quite in initial position in the alleged PSO order, but that those instances represent a periphrastic construction with a verbal noun, e.g. a participle, in initial position, while the inflected auxiliary follows. In fact, main clauses with the finite verb in initial position are outright ungrammatical. Therefore, Kennard proposes that Breton constituent order should be properly characterized as V2. The two most common instantiations of the preverbal element are the subject, leading to SPO order, and a verbal noun, which leads to what for the purposes of this study is analyzed as PSO. In addition to that, though not decisive for the criteria of the present study, embedded clauses usually follow PSO order (Kennard 2018: 155). To summarize, although Breton differs from the other Celtic languages in constituent order in that it is less thoroughly P1, it still commonly places the predicate—including the finite element—in front of core arguments without any particular pragmatic highlighting.

P1 can be traced back far in Celtic history, as it is attested in Old Irish and Old Welsh sources (Mac Cana 1973). By contrast, it does not appear that ancient Continental Celtic languages like Gaulish made regular use of P1, though this difficult to ascertain due to the scarcity of documentation: these languages are only transmitted in short (and often fragmentary) inscriptions, but not in texts. From what little documentation of Gaulish is available, it has been suggested that it followed a (flexible) V2 constituent order, most commonly instantiated as SPO (Koch 1991). Koch (1989:169) ventures to claim that P1 was “rare to non-existent” in Continental Celtic. He further points out the common use of SPO in early records of Irish (alongside PSO), and even in Middle Welsh this order is attested, where it has been traditionally referred to as the “abnormal order” (Mac Cana 1973). Following Koch’s reasoning, P1 is a relatively recent development (of the last one thousand years) peculiar to Insular Celtic, where it spread through areal convergence (remember the anomalous constituent order in Breton), though this argument is based on scarce evidence and thus subject to debate. More commonly, P1 is taken as a given for the earliest stages of the Insular Celtic languages, but reconstruction of the proto-Celtic language is understood to be beyond our reach, due to lack of documentation. The reader is referred to Fife and Poppe 1991 for various stances on the issue of reconstructing constituent order in Celtic.

Although the relationships between the surviving Celtic languages are neatly
elaborated, split into Brythonic and Goidelic, the Stammbaum of the entire Celtic branch is uncertain: MacAulay (1992a) argues that the commonly mentioned dichotomy of Insular and Continental Celtic is misleading, that ancient Gaulish may have had a closer affinity to Brittonic (i.e. Welsh, Cornish, and Breton) than to Goidelic (Gaelic), due to the common innovation of *k₁u̯ > p (as well as the reflex of syllabic nasals). However, the problem with that tree is that some dialects of Gaulish (as well as Celtiberian) retain the labiovelar in its original form, which would place Brittonic within the Gaulish language, which is not a commonly accepted scenario. The sound change *k₁u̯ > p is common enough to possibly have arisen more than once (cf. Greek and Osco-Umbrian in other branches of the Indo-European family), or to have diffused across language boundaries. In short, the Stammbaum of the Celtic branch remains uncertain.

Looking back into the earliest history of Indo-European languages, however, one notes a preference for SOP in the earlier stages. The classical languages of the family, i.e. Hittite, Sanskrit, Greek, and Latin all had a tendency toward verb-final syntax (cf. Horrocks 1997 for Greek). However, constituent order was generally more flexible in older written resources (many of which are of poetic nature). Thus, some people argue for the reconstruction of SOP in Proto-Indo-European (e.g. Lehmann 1974, Hock 2013), and others claim that no dominant constituent order can be reconstructed (e.g. Friedrich 1975, Pires and Thomason 2008).

Regardless of the basic constituent order of PIE, and regardless of the exact timing of the innovations that established P1 as an unmarked constituent order, the sequence of events that led to that end are commonly agreed upon (cf. Watkins 1963): the process is initiated by a tmesis of the verbal complex, by which preverbs (marking TAM) are planted in clause-initial position, probably through a process of cliticization. This leads to a stage of preverb in clause-initial position and finite verb in clause-final position, which is still attested in archaic Old Irish (commonly referred to as Bergin’s Law). Later, in a process of univerbation, the verbal complex is reunited, but in the position of the preverb, i.e. in clause-initial position, leading to a P1 syntax.

The Celtic branch is the only commonly cited representative of P1 in the Indo-European family, but Modern Greek also makes regular used of P1. Although Greek is generally quite flexible in constituent ordering (a fact which is often attributed to its extant case marking system), the only orders that allow a broad-focus reading are SPO and PSO (Horrocks 1983, Mackridge 1985: 234–7). Arguing on the basis of theoretical models, various authors even cite PSO as the basic constituent order for Modern Greek (e.g. Philippaki-Warburton 1985). P1 may not be as consistent in Greek as in Celtic languages, but as opposed to other European languages, where P1 may occur as a pragmatically marked structure or only in intransitives (cf. Chapter 2), P1 in Greek fulfills the standards of this study according to various sources, whether or not PSO is considered the basic constituent order.

(6.100) Modern Greek [ell, Indo-European, source: Mackridge 1985: 235]
Earlier stages of Greek allowed for even greater flexibility, but there does not appear to have been any preference for P1; rather, SPO and SOP are the most commonly used orders (cf. Cervin 1990). Similar to the Celtic case described above, the origins of P1 in Greek also stem from cliticization, although not of a verbal element but pronouns: Horrocks (1990) argues that a tension between two possible binding sites in Wackernagel (second) position or immediately following the relevant head was resolved by moving the verb in clause-initial position, thus merging the two possibilities. Later, the resulting PSO order was generalized even to cases that did not involve any pronominal clitics.

Apart from Celtic languages and Modern Greek, other Indo-European languages are either SPO in the West, i.e. in Europe, or SOP in the East, i.e. outside of Europe in the Indo-Aryan branch. A notable exception to this is the Germanic languages, most of which follow a V-2 order in main clauses, which means that the conjugated verb appears in second position after the initial constituent, which could be an argument or an adjunct. However, predicate-initial and predicate-final order are also attested in Germanic: predicate-final order is employed in dependent clauses, and predicate-initial clauses appear as a narrative strategy in Icelandic, and in the genre of joke-telling in German, albeit restricted to intransitive clauses (Askedal 2011: 48–9). The V2 order now attested almost exclusively in Germanic (and the strongly German-influenced Romansh in Switzerland) was once more widely applied, as evident in the older attested stages of various Romance languages.

To summarize, P1 within Indo-European is marginal, and mostly an independent development of Celtic, possibly specifically Insular Celtic. Within the modern Celtic languages, P1 is the norm, with all modern languages agreeing. The history of P1 in Celtic can be retraced partially through written attestation. P1 is also a common, though not quite as consistent a constituent order in Modern Greek, although it is unclear how this arose.

### 6.2.8 Mixe-Zoquean

Located between Otomanguean and Mayan, two language families with a strongly dominant P1 structure, one would expect this medium-sized language family of Oaxaca and Veracruz in southern Mexico to exhibit the same clausal syntax. However, only a minority of these languages use P1. In fact, the only language for which I could confirm regular use of P1 is Copainalá Zoque, exemplified in (6.101).

\[ čʌk-yah-u \quad pʌ'nis \quad tʌc \]
\[ make.pl-compl \quad man \quad the house \]
‘The men made the house’ or ‘The man made the houses.’
Campbell et al. (1986) cite Copainalá Zoque as a typical representative of the Zoquean branch of the family, in following POS order. However, several grammars on other Zoquean languages (e.g. Suárez on Sierra Popoluca), describe SPO as the preferred order, though it is mentioned that the ordering of constituents is fairly free. Some studies provide constituent order statistics based on narrative texts. Such studies show that although P1 is present, this order is underrepresented, accounting for only 22% of clauses with two arguments in Chimalapa Zoque (Johnson 2000: 388–9) or 13% in Sierra Popoluca (de Jong Boudreault 2009: 593). It is possible that P1 order in Copainalá Zoque is the result of borrowing from neighboring Mayan languages (cf. §6.1.2). Besides Copainalá, the only Zoquean language that is characterized as having a P1 basic constituent order is Texistepec, described as POS in Reilly 2002 (p. 35). However, Reilly’s sentential data do not support his analysis, but rather show a dominance of SPO order. Based on these accounts, it appears that constituent order in a majority of Zoquean languages is fairly free, with P1 as a rather uncommon option.

The Mixean branch of the family has been described to be a typological outlier in Mesoamerica, with predicate-final constituent order (cf. Bartholomew 1983). Indeed, this analysis is backed by grammatical descriptions on Isthmus Mixe (Dieterman 2002), Tlahuitoltepec (Martínez 2008: 40–41, Gutiérrez Díaz 2014: 165ff.), and Totontepec (Suslak 2003). Campbell et al. (1986) question this characterization of the Mixean branch, claiming PSO as the basic constituent order at least in Tlahuitoltepec, based on published data from Lyon 1980.
Although several authors disagree with Campbell et al.’s (1986) analysis of Tlahuitoltepec syntax, Lyon’s (1980) data does show several examples of PSO order, and hence it is included as a P1 language in this study. It is unclear whether Tlahuitoltepec is special within Mixe in this respect, or whether all varieties of Mixe regularly use P1 order, but an answer to this question would require corpus analysis or dedicated fieldwork, which is beyond the scope of this present study.

Reconstructing the history of this language family, Wichmann (1995) suggests the proto-language to have followed an SPO order, but his argumentation is based on verbal agreement morphology, which does not offer conclusive evidence (cf. Chapter 5). Instead, it is possible that predicate-final constituent order represents an older stage of the Mixe-Zoquean family, for two reasons: 1) predicate-final constituent order is unlikely to have arisen through contact, as the flanking families Otomanguean and Mayan are both P1, and Spanish is SPO, 2) predicate-final order is especially strong in subordinate clauses, which are not affected by topicalization processes.

6.2.9 Chapacuran

Out of some dozen recorded languages of this language family situated on the Brazilian-Bolivian border, Wari’ is the only one that still has a vibrant speaker community with nonetheless fewer than 2000 speakers reported in 1997 (Everett and Kern 1997). Most Chapacuran languages are only recorded through wordlists, but the two languages that have grammatical descriptions, Wari’ and Itene (Moré), show P1 syntax, following POS order (Everett and Kern 1997: 219;328).

\[
\begin{array}{c}
\text{On nana-on jowin hwijima} \\
\text{whistle 3P.PST-3S.M monkey.species children}
\end{array}
\]
‘The children whistled to the jowin monkeys.’

\[
\begin{array}{c}
kirik na-ʔôn kinam ʔikas \\
\text{see IPV-M jaguar shaman}
\end{array}
\]
‘The shaman sees the jaguar.’

Both languages are reported to allow for some freedom in constituent order, as participant roles are cross-referenced on verbal enclitics. However, Angenot-de Lima specifies that Itene does have certain restrictions on its constituent order. First, objects are fronted to pre-predicate position in relative clauses, questions and cleft constructions,
but they do not trigger the verbal cross-reference markers in this position, which suggests that they fall outside of the core clause. Second, the fronting of the subject to pre-predicate position is practically unattested in Itene, while PSO and SOP are completely ungrammatical (Angenot-de Lima 2002: 328). This is interesting since subject fronting is otherwise a common phenomenon in P1 languages.

In the absence of any counterexamples to a general POS order in Chapacuran, this order is the best candidate for the reconstructed Proto-Chapacuran syntax, although this hypothesis would of course need to be corroborated with data from other Chapacuran languages. However, since Wari’ and Itene both belong to the same branch of Chapacuran (though they are not the most closely related sister languages), and they are spoken in relative proximity, a lateral syntactic spread cannot be discarded as a possibility.
Figure 6.24: P1 in the Chapacuran language family
6.2.10 Cariban

This major language family distributed mostly across the Guyanas, but with branches as far as Colombia and Brazil, has a fairly rigid OV sequence (Derbyshire 1999: 55). The predicate is mostly placed in clause-final position, across the family. Although Cariban languages are fairly closely related lexically, they exhibit astonishing morphosyntactic variation, in alignment, locus of marking (head/dependent), and constituent ordering (Gildea 1998: 29). The majority of Cariban agree in a basic SOP, which is accordingly posited as the basic constituent order of proto-Cariban (Derbyshire 1981, Pachêco 2001: 246). However, there are a few exceptions to this. First, Yukpa is described as having a basic SPO constituent order (Adelaar 2004: 114), which shall not concern us any further for the purposes of this study. Perhaps the most renowned member of this family in terms of constituent order is Hixkaryana, which follows the typologically very rare OPS basic order (Derbyshire 1985). Although Hixkaryana has received much attention for this feature in the linguistic literature, other Cariban languages also make use of this order – whether that is described as their basic constituent order, as for Tiriwiyó (Carlin 2004: 477), Arára (de Souza 1993) or Apalai (Koehn and Sally 1986: 33–4), or as an alternative to the standard SOP order, as for Kuikuro-Kalapalo (dos Santos 2007: 33) or Waimiri-Atroari (Bruno 2003: 133ff.). These languages continue the old OV sequence but moved the subject after the predicate (only in main clauses in the case of Hixkaryana), resulting in an OPS constituent order. The Cariban family also has a couple of members that are described to have a basic P1 order, namely Ikpeng and Panare. Ikpeng, spoken in the Xingu province
of Brazil, follows a basic PSO order, but either core argument may precede the predicate if it is animate and when in focus (Pachêco 2001: 160).

\[
\begin{array}{ll}
t-wo-li & Korotowi tae \\
3-kill-rec K. & monkey \\
\end{array}
\]
'Korotowi killed the monkey.'

Interestingly, the predicate-final order, which is the standard for the whole language family, is completely disallowed in Ikpeng (Pachêco 2001: 158). Panare, spoken in Venezuela south of the upper reaches of the Orinoco River, is described as "predominantly verb-initial" by Payne and Payne (2013: 19–20). Payne and Payne argue that the Panare verb forms a constituent with the following transitive subject, rather than with the object, although the latter is considered the cross-linguistic norm by most theoretical syntacticians. The object may occur on either side of the PS unit: both PSO and OPS are described as unmarked orders for Panare, but PSO is more frequent in discourse, although OPS is the standard order for elicited clauses prompted in Spanish. Payne and Payne recognize OPS as the older order of the two.

It is understood that the P1 order of Ikpeng and Panare is derived directly from an OPS order, rather than from the older predicate-final order, by disintegration of the OV phrase, making the object mobile and moving to clause-final position when not in focus (Derbyshire and Pullum 1981: 204, Pachêco 2001: 248). This hypothesis is supported by the fact that Ikpeng is closely related to (the geographically quite distant) Arára, which has a basic OPS order but does not allow P1 order (Pachêco 2001: 249).

In conclusion, there is strong evidence for an old rigid OV sequence in Cariban languages. The positioning of the subject with respect to this verb phrase is flexible in so many languages across the family that it is likely that it was not fixed in the proto-language either. P1 appears to be a fairly recent phenomenon in Cariban, still alternating with an earlier OPS order in Panare, and more firmly grammaticalized in Ikpeng.

6.2.11 Tupian

This language family stretches across a large part of the South American continent, from Peru to Brazil, and from French Guiana to Uruguay. Generally speaking, Tupian languages are predicate-final, though there is some variation in this large family (Storto 1994, Jensen 1998: 555–7). The polysynthetic nature of these languages allows for freedom in constituent order. However, there is often less flexibility in dependent clauses, where the predicate-final order becomes more dominant (cf. Storto 2014). Furthermore, a preference for SOP order was apparent in Old Tupi (Tupinambá) (Navarro 2004: 62). By contrast, the most prominent member of this family, Guarani, which has official status and is widely spoken in Paraguay, follows SPO order, quite likely as a result of standardization under the influence of Portuguese (Estigarribia 2020). P1 can occur as a marked construction, which for Karitiâna is described as a verb-focus construction that requires special morphology
Language families with P1 in a minority of languages

Figure 6.26: P1 in the Tupían language family

on the verb (e.g., Everett 2006: 290ff.;321). In intransitive clauses, the predicate is often initial, as in Karitiâna.

One language in the Tupi-Guaraní stock has been described as having a P1 basic constituent order, namely Tenetehâra, also known as Guajajara or Tembê.


\[
\begin{align*}
\text{u-munyk} & \quad \text{i-azyr} & \quad \text{i-petym} & \quad \text{heraha} & \quad \text{i-zupe aʔe} \\
3-\text{lit} & \quad 3-\text{daughter} & \quad 3-\text{tobacco} & \quad 3s-to & \quad 3
\end{align*}
\]

‘His daughter lit his cigar (and took it) to him.’

This language, like other members of the stock, is flexible in its constituent order, but in a corpus analysis of 200 pages of mostly narrative text, Harrison (1986: 408–9) found that 19 of 28 sentences with two NP arguments followed the order VSO, which is equivalent to 77% (the other recorded orders being VOS, SVO, and SOV). Harrison specifies, “when assignment of subject and object is potentially ambiguous, the subject strictly precedes the object, though the verb may occur initially, medially, or finally” (Harrison 1986: 408–9). This flexibility in main clauses contrasts with dependent clauses, where Guajajara follows the typical verb-final order.

It is unclear how P1 developed in Guajarara. The only other indication of P1 in the family that I have encountered in the literature is the verb-focus construction in Karitiâna.
However, the two languages are geographically wide apart, and genetically not closely related. Further investigation into the contemporary distribution of P1 in Tupían and into its history would be worthwhile.

6.3 Language isolates and dialect networks with P1

6.3.1 Alsea

Little data is available on this language isolate on the central Oregon coast, because Alsea ceased to be spoken as a native language with the death of John Albert in 1942. Fortunately, Leo Joachim Frachtenberg documented the language in 1910, producing both a rather comprehensive grammar and a text collection. In his assessment, the object must follow the verb, while the subject is positioned more flexibly, though preferably at the very end of the sentence (Frachtenberg 1918: 421).

(6.107) Alsea [aes, isolate, source: Frachtenberg 1918: 421]

\[
\begin{align*}
tum-un'\text{hū} & \quad L-pi'k'lay-utx & \quad le'wi'-k's & \quad ats-ti'nExtsū-k' & \quad x-as \\
\text{then-now} & \quad 3.\text{OBJ-place-quickly} & \quad \text{ground-DIR} & \quad 3.\text{POSS-cane-3S.POSS} & \quad \text{AGT-the} \\
\text{mEsha'lsla-} & \quad \text{tsLō} & \quad \text{wife-AUG} & \quad & \\
\end{align*}
\]

‘then that old woman placed quickly her cane to the ground.’

The resulting POS order is characteristic of other languages of Central Oregon, including Siuslaw and the Chinookan languages. More generally, their preference for P1 places them within the Pacific Northwest Sprachbund. The likelihood of Alsea having adopted P1 at some point in its history is very high.

6.3.2 Guató

This language isolate is spoken in the state of Mato Grosso do Sul along the border of Bolivia.\(^\text{17}\) It borders the predicate-final Bororoan languages to the north and Kadiwéu (SPO) to the south. Its P1 syntax, exemplified in (6.108), is somewhat out of place for its areal setting.


\[
\begin{align*}
\text{ma-e-ro} & \quad g-\text{ępagu} & \quad g-\text{ęki} \\
\text{IPV-3-eat} & \quad \text{DET-jaguar} & \quad \text{DET-rabbit} \\
\end{align*}
\]

‘The jaguar ate the rabbit.’

The closest P1 languages in the area are the Enlhet-Enenlhet languages of the Argentinian Chaco (§6.1.6), with which they are connected by the Paraguay River (but

\(^{17}\)Guató was tentatively grouped within Macro-Jê by some authors (Kaufman 1990, Martins 2011), but this does not represent the consensus in the field (Campbell, p.c.).
beyond the Kadiwéu mentioned above). It is possible that the two groups had regular trading relations leading to a diffusion of P1, but I could not find any historical or ethnographic sources to confirm this.

6.3.3 Karirí

This language isolate was spoken in Central Eastern Brazil, in the province of Bahia, until the 1970s. The language is also known by the names of various dialects, including Kipeá, Kamurú, Dzubukuá, Sabujá, some of which are considered separate languages by some (not by Glottolog). Data on these are difficult to find, but since they are closely related, it is safe to assume that they all exhibited the POS order exemplified for Kipeá in (6.109).


\[
\text{si-pa kradzo no karai} \\
3\text{-kill cow ERG white.man}
\]

'The white man kills the cow.'

Karirí is neither anywhere near, nor demonstrably related to, other P1 languages, and little is known about its history. Typologically unusual for a P1 language is that in Karirí the adjective follows its head noun (Rodrigues 1999a: 193).

6.3.4 Kuot

This isolate of New Ireland in the Bismarck archipelago is the only language that has been described as P1 in the entire New Guinea region (cf. §4.2.1).

(6.110) Kuot [kto, isolate, source: Chung and Chung 1996: 5]

\[
a\text{-kar-oŋ kaβuna lamuana} \\
3\text{s.m.OBJ-bite-3s.m.sbj dog L.}
\]

'The dog bit Lamuana.'

Also, it is the only non-Austronesian language on New Ireland, surrounded by a sea of Austronesian (Ross 1994). As such, it has been influenced heavily by Austronesian in its structure (Ross 1994, Foley 2000: 376). Notably, it features an inclusive/exclusive distinction, as well as a decimal system, both features that are highly uncommon for the non-Austronesian languages of the area, but typical of Austronesian (Reesink and Dunn 2018: 972). Even though the Austronesian languages of the area are predicate-medial, rather than predicate-initial, the reconstructed constituent order of Proto-Austronesian, as well as Proto-Malayo-Polynesian is without doubt P1, and probably that of Proto-Oceanic was too (cf. §6.2.1). Thus, it is likely that even though the modern Austronesian languages of the Bismarck archipelago are not P1, they may have been in the past. If that was the case, it is quite probable that Kuot adopted P1 along with the typical Austronesian

---

18Karirí used to be included in the Macro-Jê group, but this affiliation is now rejected by most experts.
features stated above. Reesink and Dunn (2018: 950) state this as follows: “We leave the possibility open that [V-initial order] is a retention of an early borrowing of AN V-initial order, even after the surrounding AN languages had shifted to V-medial” (cf. Lindström et al. 2007: 132). Though not itself Austronesian, Kuot may thus very well instruct us about the syntactic history of the Oceanic languages.

6.3.5 Salinan

Salinan is a language isolate of California, which neighbours the P1 Chumash family to the south. Though sparsely attested and extinct since 1958, Salinan clausal syntax is described by Turner (1987: 204) as free, but with an unmarked POS order. Strikingly, the initial positioning of the predicate is maintained even in “emphasis and topicalization,” for which the order PSO is used. However, Turner’s assessment is not based on personal fieldwork but exclusively on previous documentation (Mason’s texts and fieldnotes by Harrington), as the last native speaker of Salinan died in 1958. In a statistical analysis of Mason’s narrative texts, Dryer (1989) reaches a slightly different conclusion: on 52 pages of text, he identified only nine sentences with both subject and object expressed as independent NPs, and six of these follow the order SPO, while only one follows the POS order understood as basic by Turner. However, in the much more common construction of sentences with one single NP argument, both subjects and objects had a strong tendency to follow the predicate (cf. (6.111)).

(6.111) Salinan [sln, isolate, source: Mason 1918]
   a. a´m-pamat’-ko pe-t-i‘itc-o
      then-chased-3s.OBJ ART-N-dog-3s.Poss
      ‘Then his dog chased him.’ (p. 95)
   b. ra´m-hala’ umck’au’yu
      then-used talons
      ‘Then he used his talons.’ (p. 62)

Dryer concludes that Salinan constituent order is most accurately described as “SVO, VS and VO”, but one could also summarize the data as showing P1 with a (loose) restriction on two NP arguments after the word, which is avoided by fronting the subject. This description is consistent with the strong tendency for subjects to precede the predicate in intransitive sentences. For the purposes of this study, we can state that Salinan makes regular unmarked use of P1, regardless of the basic constituent order(s) of the language.

19 The object of that sentence is indefinite and does thus not fulfill the criteria set for the present investigation (cf. (2.2) in Chapter 2).
6.3.6 Yokuts

Native to the San Joaquin Valley of California, Yokuts forms part of a small P1 cluster in Central California, together with Chumash (§6.1.10) and Salinan (§6.3.5). Several accounts on different varieties (Kroeber 1907: 233, Gamble 1978: 121, Rhodes 2013) analyze PSO as the basic or most common constituent order for Yokuts. Others, however, claim that the position of the subject is free (Collord 1968: 80) or usually initial (Britsch 1980: 20). The disagreement between these authors may be due to differing approaches data collection and analysis. It is also conceivable that the disagreement is reflective of dialectal variation: although Ethnologue recognizes only a single Yokuts language, Whistler and Golla (1986) have disputed the general intelligibility between all varieties, and Glottolog differentiates four Yokutsan varieties. However, the available sources do not allow for a contrastive study of different Yokuts varieties (Newman’s otherwise very insightful comparative grammar [Newman 1944] does not address constituent order), but it appears that the differences between different varieties are mostly of lexical nature (Golla 2011: 148–9), and hence Yokuts is treated as one in this study. P1 in Yokuts is illustrated in (6.112).

(6.112) Yokuts [yok, isolate, source: Rhodes 2013]

a. Ta’ash-al Nancy-’ nan
   see-PST  Nancy-SBJ 1S.OBJ
   ‘Nancy saw me.’ (p. 17)

b. Ta’sh-it na’ nim nek’et
   see-PST 1S.SBJ 1S.Poss aunt
   ‘I saw my aunt.’ (p. 19)

As an isolate or closely knit dialect network, the syntactic history of Yokuts is difficult to investigate. However, it is quite possible that Yokuts adopted P1 from its coastal neighbours, the Chumash and Salinan, are both P1. Although the Yokuts are separated from these by the high Sierra Madre Mountains, all these tribes were in regular exchange, not only for trade of marine edibles, of which the Chumash had an abundance, but also shell beads used as currency, and the Yokuts would regularly attend public ceremonies during the solstices in the Chumash territory (Gamble 2008). Thus, the contact took place in the coastal areas, with the Yokuts visiting the Chumash in their territory, rather than the other way around. With their abundance of natural resources, which left a deep impression the first Spanish explorers in the area, and their large towns, the Chumash also formed a cultural centre, which influenced the nations around them (Gamble 2008). This cultural exchange could have well led to the adoption of P1 in Yokuts.

Callaghan (1997, 2001) proposed a relationship between Yokuts and the Utian family, i.e. with Miwok and Costanoan (including Ohlone). However, this proposal was never generally accepted, because of many irregularities in the suggested related etyma. These are rather indicative of extensive borrowing rather than genetic relationship (Golla 2011: 130).
Chapter 7: Tracing predicate initiality through time

The reconstruction of syntax, and especially clausal syntax, is often speculative, and this is partly due to the lack of a commonly recognized methodological standard in the field (cf. Chapter 5). It is also because borrowing, which is a notorious source of noise in the reconstructions of any linguistic module, is particularly common in clausal syntax. Nonetheless, a general consensus about the syntactic past of many language families has been reached. For many language families, P1 is unanimously reconstructed to their proto-language; for some it can be shown to have entered through contact with other language, and to have developed language-internally. The processes and patterns that were described in Chapter 6 are here consolidated and reviewed as a whole.

The most reliable type of reconstruction is through actual documentation: direct evidence of historical constituent order from a long literary tradition leaves little room for speculation about syntactic history, and it can push the boundaries of reconstructible history back thousands of years. Fortunately, such evidence is available for Semitic languages, as well as for some other Afroasiatic branches, allowing us to reconstruct P1 to Proto-Semitic with certainty, and even to Proto-Afroasiatic with reasonable certainty, although this family has an exceptionally deep history, and although only a minority of modern Afroasiatic languages follow P1 (see §6.2.2). However, few other language families are documented thousands of years back in time. Furthermore, there are sound arguments to assume that historically documented languages, as the largest and most standardized representatives of a language family, are not necessarily the most conservative varieties: standard languages have higher numbers of L2 speakers and may be subject to standardization after another world language (cf. Chapter 5).

We can also be quite certain in reconstructing P1 to the proto-level of large language families where P1 is spread across multiple primary branches whose languages are dispersed across a large territory with little contact between the sister languages. The Otomanguean and the Salish language families, for example, are spread across large territories with non-contiguous branches. Both families are internally quite diversified but have a large majority of languages agreeing on P1 constituent order. However, both families also form part of larger P1 areas, such that the separated branches may be isolated genetically but not areally. Still, given the sheer number and geographic spread of these families, there is a consensus about the reconstruction of P1 to their proto-languages.
The same is true for other P1-dominant families like Mayan or Chumashan, although the argument becomes weaker the more geographically contiguous, historically connected, and smaller, the language family is. Given the extent of the Mayan civilization, or the compact grouping of the Chumashan family, a lateral diffusion of constituent order at some point in history is not unlikely. However, without any specific indications of such an event, the most parsimonious account is to assume reconstruction of P1 to the proto-language.

Reconstruction for smaller language families is most certain when they are areally unusual, but less so when nested in linguistic areas. For example, while the Xinkan languages of Guatemala all agree on P1, this constituent order may be reconstructible to Proto-Xinkan, but given the proximity of this family to the P1 Mayan languages, they could just as well have borrowed Mayan clausal syntax. To compare, the Enlhet-Enenlhet family is around the same size as Xinkan, and likewise consistent in P1, but unlike for Xinkan, there is no contemporarily or historically dominant P1 language in the vicinity of the Enlhet-Enenlhet languages. Hence, reconstruction of P1 is our "best bet" for Enlhet-Enenlhet, but not necessarily for Xinkan.

While the above discussed language families feature P1 in a majority of their daughter languages, such a numerical dominance does not necessarily warrant the reconstruction of P1 to the proto-language. As mentioned above, P1 can be fairly safely reconstructed to Proto-Afroasiatic despite the relative sparsity of P1 among its modern daughter languages, thanks to historical documentation. P1 can also certainly be reconstructed to Proto-Austronesian, although P1 is used as a standard constituent order in only around 27% of its daughter languages. Crucially, though, these 27% cover all but one of ten or eleven primary branches of the family, and even the last, internally most diversified, branch Malayo-Polynesian can safely be reconstructed to P1 just with internal evidence (cf. §6.2.1). P1 can also be reconstructed to Proto-Nilotic, although P1 is only spread across some 45% of this family. This is because P1 dominates two of its three branches, and it is furthermore found in subordinate clauses in the third branch (cf. §6.2.3). These examples demonstrate that a well-developed understanding of the family tree can be extremely helpful, if not necessary, in reconstruction.

Needless to say, the sparser the distribution of a feature, the bolder the claim for its reconstruction. In a few cases it has been argued that although there is general agreement in constituent order in a family, the general typological profile of a language or language family as a whole (e.g. orientation of adpositions or affixes) points toward a distinct

---

1 However, Lyle Campbell (p.c.) specifies that the Maya civilization, which emerged from speakers of languages of the Greater Cholan and Yucatecan branches (in the lowlands), only influenced neighboring Mayan languages, with very little influence in the highlands. This influence sphere results in structural differences between these regions, for example apparent in absolutive pronominal markers, which are prefixed in the highlands but suffixed in the lowlands.

2 But even for the reconstruction of Xinkan, parsimony rules, i.e., assuming P1 for Proto-Xinkan (which could also on its part borrowed this syntax from Mayan) with subsequent inheritance into the daughter languages is an easier model than assuming separate borrowing of P1 for each of the daughter languages (either all from Mayan or including lateral spread among Xinkan).
syntactic past. On such a basis, for instance, Fortescue (2009) argues that (Pre-)Wakashan adopted P1 from Salish (see §6.1.8), and Jenny et al. (forthcoming) suggest that Proto-Austroasiatic was P1, despite the fact that constituent order is reflected only in a small fraction of the modern Austroasiatic languages (see §6.2.6). Such evidence alone is rather weak, however, unless supported by further data.

Of course, the history of P1 is not just one of “erosion” from the earliest proto-languages. While syntactic borrowing has been described as a viable alternative to reconstruction in many cases above, there are also cases where P1 demonstrably spread to new language families through contact. The most conclusive cases are language families that straddle linguistic areas, with some languages inside a well-defined linguistic area, and others outside it, with little contact between the two groups. Such is the case for Nahua, which evidently represents a newcomer in the P1-dominant Mesoamerican Linguistic Area, as their origins are farther north in the Southwestern United States of America: most other Uto-Aztecan outside of this area follow a more-or-less strict predicate-final order (see §6.2.4). In this way, P1 entered some languages of the Uto-Aztecan family through language contact. In many other cases, such as in Xinkan, mentioned above, syntactic diffusion may not be as obvious as in Nahua but still a very likely scenario, namely whenever a small language family is outnumbered by a dominant family of the same syntactic type. Thus, diffusion of P1 from Nilotic has been suggested as the origin of P1 in Surmic (§6.1.4), and it is also quite possible for Kuliak, which is in fact being replaced by Nilotic (§6.1.13).

An interesting case is Kuot on New Britain: although an isolate, it can be classified as a Papuan (i.e. non-Austronesian) language, which generally follow predicate-final syntax. The general syntactic history of Melanesia is that many Austronesian languages adopted the very un-Austronesian predicate-final constituent order from Papuan languages (Foley 2000). Yet, the P1 in Kuot is not only unusual for Papuan languages, but also for the Austronesian languages of the area, which follow SPO, if they have not shifted to Papuan SOP. Crucially, Kuot is surrounded by Austronesian and shows obvious structural signs of intense borrowing from Austronesian. Since there are good reasons to assume that the Austronesian languages in the area were historically P1, the most likely scenario is that Kuot adopted P1 from Austronesian, which subsequently gave that order up again in that region. See §6.3.4 for a more detailed discussion of Kuot.

Finally, the origin of P1 in many language families is simply unknown. Especially intriguing and worthy of dedicated inquiry is the presence of P1 in areally dominant large language families with generally different syntactic profiles, like Arawakan, Austroasiatic. It should also be mentioned here that because clausal syntax is prone to change due to language contact, and language contact is triggered by intense contact between peoples, the value of non-linguistic historical information in the reconstruction of constituent order should not be under-estimated. As I am not a historian, and this is not a historical study, such information was only consulted when readily available. For example, knowledge about the Chumash control of trade and ceremonies in Central California allows us to recognize Chumashan as the probable center of diffusion of the small P1 cluster in the area.
Perhaps historical (including archeological) investigation into historical trading relations among the tribes of the Gran Chaco (possibly along the Paraguay River) or the East Africa Rift (possibly through the White Nile and tributaries) could shed more light on the history of diffusion of P1 in those areas.

Overall, we can state that although there are examples of languages that have demonstrably adopted P1 in observable history, the great majority of P1 languages most likely inherited this constituent order from their earliest reconstructible proto-language. Cumulating just the P1 languages from the six language families Austronesian, Otomanguean, Afroasiatic, Mayan, Salish, and Nilotic, for which the reconstruction of P1 is communis opinio, accounts for 652 of the 811 P1 languages in this study, or 80%. Many of the remaining P1 languages have either demonstrably borrowed P1 from neighboring non-related languages (such as Nahua or Kuot), or their families are nested in linguistic areas so that their history is difficult to assess (e.g. Chinookan or Kuliak).

There are only a few cases where P1 demonstrably developed language-internally in observable linguistic history. The best-known and perhaps only such case is the Celtic languages (§6.2.7). This begs the question why P1 does not develop independently more commonly. As suggested in Chapter 5, the synchronic and diachronic syntactic processes that would derive P1 are generally rare: languages commonly highlight constituents as a topic or focus by placing them in clause-initial position. Although this process is possible for verbal constituents in some languages (cf. Chapter 2), it is much more commonly applied to participants, and especially to core arguments, because (1) participants are more salient than actions and thus more prone to pragmatic profiling (different types of focalization), and (2) participants (especially subjects) are more continuous in narratives than actions and thus more topical. This not only means that predicates are unlikely to move to the front of the clause, but also that if the predicate occupies the initial position, it is easily demoted from that position. This is evident in the fact that alternative constituent orders with an initial core argument, especially an initial subject (i.e. SPO), are possible even in the strictest P1 languages, and often competing in frequency with P1 in the languages for which the latter is described as the basic constituent order. P1 languages differ in how often they front arguments, and this difference is perhaps less a question of grammar than one of style (although the distinction between these categories is of course all but clear). Over time, increasing appliance of argument fronting may lead to a shift in relative frequency, and to a concomitant semantic bleaching, meaning that to associated topic/focus reading of the initial argument is lost. The most common instantiation of this process is a diachronic shift from PSO to SPO order, as observed in the majority of Semitic and Austronesian languages.

Such a process is often accelerated by contact with prestigious world languages, most of which follow SPO: English (in North America), Spanish (in Middle and South America), Portuguese (in Brazil), French (in New Caledonia and the Maghreb), Swahili (in East Africa), Chinese (in Taiwan), and historically Dutch (in Indonesia) cover most of the world’s P1 areas as literary languages and have had substantial influence on the P1 languages in these areas, contributing to the shift to the basic constituent order SPO they
all share. It is important to notice, though, that such a contact-induced or contact-assisted shift in constituent order from P1 to SPO did not merely supplant the original syntax with a completely novel pattern, but it promoted an already existing construction to the norm.

That being said, there are cases where P1 appears to have solidified over time, as a response to syntactically conflicting encroaching languages. This argument has been made for Breton, for which Varin (1979) argues that the now commonly generalized PSO order represents an artificial ("chemical") construct by intellectuals, whereas SPO was much more commonly used by earlier Breton writers (cf. §6.2.7). Although this would need to be verified in a dedicated study, I have the impression that a similar process is happening with the Hawaiian language: because the order SPO is often labeled "English-like" in classrooms, L2 speakers, which constitute the majority of Hawaiian language users, learn to avoid this construction in favor of the typically Polynesian PSO. It seems that native speakers of the Ni‘ihau community, who acquired the language in a home environment, rather than at school, make more frequent use of SPO than L2 speakers.

The overall global trend appears to be that of a waning of predicate initiality. P1 was once prominent in many supraregionally influential languages and may have spread through these: Biblical Hebrew, Classical Arabic, Ancient Egyptian, Classical Mayan and Nahuatl were all primarily P1, and Malay also made common use of P1, besides SPO. Nowadays, though, the only larger P1 national languages are Arabic (only partly P1), Tamazight and Tagalog, besides some smaller official languages on Pacific Island states such as Sāmoan, Tongan, Palauan, or Chamorro.

Given the historical trend away from P1, I dare undertake a prospect into the future of this constituent order. As this study has amply demonstrated, P1 is mostly an inherited pattern, which only very rarely develops independently. However, P1 can, as any other constituent order, spread by language contact. Considering the current geopolitical distribution of P1 among the world’s language today, though, there seems to be little outlook for a big-scale spread of P1. The only supraregionally influential P1 languages are Arabic and Tagalog, and their sphere of influence is restricted to regions that already feature P1. Tagalog is spreading in the Philippines, to the detriment of other languages, but the endangered languages in the country are already P1. While the rapid population growth of the Philippines does lead to an increase of P1 language speakers in the Philippines, perhaps even a percentage increase of P1 language speakers in the world, the number of P1 languages in the Philippines is decreasing, as the linguistic diversity in the country diminishes. As for Arabic, its sphere of influence lies in the Levant, the Arab peninsula and North Africa. Arabic is already firmly established on the Arab peninsula, which means that P1 has no potential of spread there. In the Levant, there are some bilingual areas where P1 could potentially spread from Arabic: while Eastern Arabic varieties had a historical tendency to replace the traditional P1 syntax with predicate-final syntax under the influence of Turkic and Iranian languages (cf. §6.2.2), it appears that the current power balance favors Arabic as the center of diffusion especially in Syrian and Iraqi Kurdistan, where Kurdish may adapt to the local literary and supraregional language of
Arabic. In Israel, a possible diffusion of Arabic syntax into Hebrew is not unlikely, given its large, and growing, percentage of Arabic native speakers, and its geo-political location amid Arabic-speaking countries. In North Africa, though, the status of Arabic is similar to that of Tagalog in the Philippines: Arabic is spreading in the Maghreb to the detriment of other P1 languages, namely the various Berber languages (cf. §6.2.2). As in the Philippines, the spread of Arabic in the Maghreb thus leads to an actual decrease in P1 languages in the area. Whether the number of P1 language speakers is rising in the Maghreb is unclear, though, because Arabic is not as consistently P1 as Tagalog.

Hence, the potential of a P1 spread from Arabic or Tagalog to new languages appears to be rather limited. By contrast, a large portion of P1 languages, including every single one of the entire region of the Pacific Northwest, and generally most in the Americas, are endangered, many severely so. This does not mean that P1 will never again emerge from contact, but the trend is unequivocal: P1 is rare in the world’s languages and will likely become much rarer still in a hundred years.
Chapter 8: Conclusion

In this study, I have approached predicate initiality (P1) from different perspectives: from a functional-theoretical perspective in Chapter 2, from an areal-typological perspective in §4.2, and finally from a historical perspective in Chapters 6 and 7. To conclude this study, I shall summarize the main findings in this final chapter, along with some limitations and suggestions for further research.

First, I shall restate that I have treated predicate initiality as a construction type, rather than a language type (where “P1 languages” are languages that make regular, but not exclusive, use of P1). Although previous typological surveys of constituent order have been mostly concerned with the classification of languages by basic constituent order, I have also included P1 in languages that have been described by others as belonging to another type. This is reasonable because not too seldom there is disagreement in the linguistic literature about the basic constituent order of a language. P1 languages in particular make frequent use of the order subject-predicate-object (SPO), and in some cases this derived order is even statistically prevalent in texts, giving rise to discussion about the definition of basic constituent order. This has the effect that descriptions on closely related and syntactically similar languages make the languages appear more different than they are in reality. Treating P1 as a construction type, rather than as a language type, allows us to gloss over such differences in theoretical background and methodology, to achieve a more complete overview of constituent order in the languages of the world. This approach lends itself to other types of typological studies. For the purposes of this study, a language is included as a “P1 language” if it allows the pragmatically unmarked usage of predicate-initial constituent order (either PSO or POS) in declarative transitive sentences (cf. (2.2) in Chapter 2).

Under these criteria, this study has identified a total of 811 modern P1 languages. Roughly half of these (396) are confirmed with sentential data (under the criteria defined in (2.2) in Chapter 2), and the other half (415) has been described as following P1 and are closely related to confirmed P1 languages, but no data are available to confirm this. All of these are likewise considered P1 in this study. With these numbers, the present study offers the most extensive atlas of P1 to date by far. (The oft-quoted World Atlas of Language Structures identified only 134 P1 languages.) In fact, this work represents a rather unusual approach to typology: in contrast to the common approach in typology of investigating the prevalence of a structure on the basis of a representative sample, I attempt to approximate an exhaustive count of a feature (to the extent that this is possible without checking every
single language in the world). While this approach offers a high-resolution insight into various language families (to an unprecedented degree in many cases), and evades the problem of representative sampling, it faces the complication that the size of the entire set (i.e. the number of the world’s languages) is not well defined. For consistency, I have used Glottolog’s framework of classification for the present study, due to its free access and likewise exhaustive scale. Although my inclusion of languages that lost their last native speakers in the 20th century renders my comparison with Glottolog’s count somewhat hazy, this study confirms the common perception that predicate-initial order is significantly less frequent compared to predicate-medial and predicate-final constituent orders, with only 811 languages, or around 11% of the world’s languages allowing this order in pragmatically neutral declarative transitive clauses.

The resulting areal picture is that the distribution of P1 around the world is highly uneven. 714 (88%) of the 811 P1 languages are found in one of five major P1 hotspots: on Pacific Islands (including Taiwan and the Philippines), in the Pacific Northwest coast of America, Mesoamerica, the Maghreb, and the East African Rift. Two countries alone are home to almost half (47%) of the world’s P1 languages: Mexico and the Philippines.1 Another 25 (3%) of P1 languages are Austronesian languages in the Indian Ocean, or languages that have certainly adopted P1 from Austronesian, the language family that accounts for P1 on the Pacific Islands. (To these 25 could be added the five Aslian P1 languages, which belong to the Austroasiatic family.) Another 26 (3%) occur in the smaller P1 pockets of Central California (Chumashan, Yokuts, Salinan), the Gran Chaco (Enlhet-Ennlenhet), and the British Isles (Celtic). This leaves only 48 P1 languages (6% of the total) distributed over the rest of the world, of which several are scattered across the Amazon.

In the above-cited P1 hotspots, P1 order is the norm, represented by the majority of languages in a larger geographic area, either within a single language family (Austronesian in Pacific Islands, and Afroasiatic in the Maghreb), or across various neighboring language families in a linguistic area. Certainly the densest concentration of P1 is found in the Philippines, where I have not found a single exception to P1 among the 170 languages in the country. As stated, only a few dozen dispersed P1 languages are found outside of the above areas. P1 is virtually absent from all of Eurasia (with a few exceptions) and Australia, and it is notably absent from two of the major hotspots of linguistic diversity: West Africa and New Guinea. Quite notably, there is no record of any P1 sign language, and P1 is practically unattested in creole languages, as well (with only a couple of exceptions in the Philippines).

From a historical point of view, we can conclude that a big majority of P1 attestations in modern languages represents an inherited pattern that can be traced back to the earliest reconstructible stages of the respective language families, i.e. to the horizon of historical linguistics. Specifically, cumulating just the P1 languages from the six language families Austronesian, Otomanguean, Afroasiatic, Mayan, Salish, and Nilotic, for which the reconstruction of P1 is communis opinio, accounts for 649 of the 811 P1 languages

---

1 The fact that these two countries share a common history through the Manila galleon trade under Spanish occupation is nothing but coincidental, as P1 in these areas clearly predates Spanish colonization.
in this study, or 80%. The Austronesian family alone accounts for almost half (346, or 43%) of the world’s P1 languages. Many of the remaining 20% of P1 languages have also been reconstructed to P1 proto-languages, although syntactic reconstruction is generally more reliable for large wide-spread language families than for smaller areally contiguous families. Yet, the history of P1 is not just one of syntactic erosion: P1, like other constituent orders, has also clearly spread through diffusion, which gave rise to the P1 hotspots. The direction of diffusion in those linguistic areas can rarely be reconstructed, but the general assumption is that the larger and historically dominant language families exerted syntactic influence on the smaller families. Thus, it has been suggested that the larger Salish language family introduced P1 to the smaller Wakashan family, or that P1 in the seven P1 Surmic languages was introduced from the more numerous Nilotic P1 languages in the area. Only in a few cases can the historical introduction of P1 be demonstrated to have occurred through contact, as e.g. for Nahua (§6.2.4) or Kuot (§6.3.4). The former represents migration into the established P1 of Mesoamerica, while the latter was all but swallowed in the vast expansion of the P1 Austronesian languages.

Independent emergence of P1 is theoretically possible, but practically highly unlikely. Perhaps the only case where the historical development of P1 can be reconstructed in detail—thanks to a long literary tradition—are the Celtic languages and Modern Greek, which both belong to the Indo-European family. (P1 in Greek, which is less consistent than in Celtic, has received less attention in the theoretical literature, but cf. §6.2.7) Put simply, the reason for the rare independent emergence of P1 is that the cross-linguistically common syntactic processes that alter constituent order lead away from P1 order: fronting operations through topicalization or focalization typically target participants, rather than actions, because the former are more continuous in narratives and more salient. This is evident in that languages with a basic P1 order typically allow an SPO alternative, while languages with a basic SPO order do not typically allow a P1 alternative. Hence, there are only few possible syntactic pathways for P1 to develop language-internally across time. In other words, P1 not a diachronic target, or historical attractor state (Nichols 1992); rather, P1 tends toward SPO over time (Hopper 1987). As a result, P1 almost exclusively spreads through diachronic diversification of languages, and through language contact. This is the reason why P1 forms such distinct areal clusters, with relatively few typological outliers, compared to predicate-initial or predicate-final constituent orders.

As P1 languages diversify over time, the inherited P1 syntax often gives way to SPO, with that originally marked order gradually losing its pragmatic force. In many cases, this development has been accelerated by world languages, most of which (English, French, Spanish, Portuguese, Dutch, Chinese, to name just a few) notably follow an SPO structure. This process is most evident in Indonesia and northern Africa, where P1 was once dominant but has persisted only in a minority of modern descendants. On a larger scheme, this tendency of a diachronic drift away from P1 is ultimately the reason for the relative rarity of P1.

Predicate initiality is a commonly recognized language type, but unfortunately it does not feature prominently in the typology of syntax—or at least not anymore. The reason for
this is that the placing of the subject is often regarded as irrelevant to the deep structure of a language, where the main dichotomy is understood to be that of head-dependent vs. dependent-head ordering, which on the clause-level is instantiated as verb-object vs. object-verb. From this perspective, P1 languages simply fall into the verb-object category, along with SPO languages. I hope that this study gives a fresh impetus into the study of P1 as a distinct category, because there are many open questions regarding features of languages that set them apart from SPO languages. Just a few of these questions are listed below:

1. If the verb and the object universally share a closer syntactic connection than the verb and the subject, then there should be a constraint against the order PSO. Yet, PSO is more common than POS. Why?

2. Ergative alignment appears to be a common trait of many P1 languages (e.g. in Austronesian, Salish, Tsimshian, Mayan, to name just a few), presumably more so than of languages following SPO consistently. Does this apparent proclivity of P1 for ergative alignment hold up against a large sample? If so, what is the reason for this correlation?

3. P1 structure contradict the strong preference for given, topical elements to be fronted in a sentence, a principle instantiated by the strong tendency for languages to start the sentence with the subject (76% of languages in the WALS sample). Does this breach of a perhaps universal principle have any visible effect on language processing?

For such inquiries, and many other conceivable studies on the topic of P1, an appropriate representation of the world’s P1 languages is absolutely necessary. The present study, with its intentionally descriptive and theory-neutral approach, is intended to serve and be perused as a handy stepping stone and reference for further studies into the nature of P1.
References


Brustad, Kristen E. 2000. The syntax of spoken Arabic: a comparative study of Moroccan,


Campbell, Lyle. 2016. Comparative linguistics of Mesoamerican languages today. VELEIA
33. 113–134.
Cervin, Richard Stuart. 1990. Word order in Ancient Greek: VSO, SVO, SOV, or all of the above?: University of Illinois at Urbana-Champaign dissertation.
Chappell, Hilary. 2001. Synchrony and diachrony of Sinitic languages: A brief history of


Clemens, Lauren E. and Maria Polinsky. 2017. Verb-initial word orders (Primarily in Austronesian and Mayan Languages). *LingBuzz*.


Afrika und Übersee 80(1). 71–93.
de la Cruz Cruz, Victoriano. 2010. Las cláusulas relativas en el náhuatl de Teposteco, Chicontepec, Veracruz: Ciudad de México: Centro de Investigaciones y Estudios Superiores en Antropología Social MA thesis.
Dahlstrom, Amy. 1995. Topic, focus, and other word order problems in Algonquian. The Belcourt Lecture, delivered before the University of Manitoba.
Denzer-King, Ryan. 2010. Blackfoot word order in Uhlenbeck’s new series of Blackfoot
texts. Papers of the 41st Algonquian Conference.
(eds.), The Amazonian languages (Cambridge Language Surveys), 23–64. Cambridge:
Cambridge University Press.
Derbyshire, Desmond C. 1981. A diachronic explanation for the origins of OVS in some
Derbyshire, Desmond C. 1985. Hixkaryana and linguistic typology (Summer Institute of
Linguistics: Publications in Linguistics 76). Dallas, TX: The Summer Institute of
Linguistics and the University of Texas at Arlington.
Derbyshire, Desmond C. 1986. Comparative survey of morphology and syntax
in Brazilian Arawakan. In Desmond C. Derbyshire and Geoffrey K. Pullum
dissertation.
Dieterman, Julia Irene. 2002. Word order variation in Isthmus Mixe: Voice and discourse
sl.26.2.02die.
Academic Press.
Diller, Anthony Van Nostrand, Jerold A. Edmondson and Yongxian Luo. 2008. The Tai-
Dimmendaal, Gerrit Jan. 1998. A syntactic typology of the Surmic family from an areal and
historical-comparative point of view. In Gerrit Jan Dimmendaal and Marco Last (eds.),
Surmic languages and cultures (Nilo-Saharan: linguistic analyses and documentation
(ed.), Essais de typologie et de linguistique générale: Mélanges offerts à Denis Creissels,
147


Dum-Tragut, Jasmine. 2009. Armenian: Modern Eastern Armenian (London Oriental and


Early, Robert J. 1981. *Minor morphemes of the language of Tikopia:* University of Auckland
MA thesis.

dissertation.

Eatough, Andrew. 1999. *Central Hill Nisenan texts with grammatical sketch* (University of
California Publications in Linguistics 132). Berkeley and Los Angeles: University of
California Press.


University of London dissertation.

Ehret, Christopher. 1995. Do Krongo and Shabo belong in Nilo-Saharan? In Robert
Nicolai and Franz Rottland (eds.), *Actes du Cinquième Colloque de Linguistique Nilo-
Saharienne / Proceedings of the Fifth Nilo-Saharan Linguistics Colloquium, Nice, 24–29
Köln: Rüdiger Köppe Verlag.

Ehret, Christopher. 2002. Language family expansions: broadening our understanding of
cause from an African perspective. In C. Renfrew and P. Bellwood (eds.), *Examining
the farming/language dispersal hypothesis*, 163–176. Cambridge: McDonald Institute
for Archaeological Research.

Elbert, Samuel H. and Mary Kawena Pukui. 1979. *Hawaiian grammar.* Honolulu:
University of Hawaii Press.


Austin, TX: University of Texas Press.


Linguístico Francisco Marroquín. 2nd edn.

Lincoln: University of Nebraska Press.


Evans, Bethwyn and Bill Palmer. 2011. Contact-induced change in southern Bougainville.
*Oceanic Linguistics* 50(2), 483–523.

Everett, Caleb. 2006. *Patterns in Karitiana: articulation, perception, and gram-
mar.* Houston: Rice University dissertation. http://hdl.handle.net/11858/


Hanson, Rebecca. 2010. A grammar of Yine (Piro): LaTrobe University dissertation.


Harrison, Roy and Margaret Harrison. 1984. Survey of morphology and syntax for Zoque


Huber, Randall Q. and Robert B. Reed. 1992. *Vocabulario comparativo: Palabras selectas de lenguas indígenas de Colombia*. Santafé de Bogotá: Instituto Lingüístico de Verano (ILV). [http://hdl.handle.net/11858/00-001M-0000-0012-9AC4-8](http://hdl.handle.net/11858/00-001M-0000-0012-9AC4-8).


Huffmon, Herbert Bardwell. 1965. *Amorite personal names in the mari texts: a structural and lexical study*.


de Jong Boudreault, Lynda J. 2009. *A grammar of Sierra Popoluca (Soteapanec, a Mixe-Zoquean language)*. Austin, TX: University of Texas at Austin dissertation.


Kaufman, Terrence. 1990. Language history in South America: What we know and how


Mithun, Marianne. 1993. Prosodic determinants of syntactic form: Central Pomo


Karthala.


Reinholtz, Charlotte and Christoph H. Wolfart. 1996. Sentences [Sketch of Cree, an


Schachter, Paul. 1976. The subject in Philippine languages: topic, actor, actor-topic or non of the above? In Charles N. Li (ed.), *Subject and topic*, 492–518. New York / San


Smith, N. V. 1981. Consistency, markedness and language change: on the notion


Thomas, Michael F. 2014. *A grammar of Sakun (Sakur)*: University of Colorado at Boulder dissertation.


Vázquez Álvarez, Juan Jesús. 2011. *A grammar of Chol, a Mayan language*. Austin, TX: University of Texas at Austin dissertation.


