Linguistics Olympiad

Training guide

Vlad A. Neacșu

Textbooks in Language Sciences 13

language science press

Textbooks in Language Sciences

In this series:

1. Müller, Stefan. Grammatical theory: From transformational grammar to constraint-based approaches.
2. Schäfer, Roland. Einführung in die grammatische Beschreibung des Deutschen.
3. Freitas, Maria João & Ana Lúcia Santos (eds.). Aquisição de língua materna e não materna: Questões gerais e dados do português.
4. Roussarie, Laurent. Sémantique formelle: Introduction à la grammaire de Montague.
5. Kroeger, Paul. Analyzing meaning: An introduction to semantics and pragmatics.
6. Ferreira, Marcelo. Curso de semântica formal.
7. Stefanowitsch, Anatol. Corpus linguistics: A guide to the methodology.
8. Müller, Stefan. 语法理论: 从转换语法到基于约束的理论.
9. Hejná, Míša & George Walkden. A history of English.
10. Kahane, Sylvain & Kim Gerdes. Syntaxe théorique et formelle. Vol. 1: Modélisation, unités, structures.
11. Freitas, Maria João, Marisa Lousada & Dina Caetano Alves (eds.). Linguística clínica: Modelos, avaliação e intervenção.
12. Müller, Stefan. Germanic syntax: A constraint-based view.
13. Neacșu, Vlad A. Linguistics Olympiad: Training guide.

ISSN: 2364-6209

Linguistics Olympiad

Training guide

Vlad A. Neacșu

language
science
press

Vlad A. Neacșu. 2024. *Linguistics Olympiad: Training guide* (Textbooks in Language Sciences 13). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/420

ISBN: 978-3-96110-468-0 (Digital)
978-3-98554-097-6 (Softcover)
ISSN: 2364-6209
DOI: 10.5281/zenodo.10947862
Source code available from www.github.com/langsci/420
Errata: paperhive.org/documents/remote?type=langsci&id=420

Cover and concept of design: Ulrike Harbort
Translation: Pavel Iosad, Vlad A. Neacșu, Graeme Trousdale
Proofreading: Alexandra Fosså, Amir Ghorbanpour, Andreas Hölzl, Annika Schiefner, Heriberto Avelino, Annie Zaenen, Camil Staps, Christopher Green, Christopher Straughn, Elliott Pearl, Hannah Schleupner, Harold Somers, Hongyi Piao, Ikmi Nur Oktavianti, Ivelina Stoyanova, Janina Rado, Jean Nitzke, Jeroen van de Weijer, Katja Politt, Ksenia Shagal, Lachlan Mackenzie, Laura Arnold, Leonie Twente, Liam McKnight, Nicole Benker, Sarah Shan, Tom Bossuyt, Varun de Castro Arrazola, Wilson Lui, Yvonne Treis
Fonts: Libertinus, Arimo, DejaVu Sans Mono, Source Han Serif ZH, Source Han Serif KO, Source Han Serif JA
Typesetting software: XeLaTeX

Language Science Press
xHain
Grünberger Str. 16
10243 Berlin, Germany
http://langsci-press.org

Storage and cataloguing done by FU Berlin

Freie Universität Berlin

Contents

Note on the translation

This is a revised and extended translation of *Olimpiada de Lingvistică: Ghid de pregătire*, which was originally published in Romanian in 2022 with Editura Universității din București – Bucharest University Press. This translation is published by arrangement with Editura Universității din București – Bucharest University Press. The author is responsible for this translation from the original work and Editura Universității din București – Bucharest University Press shall have no liability for any errors, omissions or inaccuracies or ambiguities in such translation or for any losses caused by reliance thereon.

Preface

This book is a training guide (handbook) for the linguistics olympiads. It is the first such endeavour both nationally and internationally, since, at the moment, the only published materials contain collections of problems, rather than material to guide students and teachers, by presenting the theoretical framework and linguistic phenomena needed when solving linguistics problems, as well as different methods which can be applied when solving these problems.

All chapters follow the same structure: a short theoretical introduction followed by a detailed explanation of the less-known linguistic phenomena and further depicting these phenomena through one or more linguistics problems, solved step by step. At the end of each chapter there is a list of practice problems from national and international linguistics olympiads, accompanied by the answers, solutions, and, in some cases, additional explanations or discussions.

It is worth mentioning that this guide is not aimed at a certain level, but rather attempts to cover all relevant linguistic information from the very beginners all the way to those students who are training for the International Linguistics Olympiad (henceforth, IOL). The chapters and subchapters follow a logical order from a linguistic perspective, which does not necessarily correspond to the level of difficulty. This was done in order to attempt, as much as possible, a gradual display of terminology to avoid having to use certain linguistic terms before they are introduced and explained in detail.

When it comes to actively employing this book, it is best that students (as well as anyone else who wishes to start considering this type of problem) first attempt to solve the problems without checking the suggested solution and then read it, step by step, and check whether they reached the same results, the same rules, the same translations. Moreover, the way problems are approached in this guide is not necessarily "absolute" and many other approaches may exist which yield the same result. Therefore, I attempted as much as possible to approach each problem in a different manner, showcasing at the same time certain "tricks" or methods that might prove useful when attempting new problems. The problems in the "Practice problems" sections are mostly placed in order of difficulty, starting from problems suitable for beginners and ending with problems comparable to those from international olympiads. As such, it is perfectly normal that some

problems may seem extremely hard to solve and thus you might need to take a glance at the solution in order to get some hints. Once you have solved the problem, the most important thing is to attempt to fully understand the problem and the solution, both in terms of the linguistic phenomena that are showcased, as well as the way that you could approach similar problems in the future. This is a complex guide that presents the most common phenomena that have already been featured in linguistics problems, but it is far from being a complete guide, since new languages, features, phenomena, and even solving approaches are always discovered (and even emerging).

One of the main attributes of linguistics problems is that they can be solved without any additional information (solely based on the given data). Working through this training guide needs to be continuously complemented by solving problems in order to be able to perpetually practice the things you have learnt and to keep up with this ever-evolving competition.

Experienced linguists might notice that some aspects of the material presented herein do not always follow standard academic practices. This is because this book is not meant to be an academic work, but rather aimed at young people discovering things about language and how to approach linguistics problems. For example, readers may notice that figures are only labelled as such when they form part of the discursive text and not when they are part of the problem text or its solution. Similarly, certain linguistic concepts have been simplified in order to maximise the accessibility of the text.

Please read this book in the spirit in which it was written!

— Vlad A. Neacșu

For any suggestions, observations, or other inquiries, the author can be contacted directly at vlad.neacsu2009@gmail.com.

Foreword

This wonderful book presents the best of modern linguistics: a very attractive combination of rigour, information, insight, and fun. As chair of the UK Linguistics Olympiad, I commend it strongly to anyone with an interest in linguistics olympiads. I also congratulate Vlad Neacşu on producing such a useful guide and thank him for making its translation freely available to anyone who can read English.

I listed four attractive qualities of modern linguistics: rigour, information, insight, and fun. All four make linguistics an outstanding candidate for inclusion in the school curriculum, so I should like to take this opportunity to justify my claims.

Rigour is obvious. Our champions tend strongly to be mathematicians with a gift for spotting abstract patterns and building a logical chain of reasoning. This book takes the reader by the hand through some of the most daunting problems available, showing how each one allows just one correct solution, and sometimes just one possible route to that solution. Constructing one of these problems is an intellectual triumph in itself, so they testify to three sources of rigour: in the problem-creator, in the problem-solver, and in the problem-explainer. The explanations are brilliantly clear and untechnical.

The trouble with rigour is that for some people it doesn't come naturally. It is all too easy to find olympiad problems that look, at first sight, as though an analysis is impossible. Where to start? How to get a foothold on the data? This is where the book is particularly strong, because, after introducing a simple classification of problems, it provides tactics for each type of problem: counting, drawing graphs, constructing tables and so on. Just what a beginner needs, and a wonderful source of confidence. So instead of relying solely on native ability, we approach problems with a toolbag full of helpful ideas and skills.

Rigour is an important attribute of linguistic analysis because this is what puts linguistics on the same level as mathematics. Our schools in the UK give high status to mathematics and science, but low status to language; one of the arguments for mathematics is its rigour and its effect on thinking skills, so we now have a similar argument for language – but only if we stop thinking about language simply as a useful skill, and start viewing it as a worthy object of rigorous study.

Information is available in spadefuls. The book is an elementary introduction to the variety of human language, with plenty of opportunities for exploring this variety in datasets from an amazing range of languages. All the problems were built by professional linguisticians from across the world, and, since every language is unique, they all take us to the frontiers of research. The book takes us through the many variables that are familiar to linguistic typologists, but without straying into high theory – everything is tied firmly to the evidence in the data.

Every problem in this book introduces a language system which, in some respect, is different from English. The differences cover all the levels of language, from writing systems and phonetics to semantics, and in every problem we discover the regularities that lie behind the apparently chaotic data. The variety that emerges is truly stunning and must be part of the reason why children find linguistics olympiads so gripping.

These problems manifest in very concrete ways the claim that learning a foreign language takes us into the different mental world of its speakers. But unlike a foreign-language lesson, we reach that world and explore it in just a few minutes and without any drills or memorisation. Lip service is often paid in educational circles to the goal of Language Awareness, but this goal conflicts too often with that of teaching only for communication. Language teaching would be much more successful if our children spent as much time on linguistics olympiads as they do in communicative classrooms where they fail to learn even a single language.

Insight takes us into the workings of language. We all have expert knowledge of at least one language, but without linguistics, we don't know how it works. Struggling through one of these problems gives us insight into one small part of the great machine of language; for example, any script problem forces us to confront issues such as syllable boundaries, classification of sounds, morphological and semantic analysis, and all the possible relations between characters and linguistic units. It's all too easy for a child to believe that a word's pronunciation and spelling are the same thing, and to be surprised that *though* only has two sounds. Any activity, such as an olympiad problem, that problematises this simple view is to be welcomed.

The main insight promoted here is what we call the architecture of language – how the total structure is divided into the traditional levels of phonetics, phonology, morphology, syntax, and semantics. This rather sophisticated view is the more or less uncontroversial basis for linguistics, and should be part of the worldview of any educated citizen; but while we are waiting for our schools to catch up, the olympiad is the main, or even only, tool for teaching it.

And last, but by no means least, we have *fun*. The popularity of the linguistics olympiad shows that school children enjoy grappling with linguistic analysis. Our teachers repeatedly report that children are excited, engaged, and enthusiastic; and many thousands of them come back for more, year after year (in 2023 the UK Olympiad attracted over 5,000 competitors, all willing volunteers). Our problems offer the abstract intellectual challenge of sudoku combined with the complex strategic thinking of chess and the excitement of exploring a foreign country.

This enthusiasm for linguistics olympiads is all the more remarkable given not only the total absence of linguistics from the school curriculum but also the deep unpopularity of foreign languages among our school children, who tend to find them both difficult and boring. Maybe it's time for language teachers to pay more attention to the system of the target language as something that the learners might enjoy exploring.

Which brings us back to this book. Children enjoy cracking codes and exploring the intellectual territory behind the codes, but their enjoyment depends on making progress. Nobody is inspired by the soul-destroying experience of staring at a page of data for half an hour without making any sense of it at all; so children need intellectual tools to help them on their way. This book is just what they need.

— Richard Hudson
Chair of the UK Linguistics Olympiad
Emeritus Professor at University College London

Acknowledgments

First and foremost, to Richard Hudson, Graeme Trousdale, and Pavel Iosad, for making this version of the book happen, for their help with the English language, as well as for their scientific proofing.

For their help and assistance: Monojit Choudhury (India), Ivan Derzhanski (Bulgaria), Ksenia Gilyarova (Russia), Minkyu Kim (South Korea), Tamila Krashtan (Ukraine), Elena Muravenko (Russia), Tom Payne (USA), Jan Petr (Czechia), Dragomir R. Radev (USA), Ksenia Shagal (Finland/Germany).

To the authors of the problems: Peter Arkadiev (3.7, 4.7, 5.9, 6.3, 6.11, 7.5, 9.16), Vladimir I. Belikov (4.2), Aleka Blackwell (5.12), Michał Boroń (10.2), Bozhidar Bozhanov (7.9), Svetlana Britova (6.10), Svetlana Burlak (9.12, 9.16), Monojit Choudhury (2.8, 8.4), Theodor Cucu (6.14), Ivan Derzhanski (2.1, 6.12, 9.17), Roxana Dincă (3.6, 8.5, 9.1, 9.14), Sergey Dmitrenko (2.11), Barbora Dohnalová (10.4), Grigory Durnovo (7.7), Mirjam Fried (9.7), Rujul Gandhi (5.17), Ksenia Gilyarova (4.11, 6.2, 8.1, 9.4, 10.5, 10.8, 10.9), Elena Gurevich (3.7), Paul Helmer (4.4, 6.9, 8.6), John Henderson (4.8), Adam Hesterberg (10.6), Bill Huang (9.5), Shen-Chang Huang (7.10), Richard Hudson (7.3, 7.4), Tsuyoshi Kobayashi (4.14), Evgeniya Korovina (4.6), Tamila Krashtan (4.10), Alexey Kretov (8.8), Anton Kukhto (4.1), Michal Láznička (6.15), Tae Hun Lee (2.10), Kevin Liang (3.9), Patrick Littell (2.4, 2.7, 5.15, 10.1), Kai Low Rui Hao (5.4, 9.3), Timur Maisak (7.8), Tom McCoy (4.3), Dan-Mircea Mirea (3.5, 10.7), Danylo Mysak (7.11, 9.15), Heather Newell (6.5), Viktoria Papp (5.6), Gábor Parti (2.9), Tom Payne (5.16), Aleksejs Peguševs (8.7), Jan Petr (4.13), Alexander Piperski (3.1, 9.13), Dragomir R. Radev (2.5, 9.11), Maria Rubinstein (5.14), Michael Salter (3.3, 5.10, 5.12), Nilai Sarda (9.8), Artūrs Semeņuks (3.4), Catherine Sheard (6.7, 10.3), Ronnie Sim (6.1), Harold Somers (7.4, 7.6, 9.6), Anton Somin (4.9), Artur Corrêa Souza (8.2), Simona Strizhevskaya (5.13), Michaela Svatošová (6.8), Michael Swan (9.9), Todor Tchervenkov (5.7), Yakov Testelets (6.4, 6.6), Saujas Vaduguru (3.2, 3.8), Martin Vasev (5.4), Babette Verhoeven-Newsome (2.6, 5.11, 9.10), Alex Wade (7.1), Natalia Zaika (8.3), Alfred Zhurinsky (2.2).

Abbreviations and notes

Abbreviations used to indicate the source of the problems

APLO	Asia-Pacific Linguistics Olympiad
Elementy	published on http://www.elementy.ru
HKLO	Hong Kong Linguistics Olympiad
IOL	International Linguistics Olympiad
TurLom	Lomonosov Academic Tournament (Russia)
NACLO	North American Computational Linguistics Olympiad
ČLO	Czech Linguistics Olympiad
JOL	Japan Linguistics Olympiad
LLO	Latvian Linguistics Olympiad
RoLO	Romanian Linguistics Olympiad
MSK	Moscow Traditional Linguistics Olympiad (Russia)
UkrLO	Ukrainian Linguistics Olympiad
PLO	Panini Linguistics Olympiad (India)
Princeton	Princeton Linguistics Club (USA)
UKLO	United Kingdom Linguistics Olympiad

The year mentioned in the beginning of each problem refers to the year in which the problem was used, synced with the corresponding IOL edition, e.g., if the problem was used in the UKLO in the academic year 2002–2003, the problem will be marked as UKLO 2023.

Note about problems

Problems from APLO, HKLO, IOL, NACLO, PLO, Princeton, and UKLO were used in the original English version. All the other problems were translated by the author.

Most of the problems in the present volume were previously used in national and international competitions and were based on grammar books, grammar sketches, or dictionaries of that language, sometimes complemented by information found on the internet, official websites and so on. For example, Problem

4.6 was based on Li (1985), while Problem 6.13 was based on Refsing (1986) and Shibatani (1990).

It is important to mention that these problems were created for the purpose of linguistics contests and to test the students' logical reasoning abilities. While the data are definitely not fictitious, some small simplifications such as changes in orthography may have been made in order to make the problem accessible to the target group.

1 Introduction to linguistics problems

Linguistics problems are puzzles, logic games usually based on real languages, which allow the discovery of certain linguistic phenomena through logical reasoning. The problems usually consist of a corpus (a dataset) in an unfamiliar language accompanied by their English translations (either ordered or in random order). Solving the problem requires logical thinking and attention to detail in order to be able to decode certain aspects of the language, such as the meaning of certain words or some grammar rules.

It is important to note that all linguistics problems are internally consistent (self-consistent) and require no additional knowledge (self-sufficient), i.e., once certain rules are discovered, they will apply to all examples in the problem (of course, some rules might have exceptions) and all the information needed to solve the problem is found in the problem.

1.1 Structure of linguistics problems

All linguistics problems have the same structure, consisting of four parts (excluding the title and the author):

1. Introduction

   In the introduction, we learn about the language featured in the problem. For most problems, the introduction is simply a sentence like "Given below are some [words/structures/constructions/sentences, etc.] in [Language] and their English translations (in random order)."

   Some problems might have a more complex introduction which also includes information about the culture of the people speaking that language or information about certain characteristics of the language. Generally speaking, if the introduction is complex and contains additional information, this information is likely to be relevant to solving the problem.

2. Dataset

   This part contains the examples based on which we should solve the tasks. This part is also known as the *corpus*. If the corpus and the translations are

given in order (it is known which translation corresponds to which structure), the problem is called a *Rosetta stone problem*, while if the translations are given in random order, we are talking about a *chaos-and-order problem*.

3. Tasks

   The corpus is followed by the tasks. For chaos-and-order problems, the first task will always be "Determine the correct correspondences". Therefore, an easy way to figure out whether the problem is Rosetta stone or chaos-and-order (besides reading the introduction) is by checking the first task.

   The subsequent tasks will be "Translate into English" and "Translate into [...]". As a rule of thumb, the task asking to translate into English will precede the other one because (1) in order to translate into English, it is not always necessary to understand all the grammar rules, and (2) we might be able to use these additional examples in order to gather more information.

   Some problems might also have special tasks, which usually offer important hints about the phenomena featured in the problem.

4. Notes

   At the end of each problem, there will be some notes which provide three types of information. Firstly, there will be some data about the language featured in the problem, such as where it is spoken, how many people it is spoken by, what language family it belongs to, etc. In general, this information is not relevant to solving the problem, although, for experienced solvers, the family to which the language belongs might offer additional helpful information. In this book, this information has been removed from the notes and all the information regarding the languages can be found in Appendices A and B.

   Relevant phonetic information follows, which offers details regarding how some letters, characters, symbols are pronounced. Moreover, it might also include details regarding the existence of diphthongs, or details on additional writing notations used in the problem. Depending on the problem, this kind of information may or may not be useful.

   Finally, there might be some information about specific words, usually referring to types or species of animals or plants, traditional objects or garments, etc. Again, depending on the type of problem, this information may or may not be useful in solving it.

1.2 Classification of linguistics problems

Generally, linguistics problems are classified based on the main phenomenon that is featured (which is closely related to a specific field of linguistics). In this book, the problems are divided into seven categories, as follows:

1. Writing systems (Chapter 2)

   These problems feature words in an unfamiliar writing system together with their transliterations in Latin script. These problems can be either Rosetta stone or chaos-and-order.

2. Phonetics and Phonology (Chapters 3 and 4)

   These problems are based on the sound changes that occur in different environments or contexts (e.g., two or more forms of a word – such as singular–plural, different noun cases, declensions, etc. –, the way certain words change in different dialects, how words are transcribed phonetically, how words are stressed, etc.). Generally, these problems are Rosetta stone.

3. Morphology (Chapters 5 and 6)

   These problems are subdivided into two categories: morphology of the noun and its relation to other elements in the noun phrase (Chapter 5) and morphology of the verb and its relation to other elements in the verb phrase (Chapter 6). These problems can be either Rosetta stone or chaos-and-order.

4. Syntax (Chapter 7)

   Syntax problems contain sentences (or phrases), combining, to some extent, the morphology of the noun with the morphology of the verb. In most cases, these problems are Rosetta stone, mainly for simplicity. If the sentences were in random order, once the correspondences are made, all sentences need to be copied again – with their corresponding translations – in order to better see all the data and extract all the rules and phenomena that occur; the copying of the sentences would require a lot of time.

5. Semantics (Chapter 8)

   Semantic problems are not based on grammar rules, but rather on associations of words which share different properties related to their meaning. These problems are always chaos-and-order problems.

6. Number systems (Chapter 9)

   This kind of problem contains numbers written out in a certain language and their numeric representation. The "translations" might or might not be ordered. Sometimes, the corpus might consist solely of some mathematical equalities, in which the numbers are written out in a certain language.

7. Other types of problems

   This category is rather large and includes any problem that does not fit in the aforementioned categories. Nevertheless, even in this category, we can notice some types of problems that appear frequently in the linguistics olympiads. As a result, we can differentiate:

   7.1. Metrics and prosody (Chapter 3) – in which the corpus consists of lines from different poems and the purpose is discovering the general structure of the verse.

   7.2. Time problems (Chapter 9) – these problems highlight how the calendar dates or the time are told in different languages.

   7.3. Kinship problems (Chapter 10).

   7.4. Orientation system problems (Chapter 10) – which show how directionality is represented in different languages.

Of course, there can also be *mixed* problems, which combine two (or more) of the above categories.

Statistically speaking, based on 437 problems from different national and international linguistics olympiads, we notice that the most common type of problem is the syntax one (accounting for approximately 19.5% of the problems). Indeed, syntax problems are usually pervasive in all olympiads and each olympiad will have at least one syntax problem. Morphology problems (both nominal and verbal) are almost as common as syntax ones, accounting for 19% of the total. The other categories are phonetics and phonology (13.5%), writing systems (12%), number systems (9%) and semantics (7%).

1.3 Understanding the problem

When solving a linguistics problem, it is important to understand what exactly is expected from us. In most cases, the way the tasks are phrased (especially the special tasks, as mentioned above), but even the way the corpus is chosen, can offer important hints. For example, let us take a look at the way the following task is phrased (the whole problem is presented in Chapter 5, Problem 5.13):

Example task 1.1 Fill in the blanks:

16.	*baqra* __(7)__	'blue cow'
17.	*fjuri* __(8)__	'red flowers'
18.	*kelb* __(9)__	'brown dog'
19.	*kotba* __(10)__	'yellow books'
20.	*siġra* __(11)__	'green tree'
21.	*mwejjed* __(12)__	'purple chairs'
22.	*tuffieħa* __(13)__	'yellow apple'

The thing that strikes us the most is that the first word is always given, and we only need to be concerned about the second. Normally, in a *classical* problem, we would simply be asked to translate the structures 'blue cow', 'red flowers', etc. Since in this case the first word (which, from the full dataset – not shown here –, we can see evidently represents the noun) is already given, we infer that, most likely, in this language (or at least based on the information given) the noun plural formation is irregular (or simply too complex) and does not follow specific rules, thus not being able to infer the singular form from the plural form or vice versa. Therefore, since the nouns are already mentioned, we know that the core phenomenon of the problem focuses on the adjective, rather than on the noun.

Let us consider the following (fictitious) example which would correspond to a writing system problem:

Example task 1.2 Write in the [...] script: *Mars*, *venus*, *JUPITER*, *NePtUne*.

In this case, we deem unusual the writing of these words, some of them being just lowercase, others just uppercase, and others being written with a combination of the two. There is no plausible reason to do so unless the writing system differentiates between lower- and uppercase letters. Therefore, in this case, the choice of the tasks (and their form) offers us a valuable clue: most likely there is a difference between lower- and uppercase letters.

Another task might be:[1]

Example task 1.3 Knowing that in Turkish *dil* = 'language', translate 'linguist', 'mute'.

In this case, we are given a new word ('language') and we are asked to translate two words which belong to the same semantic field. Therefore, in order to translate 'linguist', there must be a rule which allows the formation of an agent

[1]From a problem by Bozhidar Bozhanov (UKLO 2010).

noun (the person which...) and similarly, in the case of 'mute', we need to find a "negative"-forming particle (which marks the impossibility/lack of/incapacity etc.), both of which must be inferred from the data given.

As a result, it is important to read all the data and tasks carefully and ask ourselves whenever we see something slightly peculiar: *Why is it like this?*

1.4 Solution writing

Solution writing is a core part of problem-solving. It is important to write all the rules clearly and consistently and to cover all the phenomena featured, but, at the same time, to write them succinctly enough to not waste precious time during an official competition.

Officially, in the guidelines of the IOL (and in the case of most linguistics competitions) it is stated that: "Unless stated differently, you should describe any patterns or rules that you identified in the data. Otherwise, your solution will not be awarded full marks."

The most important thing we need to understand is that we need to write *the rules that we identified* and not *how we found them*. Therefore, in the linguistics competitions, we are not asked to provide our reasoning for inferring the rules, but rather only to write the actual rules.

When writing the solution, *we should*:

- use tables, graphs, diagrams or any other kind of concise representation;
- use common abbreviations and symbols (we may also use less common abbreviations as long as we make a legend describing what they stand for);
- explain all the rules and phenomena that occur.

Briefly, through orderly and concise writing of all the rules, we ought to try and *tell the story of the language* we discovered.

At the same time, when writing the solution, *we should not*:

- explain how we inferred or deduced the rules and patterns;
- write dictionaries and explain the meaning of every single word (we will talk more about this in the next section);
- use connectors and excessive words, such as: "I think", "we deduce", "it is obvious that", "since", "therefore", "because", "it is possible that", etc.

Therefore, when providing a solution, we should not write something like "Since examples 1, 3, and 5 all contain the word *mi* and all the English translations of these examples end in a question mark (thus being interrogative constructions), we most likely can infer that this word marks the fact that the structure is an interrogative one" since the same explanation can (and should) be briefly written as: "*mi* = question".

1.5 Dictionary vs. rules

We mentioned above that a solution should not include the *dictionary*. By dictionary we mean the base words (or stems) such as nouns, pronouns, adverbs, etc., *whose form does not change*.

On the other hand, the *rules* (which we need to write) explain the alternations that occur in the language. Therefore, they explain the word order, the way words change depending on their number, gender, tense, etc. We also include here all the words that do not have a direct English translation (usually they represent words that have a function rather than a meaning). For example, in Chinese, the character 吗 placed at the end of the sentence signals that it is a direct question (requiring a yes/no answer). Therefore, since this character has a function (marks the interrogation) and not a meaning (it does not mean anything and it would not be found in a dictionary), it must be included in the solution writing.

Let us consider the following dataset from the Turkish language, and imagine we are asked about how possession is marked in Turkish:

babam	'my father'
kedin	'your_{SG} cat'
kedimiz	'our cat'
baban	'your_{SG} father'
kedi	'his cat'

In this case, we can easily observe that the possessive is marked with a suffix (attached at the end of the word) as follows: *-m* for 'my', *-n* for 'your_{SG}', *-miz* for 'our', while for 'his' nothing is added – in fact, it is important to specify that "zero" is added or, in other words, that a null morpheme is used to mark the equivalent of 'his' in English.

For this problem, the dictionary is: *baba* = 'father' and *kedi* = 'cat' (these words are invariable). Therefore, a correct and complete way of writing the solution is:

(1) possession suffixes: *-m* for 'my', *-n* for 'your_{SG}', *-miz* for 'our', ∅ for 'his'

Note

The symbol ∅ marks the fact that nothing is added (it represents the null morpheme). This symbol does not need an explanation/legend when used.

Another, briefer, way to write the solution is:

(1′) possession suffixes: *-m* = 1SG, *-n* = 2SG, *-miz* = 1PL, ∅ = 3SG

The simplest way to write these suffixes is by creating a table which includes the persons (1, 2, 3) in the rows and the numbers in the columns (singular, plural). Therefore, we can also write:

(1″) possession suffixes:

	SG	PL
1	*-m*	*-miz*
2	*-n*	
3	∅	

In this case, we can see the importance of using the null morpheme (∅). Thanks to it, we can deduce (based on the table above) that there is a difference between the possessive suffix for 2PL (the cell is blank – therefore we cannot deduce it based on the data given) and for 3SG (the cell is not empty, it contains the symbol ∅, thus proving that we discovered the way it is marked).

Let us now consider the following three sentences in Spanish, and imagine we are asked to describe the word order:

Tu marido corre. = 'Your$_{SG}$ husband runs.'
Él ve a tu marido. = 'He sees your$_{SG}$ husband.'
Mi novio ve a él. = 'My boyfriend sees him.'

The rules we need to write concern the order of subject, verb and object and the order of the possessor and the possessed. Therefore, the solution is:

(2) Word order: S V (*a* O); Possessor – Possessed

The rules above contain a lot of relevant information, written very briefly:

- The word order is S(ubject), followed by V(erb), followed by the particle *a* and finally followed by the O(bject);
- The particle *a* only appears together with the object; if the sentence has no object, the particle *a* is not used, a fact marked by the use of brackets around the structure;
- In a possessive construction, the possessor (owner) is placed before the possessed object.

We need not mention anything about the verbs since all of them are in the third person singular (3SG) present tense: therefore, we do not know how (or if) they change form in any way.

Let us now consider three more sentences (as a supplement to those above):

Él ve tu casa.	=	'He sees your$_{SG}$ house.'
Ella ve a su padre.	=	'She sees her father.'
Yo veo tu libro.	=	'I see your$_{SG}$ book.'

Based on these extra sentences, it is important to check, first of all, whether the rules we wrote before still hold for these examples as well. Therefore, we now see that *a* does not appear every time before the object, but it is only used when the object is human. Therefore, the rules become:

(2′) Word order: S V O; Possessor – Possessed
If O = person, add *a* before it.

Moreover, we see this time that the verb changes, having the pair *ve* ('he/she sees') and *veo* ('I see'). Therefore, we also need to pay attention to verbal morphology. If we do so, we can deduce the conjugation rules for the verb: *-o* = 1SG, ∅ = 3SG.

Thus, the final rules are:

(2″) Word order: S V O; Possessor – Possessed
If O = person, add *a* before it.
Verb: 1SG = *-o*, 3SG = ∅

Let us now consider the following possible task:

Example task 1.4 There is an error in the following data. What is it?

Mi padre correo.	=	'My father runs.'

Since we are told that there is an error in the sentence, we need to check the rules we have in order to see if any of them could justify this task. We remember that the *-o* ending of verbs is for 1SG subjects, but here the subject is 3SG, meaning there should be zero inflection on the verb, so the Spanish sentence should read *Mi padre corre*.

In the following chapters, each problem will be accompanied by a solution so that the reader can get accustomed to different ways of writing the solution.

2 Writing systems

2.1 Introduction

Writing systems (or scripts) are collections of symbols and the rules for combining them in order to represent language. By some counts, there are now over 3,500 writing systems in the world. We will avoid the term *letter* to refer to these symbols since they might represent not only individual sounds (see Chapter 3, *Phonetics*) but also syllables or even whole words. Therefore, the term *character* is preferred.

2.2 Pictographic and ideographic systems

Etymologically, the terms *pictographic* and *ideographic* are compounds formed from *picto-* ('picture'), *ideo-* ('idea'), and *grafos* ('writing'). In these systems, each character represents a word, an idea or a concept. Moreover, the characters are similar in appearance to the real-life representation of these concepts.

Although there are notable differences between the picto- and ideographic systems, these are not very relevant when it comes to linguistics problems. For this reason, we will combine these two types of script in a single category.

We often encounter these kinds of systems in contexts where there is a need to convey some idea that is not specific to a particular language. Many such symbols are widely used across cultures, such as the *P* symbol used to represent a parking lot. A good example of a pictographic system is traffic and public signage (such as a crossed-out ice-cream cone to show that it is forbidden to eat food).

In linguistics problems these types of system are quite rare: usually, the relationship between the character and its meaning is so straightforward as to make the solution essentially effortless.

2.3 Logographic systems

In some ways, such scripts are highly similar to picto- and ideographic systems. The word *logographic* is made up of two etymons: *logos-* 'word' and *grafos-* 'writing': in these systems, at least in principle, each character represents a word. In

most cases, the logographic systems have their origin in picto- or ideographic systems, but the characters have evolved over time, thus losing their similarity with the real-life representation of the words they designate. For instance, many Chinese characters have developed from pictographic representations to logographic ones. Thus, the character for 'sun', which used to be represented as something like ⊙, is nowadays (in Modern Chinese) written as 日. Similarly, the character for 'moon/month', initially represented as something like ☽, has become 月 in Modern Chinese.

2.4 Syllabic systems (syllabaries)

Here, each character represents a syllable. Crucially, unlike some scripts that we will present below, each syllable is represented as a whole, without any clear relationship between characters denoting syllables that share some vowels or consonants.

One example of a syllabary is *katakana*, one of the writing systems in use in Japan.

The following table gives some examples of katakana characters. Each character represents a syllable, consisting of a consonant and a vowel. Conventionally, these are written in a table in which the vowel spans across columns and the consonant across rows:

	a	*i*	*u*	*e*	*o*
∅	ア	イ	ウ	エ	オ
k	カ	キ	ク	ケ	コ
s	サ	シ	ス	セ	ソ
t	タ	チ	ツ	テ	ト

We can easily observe that the syllable is treated as a whole, and cannot be further divided into smaller components. For instance, the first line in the table corresponds to the absence of consonant (the character ウ represents the syllable *u*), but the other characters do not look anything like these symbols. Similarly, from knowing the characters for *ka* (カ), *ki* (キ), and *si* (シ) we cannot deduce the character for *sa* (サ).

2.5 The Chinese writing system

There are a lot of misconceptions about the Chinese writing system. It is erroneous to consider it either pictographic or ideographic: Chinese characters are

not “pictures” representing meanings directly. Neither is it entirely logographic. In fact, it is a logo-syllabic system (a logo-syllabary). Some characters are indeed purely logographic (as we mentioned before), but they make up a very small percentage of the total number of characters. Most of the characters are formed by what we call the *rebus principle*, where each character is formed by two components: a logographic part, approximately showing the meaning of that character (called the *semantic* component) and a syllabic part, giving clues about the pronunciation of that character (called the *phonetic* component).

For example, the character *ma*^{3} (马) is purely logographic and represents the word ‘horse’, while the character for ‘mother’ (妈, *ma*^{1}) is logo-syllabic.[1] It is formed from the semantic component (cf. the logographic 女 *nü*^{3} ‘woman’) and the phonetic component 马. Therefore, 妈 is a character whose meaning is related to ‘woman’ and whose pronunciation is similar to that of 马. Some more examples of logographic characters created by the rebus principle are:

机 (*ji*^{1}, ‘machine’) = 木 (*mu*^{4}, ‘wood’) + 几 (*ji*^{3}, ‘some’)
唱 (*chang*^{4}, ‘to sing’) = 口 (*kou*^{3}, ‘mouth’) + 昌 (*chang*^{1}, ‘prosperity’)

2.6 Alphabetic systems (alphabets)

These are the most common systems in Europe and each character represents either a consonant or a vowel. In other words, each character represents a sound. Examples of these scripts are the Latin alphabet (used to write languages such as English, Romanian, Spanish, Italian, German, Polish, Turkish), the Greek alphabet (used for writing Greek), the Cyrillic alphabet (used to write some Slavic languages such as Russian, Ukrainian, Belarusian, Bulgarian, as well as many languages spoken in Russia), the Georgian alphabet (used for the Georgian language), and the Armenian alphabet (used for the Armenian language).

2.7 Abjad systems

In these systems, each character represents a consonant, while vowels are not written at all or shown as diacritics (i.e., small marks attached to the character). In broad terms, this type of system can be compared to syllabic systems, in the

[1] To transcribe Chinese words, we use a modified version of the *pinyin* transcription system: the letters indicate the sounds and the digit indicates TONE, a distinctive property of the entire syllable that can distinguish meaning (see Section 3.6). In Chinese, syllables with different tones differ in the level and the trajectory of the voice’s PITCH.

sense that each "complete" character represents a combination of a consonant and a vowel; but, unlike syllabaries, each component is represented separately, so the character can be broken down into subparts.

Some well-known abjads are used for Semitic languages, such as the Arabic and Hebrew scripts. Below we show some Arabic characters:

	∅	*u*	*ū*	*a*	*ā*	*i*	*ī*
∅		◌ُ	◌ُو	◌َ	◌َا	◌ِ	◌ِي
b	ب	بُ	بُو	بَ	بَا	بِ	بِي
s	س	سُ	سُو	سَ	سَا	سِ	سِي

The dotted circle is used to show the placement of the diacritics for *u*, *a*, and *i* relative to the consonant character: in Arabic, the same diacritic is used to mark both *a* and *i*, but in the former case it is placed above the consonant, and in the latter case it goes below the character marking the consonant.

2.8 Abugida systems

These systems are very similar to abjads, in the sense that each "complete" character represents a combination of a vowel and a consonant. However, in an abjad each character (without diacritics) represents a consonant and the diacritics, when used, append the corresponding vowels. In abugidas, a character without any diacritics or modifications represents a consonant followed by some vowel of the language (called the "default" or "inherent" vowel), and additional modifications are used to change that vowel to another one. Moreover, most abugidas will have some device to "delete" the vowel, so as to represent the consonant on its own.

Many languages of South and South-East Asia are written using abugidas. Here are some Burmese characters:

	a	*aa*	*i*	*ii*	*u*	*uu*	*e*	*ai*	∅
∅	◌	◌ာ	◌ိ	◌ီ	◌ု	◌ူ	◌ေ	◌ဲ	◌်
k	က	ကာ	ကိ	ကီ	ကု	ကူ	ကေ	ကဲ	ကင်
s	ဆ	ဆာ	ဆိ	ဆီ	ဆု	ဆူ	ဆေ	ဆဲ	ဆင်

We can observe that the default vowel is *a* (representing the basic form of each consonant). Adding a loop will change the vowel from *a* to *aa* (long *a*) while adding an L-like symbol below changes the vowel from *a* to *u*, etc.

A very interesting abugida system is Cree Syllabics, used to write aboriginal languages of Canada. In this system, the vowel change is not shown by diacritics, but rather by rotating the character (by 90°, 180° or 270°) or by mirroring it, depending on its shape. Below we show a table with some Cree characters.

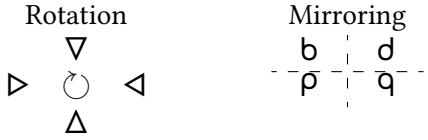
Rotation

	a	*e*	*i*	*o*
∅	ᐊ	ᐁ	ᐃ	ᐅ
p	ᐸ	ᐯ	ᐱ	ᐳ
t	ᑕ	ᑌ	ᑎ	ᑐ

Mirroring

	a	*e*	*i*	*o*
k	ᑲ	ᑫ	ᑭ	ᑯ
m	ᒪ	ᒣ	ᒥ	ᒧ
n	ᓇ	ᓐ	ᓂ	ᓄ

We can observe, conceptually, that the preference is for mirroring (right-hand table) since sometimes the same shape is used for two different consonants (*k* ᑲ, vertically, and *n* ᓇ, horizontally). The characters obtained by rotation are used only where mirroring is not possible due to symmetry reasons (a mirrored ᐁ is still ᐁ). Therefore, we can deduce that if a character cannot be mirrored (it has a symmetry axis), the vowel change will be represented by a clockwise 90° rotation (vowels change in the sequence $a \longrightarrow i \longrightarrow o \longrightarrow e \longrightarrow a$). The change of vowel by mirroring is performed either vertically ($a \Longleftrightarrow o$ and $e \Longleftrightarrow i$), or horizontally ($a \Longleftrightarrow i$ and $e \Longleftrightarrow o$). The rotation and mirroring patterns of the Cree characters are shown below.



2.9 Featural systems

This is the last type of writing system and probably the least common. In this system, each character highlights the phonetic or phonological features of the sounds it designates. For example, in the Korean writing system (*hangul*), the characters corresponding to the sounds *p*, p^*, p^h, and *m* (ㅂ, ㅃ, ㅍ, ㅁ, respectively) are highly similar: all of them derive from a square. This is because all these sounds are bilabial (as we will see in the next chapter). When pronouncing them

we use our lips (see the next chapter for more details on how to describe different sounds) and the square is meant to evoke the mouth seen from the front.

The property of being featural cuts across the other criteria we have used: the main characteristic of these systems is the fact that the properties of symbols are based on features of the sounds they represent, not the nature of the mapping between character and sound. For instance, the hangul writing system marks each character, whether vowel or consonant, individually, so it can also be considered an alphabet.

Problem 2.1

Hmong (Ivan Derzhanski, MSK 2003)

Here are several words and phrases in the Hmong Daw language written in Shong Lue Yang's script and the missionaries' alphabet, as well as their English translations:

1.	𖬀 𖬯𖬇 𖬵𖬍	*kev ntsuas no*	'degree'
2.	𖬊𖬏	*hauv*	'inside'
3.	𖬊𖬍 𖬮𖬍 𖬵𖬀	*raug raws cai*	'legal'
4.	𖬄𖬇 𖬰𖬏	*hloov mus*	'transfer'
5.	𖬴𖬆	*qhua*	'guest'
6.	𖬶𖬄 𖬶𖬊 𖬷𖬍	*yog los nag*	'it is raining'
7.	𖬑 𖬳𖬄	*kwv yees*	'guess'
8.	𖬅𖬍 𖬁𖬀 𖬊𖬊	*ris ceg luv*	'Bermuda shorts'
9.	𖬷𖬍	__(1)__	'bird'
10.	𖬑𖬀	__(2)__	'lobster'
11.	𖬶𖬏 𖬰𖬊	__(3)__	'speak'
12.	𖬶𖬆 𖬷𖬏 𖬅	__(4)__	'dizzy'
13.	__(5)__	*hluav*	'ash'
14.	__(6)__	*li cas*	'how?'
15.	__(7)__	*neeg ntse*	'smart, wise'
16.	__(8)__	*yawg*	'grandfather'

Problem 2.1a Fill in the blanks.

In the missionaries' alphabet, the letter *w* represents a specific vowel. The letters *g*, *s*, *v* at the ends of the syllables are not consonants; instead, they denote tones (specific ways of pronouncing the vowels).

Solution

First thing we need to notice is that we do not have to provide English translations. This is one of the main characteristics of writing system problems. Moreover, the fact that we are not asked to provide English translations means that the translations are probably not relevant to solving the problem.

Note

This is not always true; it is just a rule of thumb. It is possible for some problems that, although no translations are required, they can still be relevant – for example, based on semantic considerations.

We also need to keep in mind that for writing-systems problems the writing direction is relevant (from left to right or from right to left). Moreover, we notice that in Hmong the characters are grouped in clusters of one or two, while in the Latin transcription, they are grouped in syllables. Therefore, we can deduce that each group of characters represents a syllable.

We can begin by noticing the diacritics placed above some syllables. Since the mark ◌̈ appears the greatest number of times, we can begin with it and observe that it is transliterated by the letter *g* at the end of the syllable. Moreover, reading the footnote, we find out that this letter does not represent a consonant or a vowel per se, but rather the syllable tone. Finally, based on example 3, in which the syllable containing the *g* tone appears first, we deduce that the writing direction for syllables is from left to right.

Using a similar reasoning, we identify the four possible tone marks: ∅ (◌̇), *g* (◌̈), *s* (◌̄), *v* (◌). It is important to notice that the lack of a diacritic mark in Hmong is not equivalent to the lack of tone in the Latin transcription. If there is no diacritic above the syllable in Hmong, the Latin correspondent is the tone *v*,

while if there is no tone marking in the Latin transcription (the syllable ends in a vowel and not in *g*, *s*, or *v*), then the Hmong syllable will have a dot on top of the syllable. Therefore, we can consider the tone marking, to some extent, as an abugida system, in which the default tone is *v* and the change in tone or the lack thereof is marked by diacritics.

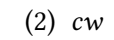
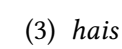
We are left to find out how the syllable is formed, i.e., which character represents the consonant and which character represents the vowel. Comparing examples 2 and 3, we notice that the first character of the first syllable is identical (except for the tone, which we already identified) and the two syllables (*hauv*, *raug*) have the same vowel (or sequence of vowels); thus the first symbol represents the vowel and the second one the consonant.

Another indication towards this order between the vowel and the consonant is that the tones, which are described as "specific ways of pronouncing the vowels", are generally marked above the first character, suggesting that the first character indeed refers to the vowel.

Based on this information, we can easily identify all the characters in this script. An important observation is the way the consonant *k* is marked. Based on examples 1 and 7, we notice that this consonant is not written, but rather treated as a default consonant. As a result, if in the Hmong script no consonant is written, we understand that the consonant is *k*. This is a particularly interesting writing system which cannot be easily fitted into any of the aforementioned categories of scripts, having characteristics of different types. On the one hand, we can consider it an alphabet since each consonant (consonant cluster) and vowel (or sequence of vowels) have individual characters (however, there are no characters for vowel-consonant combinations). On the other hand, the consonant *k* is not written, so it can be considered a default consonant and all the other symbols are used to change this consonant to another one (resulting in an abugida-like system in which there is a default consonant, rather than a vowel), just as in the case of tones (in which the abugida characteristics are much more obvious, having a default tone and a diacritic to remove it).

Based on all of the above, we can write the solution and solve the tasks.

Solution 2.1a

(1) *noog*
(2) *cw*
(3) *hais lus*
(4) *qhov muag kiv*
(5) 
(6) 
(7) 
(8) 

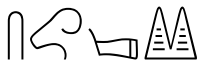
Rules:

- Syllables are written from left to right.
- Syllable structure: $\overset{3}{12}$
- 1 = vowel / vowel cluster (syllable nucleus) – *a* (), *e* (), *i* (), *o* (), *u* (), *w* (), *ai* (), *au* (), *aw* (), *ee* (), *oo* (), *ua* ()
- 2 = consonant (beginning of the syllable, onset) – *c* (), *h* (), *k* (∅), *l* (), *m* (), *n* (), *r* (), *y* (), *hl* (), *nts* (), *qh* ()
- 3 = tone – ∅ (), *g* (), *s* (), *v* ()

Problem 2.2

Luwian (Alfred Zhurinsky, MSK 1979)

The following are some inscriptions in the Luwian language. They correspond to some names of regions: *Khamatu*, *Palaa*, names of cities: *Kurkuma*, *Tuvanava* and names of kings: *Varpalava*, *Tarkumuva*.

1. 
2. 
3. 
4. 
5. 
6. 

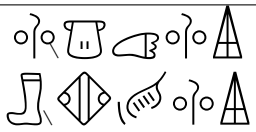
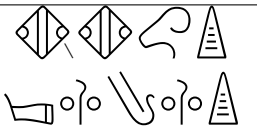
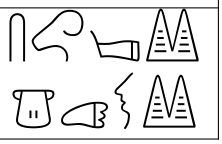
Problem 2.2a Determine the correct correspondences.

Problem 2.2b Write in Luwian:

7. *king Parta*
8. *king Artur*
9. *city Tartu*
10. *region Tuva*
11. *city Narva*

Solution

According to the introduction, the six inscriptions correspond to three categories of words: names of kings, cities, and regions. Moreover, we notice that the last character of each inscription does not appear anywhere else inside the inscription. Therefore, we can assume that these characters denote the idea of 'king', 'city', and 'region'; thus we can divide the six inscriptions into three categories based on the last character:

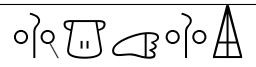
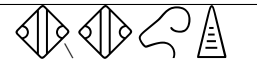
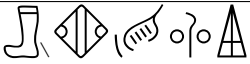
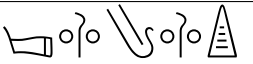
I	II	III
		

We can safely assume that this writing system is not alphabetic since each inscription has four or five characters, while their transcriptions have between five and nine characters. Moreover, it is unlikely that this system is an abugida or an abjad since there do not seem to be any diacritics appended (or similar characters). Therefore, it is most likely a syllabic system. To check this, we can try to divide the words in Latin transcription into syllables to check if the number of characters matches the number of syllables. (If you are unsure how to do this, see the discussion in Chapter 3.)

Kha-ma-tu and *Pa-la-a* each have three syllables. Therefore, we know for sure they correspond to group III (because it is the only group in which both inscriptions have four syllables – three for the actual name and one to show the category).

Kur-ku-ma and *Tu-va-na-va* have three and four syllables and the only category that matches it is II, so we deduce that this corresponds to the cities. Moreover, since the number of syllables is different, we can already make the correct correspondences: 2 – *Kurkuma* and 3 – *Tuvanava*.

The last group is that for kings and both words have indeed four syllables (*Var-pa-la-va* and *Tar-ku-mu-va*). We get:

Kings	Cities	Regions
	 *Kurkuma*	
	 *Tuvanava*	

Looking at the script representation of *Kurkuma*, we notice that the first two characters are very similar and they only differ by a little line placed on the bottom-right. Therefore, most likely, those two characters represent the syllables *kur* and *ku* and the little line on the bottom-right marks the consonant *r* at the end of the syllable. This is also confirmed by the fact that the names of the two kings both start with a syllable ending in *r* (*Va**r**-pa-la-va* and *Ta**r**-ku-mu-va*) and in both cases the first character has that line on the bottom right. Therefore, we deduce that the syllables are written from left to right and at the end we write the character showing the category they belong to (regions, cities or kings).

The rest of the correspondences are easily determined: the first character of *Tuvarnava* corresponds to the syllable *tu* and this syllable is also found in one of the regions' names (third character). The only region that contains the syllable *tu* is *Khamatu*, thus 6 – *Khamatu* and 4 – *Palaa*.

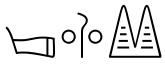
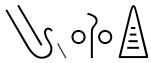
Among the two kings' names, one begins with *var* (which we already know from the city *Tu-**va**-na-**va***, adding the line representing final *r*), so we get the correspondences 1 – *Varpalava* and 5 – *Tarkumuva*.

Now we have all the information needed to solve the tasks.

Solution 2.2a

1. king *Varpalava*
2. city *Kurkuma*
3. city *Tuvanava*
4. region *Palaa*
5. king *Tarkumuva*
6. region *Khamatu*

Solution 2.2b

7. 
8. 
9. 
10. 
11. 

Rules: Syllabic system, left-to-right. In the end, there is a character showing the category to which the names belong:

 for kings,  for cities, and  for regions.

Each character represents a combination of a consonant and a vowel and adding a line on the bottom-right () marks the addition of the consonant *r* at the end of the syllable:

	a	*u*
∅		
k		
kh		
l		
m		

	a	*u*
n		
p		
t		
v		

Discussion (not part of the solution)

This is a syllabic type of system since each character represents a syllable. The only exception is the final character which is logographic (maybe even pictographic, considering that the symbol for 'region' is composed of two symbols for 'city', reflecting that a region is made up of more cities).

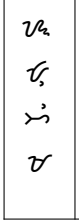
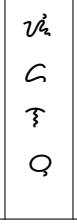
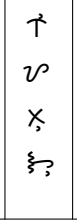
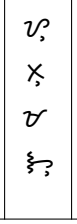
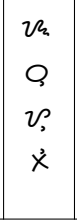
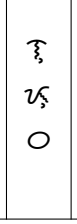
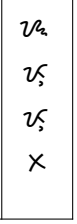
A possible issue when solving the problem is the syllabification of the word *Palaa*. Normally, we would have been tempted to syllabify it as *Pa-laa*, assuming the *aa* represents a long vowel (similar to problem 2.1). Nevertheless, if we had done that, the region Palaa would have had only three characters (two characters for the syllables *pa* and *laa* and one logographic character for region). Since we do not have any three-character inscriptions, we chose to syllabify the word as *Pa-la-a*.

Problem 2.3

Tagbanwa (Vlad A. Neacșu, JOL 2022)

Here are some words related to the mythology of the Tagbanwa people, written in the traditional script. They represent deities (*Mangindusa*, *Bugwasin*, *Tungkuyanin*, *Tumangkuyun*), names of spirits (*Kiyabusan*), rituals and words related to rituals (*Kapupusan*, *kadiyang*), as well as mythical places (*Balugu*). Their Latin transcriptions are given *in random order*:[2]

[2] *Note*: Due to the contest taking place online, a slightly different format of the problem was used.

1.	2.	3.	4.	5.	6.	7.	8.
							

A. *balugu*
B. *bugawasin*
C. *kadiyang*
D. *kapupusan*
E. *kiyabusan*
F. *mangindusa*
G. *tumangkuyun*
H. *tungkuyanin*

Problem 2.3a Determine the correct correspondences.

Problem 2.3b Write in Tagbanwa:

9. *mapintatan* (‘to charm’)
10. *panalangin* (‘prayer’)
11. *supisinti* (‘lifestyle’)

ng = ‘ng’ in ‘king’.

Solution

We begin again by attempting to deduce what type of writing system this could be. We know for sure that it is not an alphabet since we do not have any three-letter words (corresponding to examples 6 or 7). Moreover, it is extremely unlikely that this is a picto-, ideo- or logographic system since (1) the characters are rather simplistic and similar to one another (by adding semicircles ◌ᩮ or ◌ᩯ), which can be considered as diacritics), (2) we do not know the specific meaning of the words so we cannot correlate them with some pictographic or ideographic characters, and (3) using four characters to represent a single word would be rather many.

We are left with the possibilities of a syllabary, abjad, or abugida, in which each character would represent a syllable or a consonant-vowel pair (CV).

We can start by assuming it is a syllabic system and we syllabify the words. Based on this, the eight words are: *ba-lu-gu*, *bu-ga-wa-sin*, *ka-di-yang*, *ka-pu-pu-san*, *ki-ya-bu-san*, *ma-ngin-du-sa*, *tu-mang-ku-yun*, *tung-ku-ya-nin*.

Note

At first sight, it may seem more likely for an English-speaking person that the word *mangindusa* be syllabified as *mang-in-du-san* and not *ma-ngin-du-sal* since the sound *ng* is not found at the beginning of the syllable in English. Either way, the number of syllables does not change so we can create a frequency table with the number of characters and syllables.

# syllables	# words
3	2
4	6

If we first assume that each character represents a CV group, the resulting word-splitting would be *ba-lu-gu*, *bu-ga-wa-si-n*, *ka-di-ya-ng*, *ka-pu-pu-sa-n*, *ki-ya-bu-sa-n*, *ma-ngi-n-du-sa*, *tu-ma-ng-ku-yu-n*, *tu-ng-ku-ya-ni-n*, resulting in the following frequency table:

# CV groups	# words
3	1
4	1
5	4
6	2

Since we do not have any word represented by five or six characters, we deduce that this system cannot be based on CV groups.

The first observation is that in example 8 we have two consecutive identical characters (second and third). If we look at the given transcriptions, only one has two identical syllables, *ka-pu-pu-san*. Thus, we deduce 𑀯𑀸 = *pu*.

We must not forget that we have not yet confirmed the writing direction: it can be either top to bottom or bottom to top. Knowing that 8 = *kapupusan*, we deduce that the two other characters represent *ka* and *san* (not necessarily in this order). In order to find out the writing direction, we look at the last character (from top to bottom). This also appears as the last character in word 7. Thus, we have two possible cases:

Case 1. Writing from top to bottom ⟹ ᜐ = *san*. None of the three-syllable words contains the syllable *san*, therefore this case is impossible.

Case 2. Writing from bottom to top ⟹ ᜐ = *ka*. Looking at the three-syllable words, we notice that one of them starts with *ka* (*kadiyang*). Thus, we deduce that the writing system is from bottom to top and ᜏ᜴ = *san*.

In order to make the remaining correspondences, we begin by noticing that we have only one three-syllable word left, therefore 6 = *balugu* and we find out the characters for *ba*, *lu*, and *gu*.

Next, we notice that we have two words (3 and 4) which begin with the same character and among the words we have left, the only two that begin similarly are *tumangkuyun* and *tungkuyanin* (although they do not begin with exactly the same syllable – one begins with *tu* and the other with *tung*). Therefore, we deduce that {3, 4} = {G, H}. Moreover, we see that both words also have another syllable in common, *ku*, and it has different positions: third in *tumangkuyun* and second in *tungkuyanin*. In the Tagbanwa script, there is a character that appears in both words (except for the first character), ᜓ. Thus, 3 = *tungkuyanin* (since *ku* is the second syllable and ᜓ is the second character) and 4 = *tumangkuyun*. Moreover, it seems that syllables *tu* and *tung* are represented by the same character. The syllable *ya*, which appears in word 3, also appears in word 5, and the only word that also contains the syllable *ya* is *kiyabusan*. Thus, 5 = E.

Based on word 5, we deduce the character for *bu*, which also appears in word 2, and the only word that contains this syllable is *buguwasin*. Thus, 2 = B. The only word left is *mangindusa*, so 1 = F.

We noticed previously that the same character represents both syllables *tu* and *tung*. Comparing the word pairs 1-3 (syllables *sa-san*), 3-7 (syllables *ya-yang*) and 1-4 (syllables *ma-mang*), we infer that if the syllable ends in a consonant (which can only be *n* or *ng* based on the data given), this is not transcribed. Therefore, word 4, for example, is read *tu-ma-ku-yu*, although it represents the word *tu-mangkuyun*. In reality, the reader is able to fill in the missing consonants when reading the word, even though they are not written.

In order to find out the way the vowel (or consonant) is marked, we can make a table for all characters in order to check for any patterns. We exclude the consonant that may occur at the end of the syllable based on what was said above.

C/V	*a*	*i*	*u*
b			
d			
g			
k			
l			
m			
n			

C/V	*a*	*i*	*u*
ng[a]			
p			
s			
t			
w			
y			

[a] Another possible explanation is that the character represents the lack of an initial consonant, in which case the word *mangindusa* would be syllabified as *mang.in.du.sa*. Both explanations are correct and yield the same solution. We chose the explanation in which the character represents the consonant *ng* since this is the real explanation and, furthermore, the syllabification respects the rule $VCV \rightarrow V.CV$.

We can easily observe that the basic form of the consonant is that with vowel *a*, while the other vowels are formed by adding the diacritic in different positions around the base character (above for vowel *i* and on bottom-right for the vowel *u*). Therefore, this system is an abugida, save for the fact that there is no diacritic for vowel deletion, but rather if a consonant appears alone (hence at the end of the syllable) it is not written. Based on the table above and our observations, we can easily deduce all the other characters and solve the tasks.

Solution 2.3a

1. F
2. B
3. H
4. G
5. E
6. A
7. C
8. D

Solution 2.3b

9.	10.	11.
		
		
		
		

2.10 Sign language

Sign languages are used by Deaf people and are based on movements and gestures rather than spoken sounds. Their grammar is entirely different from that of spoken languages, and they generally bear no relation to the languages of hearing communities: for instance, the United States and the United Kingdom share a majority spoken language (English), but their most widely used sign languages (American Sign Language/ASL, British Sign Language/BSL) are completely distinct and not mutually intelligible.

In addition to their own grammar, many sign languages have a system of *finger spelling* to represent the written forms of other languages. Thus, some gestures represent certain ideas or words (sometimes in an iconic way, thus having a *semantic* purpose), while others do not have any meaning and are representational of letters in written language, allowing the possibility to spell out items such as novel words or names.

2.11 Braille alphabet

Braille is a writing system developed for blind people in which each character is represented by raised dots which the reader can feel with their finger tips. Each character is represented by a 3×2 grid

where each circle can be either empty (marking the lack of a raised dot) or full (marking a raised dot). This is just a way of representation in order to make it easier to identify. In reality, only the raised dots (full circles) are used. Just as in the case of sign languages, each country or language may develop their own Braille system, not necessarily mutually intelligible.

Problem 2.4

Japanese Braille (Patrick Littell, NACLO 2009)

Below are some Japanese words written in the tenji system (a Japanese version of the Braille system), together with their Latin transcriptions *in random order*:

a. 

b. 

c. 

d. 

e. 

f. 

atari, *haiku*, *katana*, *kimono*, *koi*, *sake*

Problem 2.4a Determine the correct correspondences, knowing that:

karaoke = 

Problem 2.4b Write in Latin script:

 and  .

Problem 2.4c Write in tenji: *samurai* and *miso*.

Solution

We start, again, by determining the type of writing system. We already know this is not alphabetic (since we have words represented by two tenji characters and we have no two-letter words), and it is obvious we cannot talk about a picto-/ideo- or logographic system. Therefore, it most likely is a syllabic system in which each tenji character represents a syllable (or a CV group – in this case, the two are equivalent since each syllable of the given words has the structure (C)V). We infer that the four characters in *karaoke* represent *ka*, *ra*, *o*, and *ke*, but we still do not know in which order.

We notice that *ka* appears one more time as a first syllable in the word *katana*, while *ke* appears as a last syllable in the word *sake*. Since the first character of *karaoke* is the same as the first character of c., we have two possibilities:

1. this character represents *ke* and c. = *sake*, which is impossible since c. has three characters, not two;

2. this character is *ka*, the writing direction is left-to-right, and c. = *katana*. Moreover, b. = *sake* and we can deduce the characters for *ta*, *na*, *sa*.

Based on this information, we can easily make the rest of the correspondences as follows: *ta* appears in only one other word (*atari*), so f. = *atari* and we deduce the characters for *a* and *ri*. We are left with three words to match (*koi*, *haiku*, *kimono*). Out of them, only one has two syllables and, therefore e. = *koi* and we deduce the characters for *ko* and *i*. Knowing the character for *i*, which must also

appear in the word *haiku*, we can make the last two correspondences: a. = *haiku*, d. = *kimono*.

We again make a table to check whether there are any patterns based on the vowel or consonant in the syllable structure.

	a	*e*	*i*	*o*	*u*
∅					
h					
k					
m					
n					
r					
s					
t					

In this case, we notice that each character represents a combination of a vowel and a consonant, the vowel being marked on the first three dots and the consonant on the last three dots. We can better illustrate this as follows:

	∅	*a*	*e*	*i*	*o*	*u*
∅						
h						
k						
m						
n						
r						
s						
t						

where the first row and column (highlighted) represent the individual characters and in order to obtain a CV syllable, we simply overlap the two components. Based on these rules, we can solve all the tasks.

Solution 2.4a

a. *haiku*
b. *sake*
c. *katana*
d. *kimono*
e. *koi*
f. *atari*

Solution 2.4b

⠓⠁ ⠗⠁ ⠞⠑ = *karate*

⠁⠝ ⠊⠍ ⠑ = *anime*

Solution 2.4c

samurai = ⠎⠁ ⠍⠥ ⠗⠁ ⠊

miso = ⠍⠊ ⠎⠕

2.12 Practice problems

Problem 2.5

Armenian (Dragomir R. Radev, NACLO 2010)

On her visit to Armenia, Millie has gotten lost in Yerevan, the nation's capital. She is now at the metro station named *Shengavit*, but her friends are waiting for her at the station named *Barekamutyun*. Other names of stations that can be found on the map below are: *Gortsaranayin*, *Zoravar Andranik*, *Charbakh* and *Garegin Njdehi Hraparak*.

Problem 2.5a Assuming Millie takes a train in the correct direction, which will be the first stop after *Shengavit*? Write the name transcribed into English.

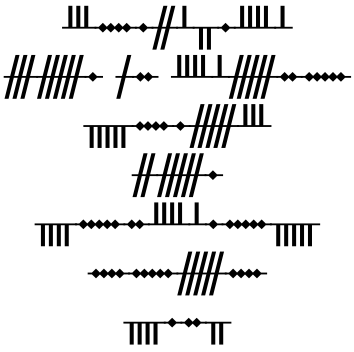
Problem 2.5b After boarding at *Shengavit*, how many stops will it take Millie to get to *Barekamutyun*? Don't include *Shengavit* itself in the number of stops.

Problem 2.5c What is the name (transcribed into English) of the end station on the short, five-station line that is currently under construction?

Problem 2.6

Ogham (Babette Verhoeven-Newsome, UKLO 2021)

Here are some Irish words written in the Ogham alphabet and their transcriptions in the Latin alphabet (together with their English translations) *in random order*:



A. *grá* ('love')
B. *teaghlach* ('family')
C. *Éire* ('Ireland')
D. *neart* ('strength')
E. *saol* ('life')
F. *síocháin* ('peace')
G. *grá mo chroi* ('love of my heart')

Problem 2.6a Determine the correct correspondences.

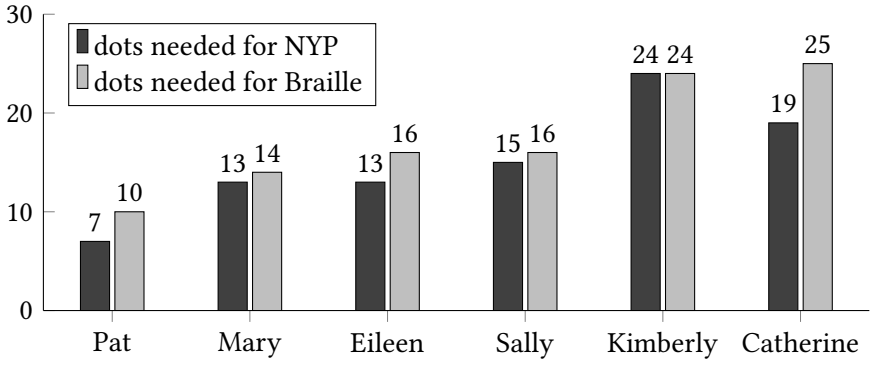
Problem 2.6b Below is the Ogham spelling of the Irish for 'I love you'. Write it down in Latin alphabet transliteration. You can ignore accents for this task.

Problem 2.7

New York Point (Patrick Littell, UKLO 2011)

Before the Braille tactile writing system was well established in the United States, the New York Point system (NYP) was widely used in American blind education. NYP was developed in the 1860s by William Bell Walt for the New York Institute for the Blind and was intended to fix the shortcomings he perceived in the French and English Braille standards. The next six decades in blind education became known as the "War of the Dots", as bitter feuds developed between proponents of this homegrown system and more international Braille-based systems. NYP finally met its end after a series of public hearings convinced educational authorities that there should be a single standard for the entire English-speaking world.

Experts from both sides weighed in on the systems' merits. The proponents of NYP argued that allowing letters to vary in size (from a 2x1 grid to a 2x4 grid, rather than a fixed 3x2 grid) allowed the most frequent letters to use fewer columns, resulting in space (and cost!) savings when publishing texts for the blind. For example, the number of dots needed to write the following names in each system:



They also pointed out that NYP had a distinct series of capital letters, whereas Braille only had a "capital" punctuation mark.

On the Braille side, experts such as Helen Keller wrote that the NYP capitalization system was unintuitive and confusing ("I have often mistaken *D* for *j*, *I* for *b* and *Y* for double *o* in signatures, and I waste time looking at initial letters over and over again"), and that using Braille allowed her to correspond with blind people from all over the world.

The following 12 words in NYP represent, *in random order*, the names: *Ashley*, *Barb*, *Carl*, *Dave*, *Elena*, *Fred*, *Gerald*, *Heather*, *Ivan*, *Jack*, *Kathy*, *Lisa*.

1. 
2. 
3. 
4.
5.
6.
7. 
8. 
9. 
10. 
11. 
12. 

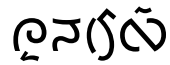
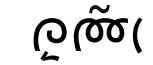
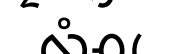
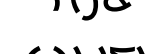
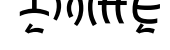
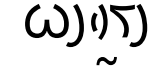
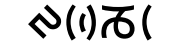
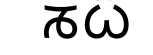
Problem 2.7a Determine the correct correspondences.

Problem 2.7b Write in NYP: *Billy*, *Ethan*, *Iggie*, *Orson*, *Sasha*, *Tim*.

Problem 2.8

Lepcha (Monojit Choudhury, PLO 2015)

Sikkim state in India has 11 official languages. Amongst these, ten are given below (the eleventh one is English) written in the Lepcha script, as well as in the Latin script:

nepaalii		*newaar*	
lepchaa		*raai*	
sikkim		*gurung*	
taamaang		*magar*	
liimbu		*sunwaar*	

Problem 2.8a One of the languages above has, in reality, two names, and its name written in the Latin script does not match the name written in the Lepcha script. Its name, transliterated from Lepcha, is *drenzoongkee*. Which language is this?

Problem 2.8b The Lepcha speakers, who call themselves *roong haagiit* (or, in the Lepcha script, ᰛᰩᰵ ᰝᰦᰃᰧᰳ) are composed of four main distinct communities: ᰛᰬᰵᰈᰩᰵᰕᰪ, ᰊᰮᰠᰵᰕᰪ, ᰣᰧᰜᰯᰕᰪ, and ᰎᰥᰨᰕᰪ. Transcribe these four community names into the Latin script.

Problem 2.8c Sikkim boasts the *Kaangchenzoonggaa*, the third highest peak in the world, which, in Tibetan, means ‘the five treasures of the high snow’. Transcribe the name of this peak in Lepcha.

> Vowel doubling denotes length. *ch* = ‘ch’ in ‘chop’; *ng* = ‘ng’ in ‘king’; *w* = ‘v’ in ‘van’.

Problem 2.9

Arabic – Hebrew (Gábor Parti, HKLO 2020)

Arabic and Hebrew are today two different, mutually unintelligible languages. However, they share both grammatical similarities and several lexical correspondences. Besides loanwords (mostly from Arabic to Hebrew) from different historical periods, scholars have identified over a thousand *cognates* (words with a common etymological origin from Proto-Semitic, which is the partially reconstructed common ancestor language spoken around 6,000 years ago).

The two lists below show pairs of cognates, but the pairs are mixed: Arabic on the left, and Hebrew on the right. The first match (1–A) is given for you.

شمس	1.	⟷	A.	שמש
غضب	2.		B.	כלב
ولد	3.		C.	מלך
أرض	4.		D.	עבד
بطن	5.		E.	ארץ
كلب	6.		F.	עצב
حبل	7.		G.	קרן

ملك	8.	H.	ילד
قرن	9.	I.	חבל
عبد	10.	J.	בטן

Thanks to regular and consistent sound changes, we have some easily identifiable patterns. Take for example the following eight words from the list above (transcribed in the Latin script):

Arabic	Hebrew	Translation	Arabic	Hebrew	Translation
kalb	*kelev*	‘dog’	*shams*	*shemesh*	‘sun’
malik	*melekh*	‘king’	*qarn*	*qeren*	‘horn’
ʾarḍ	*ʾerets*	‘land, earth’	*ghaḍab*	*ʿetsev*	‘anger, sadness’
ʿabd	*ʿeved*	‘slave’	*walad*	*yeled*	‘child’

Problem 2.9a What is the transliteration (in both Arabic and Hebrew) of the two word pairs from the lists 1-10 and A-J not included in the table showing transliterations?

Problem 2.9b The chart below shows a few letters in both scripts with their transliterations. Note that some letters may have different forms depending on the context they appear in.

Hebrew:

ר	צ / ץ	ע	נ / ן	כ / ך	י
(1)	(2)	(3)	(4)	*k*	(5)

ט	ח	ד	ב	בּ	א
t	*ḥ*	(6)	(7)	*b*	ʾ (alef)

Arabic:

ط / ـطـ / ـط	ر / ـر	د / ـد	ح / ـحـ / ـح	ب / ـبـ / ـب
ṭ	(8)	(9)	*ḥ*	(10)

أ	و / ـو	ن / ـنـ / ـن	ل / ـلـ / ـل	ك / ـكـ / ـك	ض
ʾ (alif)	(11)	(12)	(13)	(14)	(15)

Fill in the gaps (1–15).

Problem 2.9c Pair the matching cognates 2-10 and B-J from the first list (words transcribed in Arabic and Hebrew).

Problem 2.9d If 'thousand' in Arabic is ألف and the final letter *f* in Hebrew is ף, what is the transliteration of אלף – also meaning 'thousand' in Hebrew?

The apostrophe (') in both languages represents a glottal stop /ʔ/. In Arabic it is written by a *hamza*, which often 'sits' on top of an *alif*; in Hebrew, it is represented by an *alef* and often omitted in pronunciation. Here, you should just consider it a consonant and treat it like you would treat any other consonant!

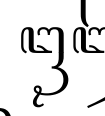
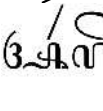
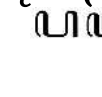
The symbol ʿ represents a voiced pharyngeal fricative /ʕ/, which is an odd sound made by contracting the muscles in the throat. It gives Arabic its unique flavour we can easily hear. In Modern Hebrew, it is silent and almost only ever appears in writing. Just think of it as an ordinary consonant!

Problem 2.10

Javanese (Tae Hun Lee, NACLO 2016)

Here are some Javanese words in the Javanese script, Latin script, and their English translations:

1.  *penyakit* 'disease'
2.  *Inggris* 'England'
3.  *traktor* 'tractor'

4.		*panyumbang*	‘donor’
5.		*rembulan*	‘moon’
6.		*tansah*	‘always’
7.		*Amérika*	‘America’
8.		*ngrebut*	‘to grab’
9.		*ibukota*	‘capital’
10.		*Argentina*	‘Argentina’
11.		*srengéngé*	‘sun’
12.		*palsu*	‘false’
13.		*rerenggan*	‘decoration’
14.		*angsal*	‘to acquire’
15.		*inggih*	‘yes’
16.		__(1)__	‘often’
17.		__(2)__	‘letter, script’

18.		__(3)__	‘to unload’
19.		__(4)__	‘to examine’
20.		__(5)__	‘to cancel’
21.	__(6)__	*nyolong*	‘to steal’
22.	__(7)__	*sepalih*	‘half’
23.	__(8)__	*trengginas*	‘lively’
24.	__(9)__	*Antartika*	‘Antarctica’
25.	__(10)__	*Istanbul*	‘Istanbul’

Problem 2.10a Fill in the blanks.

ny and *ng* are consonants; *é* is a vowel.

Problem 2.11

Thai (Sergey Dmitrenko, MSK 2001)

Here are some Thai words written in the Thai script and their Latin transcriptions (together with their English translations) given *in random order*:

1. หวาย	4. ว่าย	7. กาว	10. วัย	13. หลัง
2. ท่าน	5. กาย	8. ถาด	11. หลาว	14. ลาว
3. วาย	6. ทัน	9. ถาก	12. ทาน	

A.	*tʰà:k*	‘to clear (a field)’	H.	*ka:w*	‘glue’
B.	*vâ:y*	‘to swim’	I.	*la:w*	‘Laotian’
C.	*lă:w*	‘javelin’	J.	*tʰâ:n*	‘you (formal)’
D.	*va:y*	‘to end’	K.	*tʰa:n*	‘charity’
E.	*vă:y*	‘rattan’	L.	*vay*	‘age’
F.	*tʰan*	‘to have time’	M.	*tʰà:t*	‘tray’
G.	*lăŋ*	‘back’	N.	*ka:y*	‘body’

Problem 2.11a Determine the correct correspondences.

Problem 2.11b Write in Thai:

15.	*vă:n*	‘sweet’	17.	*tʰàk*	‘to knit’
16.	*ya:ŋ*	‘rubber’	18.	*vâ:w*	‘kite’

A colon (*:*) after a vowel indicates length. The marks above vowels denote tones. This problem features four tones: medium (*a*), rising (*ǎ*), falling (*â*), low (*à*).

tʰ and *ŋ* are consonants.

2.13 Solutions of practice problems

Solution for practice problem 2.5. Armenian

Solution 2.5a *Gortsaranayin*

Solution 2.5b 7

Solution 2.5c *Avtogortsaran* (the character resembling the letter S can be inferred to mean *t* from the title of the map, where the last word is *metropoliten*).

Solution for practice problem 2.6. Ogham

Solution 2.6a 1. B 2. G 3. D 4. A 5. F 6. C 7. E

Solution 2.6b *ta me i ngra leat* (in reality it is *Tá mé i ngrá leat*).

Rules: We can classify the characters depending on the number of dots or lines as well as their position or direction (vertical or diagonal, above or below the horizontal line):

	1 line	2 lines	3 lines	4 lines	5 lines
vertical, below		*l*		*s*	*n*
vertical, above	*h*		*t*	*c*	
diagonal	*m*	*g*			*r*
dots	*a*	*o*		*e*	*i*

The accents are not marked (*á* = *a*).

Solution for practice problem 2.7. New York Point

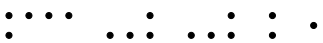
Solution 2.7a

1. *Kathy*
2. *Elena*
3. *Ivan*
4. *Carl*
5. *Jack*
6. *Gerald*
7. *Lisa*
8. *Fred*
9. *Heather*
10. *Barb*
11. *Ashley*
12. *Dave*

Solution 2.7b

Billy –


Ethan –


Iggie –


Orson –


Sasha –


Tim –


Rules: Forming the capital letter: all capital letters are four columns long and are formed by appending dots to the lowercase letter until it is four columns long, according to the following pattern:

- If the last column of the lowercase letter has a dot in the upper row, add the extra dots on the lower row.
- If the last column of the lowercase letter has a dot in the lower row or both dots, add the extra dots on the upper row.

Discussion (not part of the solution)

Figuring out the character for *o*: in the introduction it is mentioned that *Y* (⠊⠊) can be mistaken for a double *o*, and *Y* is formed by the same pattern repeated twice, so that pattern must represent the letter *o* (⠊).

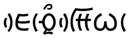
Figuring out the characters for *t* and *m*: based on the graph given in the introduction, we can deduce that *t* must contain a single dot (*Pat* has seven dots, *a* is two dots and *P*, since it is uppercase, must have at least four dots). Since we already know that *e* is ⠁, *t* can only be ⠂.

From the graph, we deduce that *M* has five dots and *m* three, therefore the three dots must be distributed on only two columns (since two columns are needed for the extra dots to form the uppercase). There are four options (⠚, ⠋, ⠓, ⠙). Since three of the characters are already used (⠚ = *r*, ⠓ = *l*, ⠙ = *d*), there is only one option left to place the dots (⠋) such as the final pattern does not coincide with other letters.

Solution for practice problem 2.8. Lepcha

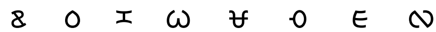
Solution 2.8a Sikkim language

Solution 2.8b *renzoongmu* *taamsaangmu* *hilaammu* *proomu*

Solution 2.8c 

Rules:

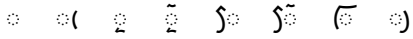
- Abugida script, left-to-right.
- Consonants at the beginning of the syllable:



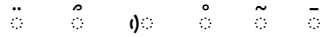
∅ *b* *d* *g* *h* *ch* *k* *l*

m *n* *p* *r* *s* *t* *w* *z*

- Vowels are marked by diacritics. The default vowel is *a*.



- Consonants at the end of the syllable (codas) are also marked by diacritics:



- If the syllable onset has the structure *Cr*, that *r* is marked as ◌ྲ. Compare:

Solution for practice problem 2.9. Arabic and Hebrew

Solution 2.9a *ḥabl* ⟶ *ḥevel*
baṭn ⟶ *beten*

Note

In Hebrew all vowels shown are *e*. In Arabic it is impossible to deduce the vowels based on the data, so alternative versions are also accepted (such as *ḥabal* or *ḥabil*).

Solution 2.9b

(1) *r*	(6) *d*	(11) *w*
(2) *ts*	(7) *v*	(12) *n*
(3) ʕ (ayin)	(8) *r*	(13) *l*
(4) *n*	(9) *d*	(14) *k*
(5) *y*	(10) *b*	(15) *ḍ*

Solution 2.9c

1. A	3. H	5. J	7. I	9. G
2. F	4. E	6. B	8. C	10. D

Solution 2.9d *'elef*

Solution for practice problem 2.10. Javanese

Solution 2.10a

(1) *kerep*

(2) *aksara*

(3) *mbongkar*

(4) *mrikso*

(5) *murungaké*

(6) 

(7) 

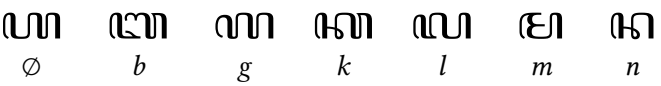
(8) 

(9) 

(10) 

Rules:

- Abugida script, left-to-right, with the default vowel *a*.
- Syllables have the structure $C_1(C_2)V(C_3)$, taking into account the following syllabification rules:
  - $...VC^aC^bV... \rightarrow ...VC_3^a - C_1^bV...$ if C^a is *ng*, *h*, or *r*;
  - $...VC^aC^bV... \rightarrow ...V - C_1^aC_1^bV...$ otherwise;
- C_1



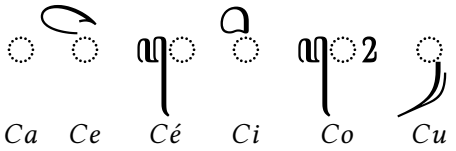
∅ *b* *g* *k* *l* *m* *n*

ng *ny* *p* *r* *s* *t*

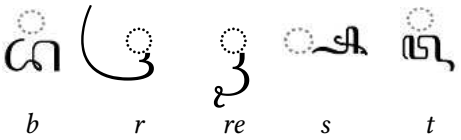
Note

The combination *re* has its own character: 

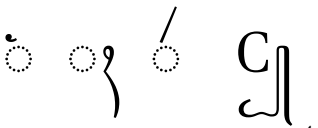
- The vowel change is shown using diacritics:



- C_2:[3]

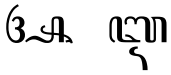


- C_3:



[4]

- Special characters for capital letters:



Solution for practice problem 2.11. Thai

Solution 2.11a 1. E 2. J 3. D 4. B 5. N 6. F 7. H 8. M 9. A 10. L 11. C 12. K 13. G 14. I

Solution 2.11b 15. หวาน 16. ยาง 17. ถัก 18. ว่าว

[3] In fact, this special form of the consonant marks the fact that the vowel of the preceding consonant is deleted.

[4] This is basically the vowel-removing mark, showing that the previous consonant does not have a vowel.

Rules:

- Writing direction from left to right.
- The character ว represents two consonants: *v*, if in the beginning of the word; and *w*, if at the end of the word.
- The sound t^h corresponds to two Thai characters: ท and ธ. The latter is used to mark the low tone.
- Vowels: *a* = ◌ั (placed on the first consonant), *a:* = า
- Tone:
  - Medium – default tone (unmarked);
  - Rising – ห placed at the beginning of the word;
  - Falling – ◌่ placed on the first consonant;
  - Low – appears only in words starting with t^h. In this case, the tone is marked by using the character ธ to mark the consonant t^h (rather than ท).

Discussion (not part of the solution)

The way that the Thai writing system represents tone is very complex. To know which tone to pronounce the syllable with, we need to know the initial consonant and the type of syllable. Initial consonants are split into three classes: low, medium (mid), and high, while the syllables are either "live" (ending in *n*, *m*, *ŋ*, *w*, *y*) or "dead" (ending in *p*, *t*, *k*). Moreover, there are some diacritics which may be added to further modify the tone. The following table details the rules relevant to the data in the problem:

		Type of first consonant		
Diacritics	Syllable type	*low*	*mid*	*high*
no mark	*live*	medium (*a*)	medium (*a*)	rising (*ǎ*)
no mark	*dead*	*	low (*à*)	low (*à*)
◌่	*live*/*dead*	falling (*â*)	low (*à*)	low (*à*)

* The choice of tone further depends on the vowel length: if the vowel is long, the syllable will have falling (*â*) tone; if the vowel is short, the syllable will have high (*á*) tone.

Most words in the problem start with *low* consonants (ล, ท, ว, ล). There is a single *mid* consonant at the beginning of the words (*k* = ก), but, since this one only appears in *live* syllables, its tone is identical to that of a syllable starting with a *low* consonant.

Most Thai consonants have two forms: a low form and a high form (in the problem we have the consonant t^h, which, in its low form, is written as ท, while in its high form becomes ถ). If some consonants do not have two different characters to differentiate between low and high, a low consonant can become high by adding the character ห before it (the character represents the high consonant *h*, but it is not pronounced).

Based on this information, the rules in our problem can be explained as follows:

- the default tone is the medium one because most of the words contain live syllables and start with a low or mid consonant.
- the falling tone is shown by the diacritic ◌่ and by the fact that these syllables all start with a low consonant (otherwise they would have a low tone).
- the character ถ at the beginning of the word changes the tone to low because the syllable is dead (it is the only occurrence of a dead syllables in the problem) and the first consonant is high.
- the rising tone is shown by adding the character ห at the beginning of the word since it will transform the initial consonant from low to high.

Further reading

Coulmas, Florian. 1999. *The Blackwell encyclopedia of writing systems.* New York: Wiley-Blackwell.

Coulmas, Florian. 2003. *Writing systems: An introduction.* Cambridge: Cambridge University Press.

Cruttenden, Alan. 2021. *Writing systems and phonetics.* London: Routledge.

Daniels, Peter T. & William Bright. 1996. *The world's writing systems*. Oxford: Oxford University Press.
Haarmann, Harald. 2004. *Geschichte der Schrift*. München: C. H. Beck.
Robinson, Andrew. 1999. *The story of writing*. New York: Thames & Hudson.

3 Phonetics

3.1 Introduction

The field of phonetics is concerned with how speakers of different languages make the sounds of their speech (*production*, or *articulation*), and also with how they are heard by listeners (*perception*; this is sometimes known as *auditory phonetics*). Many linguistics problems will involve phenomena of *phonology* – for now, we can define these as changes of sounds that depend on the structure of words. We will see many examples of these in Chapter 4. Before we do so, it is useful to look at the different types of sound, and the ways in which they are classified, in order to understand these types of change.

In this chapter (and indeed throughout the book), we will use a notation system called the International Phonetic Alphabet (IPA). This is an alphabetic writing system, formed primarily of Latin characters and symbols derived therefrom and developed in order to make it possible to transcribe any word, from any spoken language, using this single set of conventions. Thus, each sound will have its own notation.

Writing down the sounds of a language is referred to as (phonetic) *transcription*. To distinguish words that are transcribed from how they are represented in spelling, we often use square brackets (as in [tʰɹənsˈkɹɪpʃn̩]) or slashes (as in /trænsˈkrɪpʃən/), both corresponding to the spelling *transcription*. The precise difference between the two kinds of transcription shown here is not that important; very roughly, slashes are generally used for a kind of notation that only aims to differentiate the sounds found in that specific language (and thus omit some detail), while square brackets tend to contain quite *narrow* transcriptions, which contain a lot of specific information. For example, the sound usually written as /r/ can, in fact, be pronounced differently in different languages, and those differences will be reflected in narrower transcriptions: in English, it can be represented as [ɹ] (for instance, in many dialects of England), [ɻ] (in North America or Northern Ireland), or [ɾ] (in Scotland); in Romanian, it is generally [r], and in French or German it is usually [ʁ]. Narrow transcriptions aim to reflect this detail that depends on each language, but as often as not we only need to worry about the broad outlines of the system and will just use /r/ for all of these sounds,

since most languages have only one type of *r*. This distinction rarely matters in practice for solving linguistics problems.

3.2 Classification of sounds

Probably the most fundamental sound distinction that we need to keep in mind is the difference between vowels and consonants. The basic characteristic of vowels is that when we pronounce them, the air coming up from the lungs does not meet any obstacle (or, as we sometimes say, there is no constriction, either total or partial) in the vocal tract. Conversely, consonants are formed with some kind of obstacle, or *stricture*.

3.2.1 Consonants

We can describe most consonants with reference to three main characteristics, which are referred to as *place of articulation*, *manner of articulation*, and *voicing*. ("Articulation" is just what we call the movements involved when we make the different sounds of speech.)

3.2.1.1 Place of articulation

This describes where in the vocal tract the main constriction is located, i.e., what organs are primarily involved in the articulation of each consonant. Going from front (the lips) to back (the back of the throat), we can identify the following places of articulation for consonants:[1]

- Labial, in which the lips are involved. In English, these are:
  - Bilabial consonants, produced using both lips (*bi-* = 'two', *labium* = 'lip' ⟹ 'two lips'): [b], [p], [m];
  - Labiodental consonants (*labium*, *dental* = 'tooth' ⟹ the lower lip and the upper teeth): [f], [v];
- Coronal, pronounced by using the tip of the tongue:
  - (Inter)dental consonants (*inter* = 'between', *dental* ⟹ the tongue is placed between, or just on, the teeth) – [θ] (*th* in *thin*), [ð] (*th* in *that*);

[1] A diagram of the vocal tract can be found on page 7 of the International Phonetic Association's *Handbook of the International Phonetic Association*, published by Cambridge University Press in 1999.

  - Alveolar consonants (the tip of the tongue is placed on or near the alveolar ridge, the little protrusion just behind the upper teeth) – in English, [t], [d], [s], [z], [n], [l] are all usually alveolar, as is [r] in some languages other than English;[2]
  - Post-alveolar consonants (*post* = 'after' ⟹ the tongue is placed further behind than the alveolar ridge) – [ʃ] (*sh* in *shop*), [ʒ] (*s* in *vision*), [tʃ] (*ch* in *church*), [dʒ] (*j* in *jam*); in many varieties of English, [ɹ] is also post-alveolar;

- Dorsal consonants, articulated primarily using the blade of the tongue:
  - Palatal consonants (the tongue is placed against the hard palate of the oral cavity): one example from English is [j] (note that in the IPA this symbol refers to the *y* of *yellow*, not the *j* of *jam*!). Other palatal consonants are [c] (similar to *cc* in *accute*, but not as the *c* in *cat*), [ɟ] (similar to *g* in *geese*), [ɲ] (*ni* in *onion*), [ʎ] (as in Italian *figlio*, or in English *mi**ll**i**o**n*);
  - Velar consonants (the back of the tongue is placed against the soft palate, also known as the velum) – [k] (*c* in *cat*), [g], [ŋ] (*ng* in *king*), [x] (as in Scottish *loch* or German *Bach*);

- Glottal consonants, which generally lack any constriction in the mouth, but some noise is created as air passes through the vocal folds (the glottis is the gap between the two vocal folds) – [h] and the glottal stop, written [ʔ] and heard in *uh-oh*, or, for some speakers, in the middle of words like *butter*. (Pronouncing *butter* with a glottal stop is sometimes called "dropping one's t's", but that is clearly wrong: the *t* isn't dropped, it's just pronounced differently!)

These are the main places of articulation we encounter, but in the world's languages, there are quite a few more (*alveolo-palatal*, *retroflex*, *uvular*, *pharyngeal*, *epiglottal*, and so on). In practice, though, if these kinds of sounds are featured in a linguistics problem, they will likely be described in the note at the end: yet another reason to read these notes carefully. If the consonants are described in apparently unnecessary detail, this might be a clue that the information is important!

[2] In most linguistics problems, the sound [r] is treated as an alveolar sound, unless the footnotes state differently.

3.2.1.2 Manner of articulation

Manner refers to the precise way in which the articulators move and the stricture is formed. Here is how the consonants can be classified in terms of manner:

- Plosive consonants, also known as occlusives or stops, formed with a total closure of the vocal tract, followed by a sudden release, similar to an explosion. Examples of these consonants are: [p], [b], [t], [d], [k], [g];
- Fricative consonants: there is a partial stricture of the vocal tract that leaves a very narrow opening. As a result, the constant airflow through the opening produces turbulent noise. Examples include [f], [v], [s], [z], [θ], [ð] [x], [ɣ];
- Affricate consonants are a combination of a plosive and a fricative: at the beginning of their articulation, the closure of the vocal tract is total, but the release is gradual, causing a fricative-like flow of air. In the IPA, they are notated by the symbol for the stop followed by the corresponding fricative, sometimes joined by an arc: [t͡s], [d͡z], [t͡ʃ], [d͡ʒ].[3]
- Nasal consonants: in this case, the oral tract is closed (as it is for stops), but the airflow is released through the nose. Most IPA symbols for nasals resemble the letter *n*, e.g.: [n], [m], [ŋ], [ɳ];
- Liquid consonants. This is an umbrella term for many different manners of articulation, but as far as the linguistics problems are concerned, we need not go into any more details. This category includes the consonants [l] and [r], as well as, similar to the nasals, most consonants whose IPA symbols resemble *l* and *r* (e.g., [ɬ], [ɭ], [ʎ], [ɺ], [ɽ], [ʀ]).
- Glides (also known as semivowels). There is still partial closure of the vocal tract, but it is not as narrow as for fricatives: these consonants are in between fricative consonants and vowels. Commonly encountered glides are [w] (*w* in *week*) and [j] (*y* in *year*).

[3]The addition of the arc is necessary because some languages distinguish between the affricate and the corresponding stop-fricative sequence (as in Polish *trzy* [tʃɨ] 'three' but *czy* [t͡ʃɨ] 'whether'). Nevertheless, this rarely happens in linguistics problems, and you should not worry about the presence or absence of the arc. Note also the letter *c*, which in many languages' orthographies is used for the affricate [t͡s], but has a different meaning in the IPA!

Technically, [w] is what is known as a "labial-velar" glide because it is articulated using both the back of the tongue and the lips. In real problems that you might encounter, it might behave both as a velar and as a labial, so we have put it in both columns in Table 3.1.

There are also umbrella terms that can be useful with respect to the manner of articulation. The cover term for plosives, fricatives and affricates is OBSTRUENT, while all the other consonants (nasals, liquids, and glides) can be together referred to as SONORANTS.

Moreover, there is a useful term that combines a manner of articulation with a place of articulation: alveolar and post-alveolar fricatives and affricates can be called SIBILANTS.

3.2.1.3 Voicing

Voicing refers to the involvement of the vocal folds in the articulation of the consonants. A common distinction is between VOICELESS (vocal folds are not involved) and VOICED (vocal folds are vibrating) sounds. Vowels are almost always voiced, but this parameter can make a difference for consonants. Specifically:

- sonorants (i.e., nasals, liquids, and glides) are almost always voiced;
- stops, fricatives, and affricates can be either voiceless ([p], [s], [ts], [k]) or voiced ([b], [z], [dz], [g]).

Considering all of the above, we can lay out the consonants in a table summarising all these properties (Table 3.1). Across columns, we will write the places of articulation (left to right from anterior to posterior), while rows show the different manners of articulation. To show voicing, we can use alignment within the table cell: the voiceless consonant is on the left and the voiced one is on the right. For nasals, liquids and glides (for which there is usually no voiceless correspondent), the symbol will be placed in the centre.

We have also included in this table some other consonants that we have not discussed, but are commonly featured in linguistics problems: the glottal stop ([ʔ]), as well as the velar fricatives ([x] and [ɣ]). We can describe every consonant with reference to the properties we've just outlined, for example:

[f] = voiceless labiodental fricative

[m] = bilabial nasal

[g] = voiced velar stop

In doing this, the usual order is voicing – place – manner.

Table 3.1: Consonants

	labial				coronal						dorsal					
	bilabial		labio-dental		(inter) dental		alveolar		post-alveolar		palatal		velar		glottal	
stops	p	b					t	d			c	ɟ	k	g	ʔ	
fricatives			f	v	θ	ð	s	z	ʃ	ʒ			x	ɣ	h	
affricates							ts	dz	tʃ	dʒ						
nasals		m						n				ɲ		ŋ		
liquids																
lateral								l								
rhotic								r								
glides		(w)										j		(w)		

3.2.1.4 Other characteristics of consonants

Besides the three main characteristics mentioned above, consonants can also have other features, such as:

- Aspiration – some consonants, especially stops, can be aspirated, i.e., pronounced with a little puff of air. This feature is marked by a superscript letter *h* after the consonant symbol. Thus, an aspirated, voiceless, bilabial stop can be written as [pʰ], since the sound is similar to the pronunciation of the consonant, followed by an [h]. Although in most languages of Europe consonants with the same place and manner of articulation are usually differentiated by voicing ([p] vs. [b], [f] vs. [v], etc.), many others differentiate these sounds based on aspiration; there are languages that have a three-way distinction between aspirated voiceless, unaspirated voiceless and unaspirated voiced versions of the same stop. For example, Mandarin Chinese does not have any voiced stops, but it has aspirated and unaspirated stops. Although the standard Chinese transliteration system (*pinyin*) uses the letters *p* and *b*, in reality, they correspond to the sounds [pʰ] and [p], respectively.
- Labialisation is similar to aspiration in the sense that consonants generally are non-labialised, but can exist in a labialised variant (called labialised consonants). An alternative term is rounding, which is also used for vowels (see below). Labialised consonants are pronounced with rounded lips, and they are marked in the IPA by a superscript [ʷ] symbol. Thus, a labialised voiced velar stop is written as [gʷ].

- PALATALISATION is another optional feature, similar to labialisation: in palatalised consonants, the back of the tongue is raised upwards, towards the hard palate. It is marked by a superscript [ʲ], though palatalisation can sometimes be indicated by an added apostrophe, e.g. [t'].

There are many other signs and symbols that can be added to a consonant to mark different alterations, but, if they are featured in a linguistics problem and are relevant to solving that problem, they will be described in the footnotes.

Besides palatalisation, there are other symbols that show the shifting of the tongue, e.g., velarisation (superscript [ˠ]) and pharyngealisation (superscript [ˤ]).

For example, while nasals and liquids are usually voiced, some languages also possess voiceless sonorants, for which there are no dedicated IPA symbols. To mark the voicelessness of a sonorant consonant or a vowel, a small ring can be placed under (in some cases, above) the corresponding symbol. For example, a voiceless alveolar nasal is written as [n̥].

3.2.2 Vowels

Vowels are produced without any constriction of the vocal tract, but with a subtle narrowing of the oral cavity by the tongue. The main features relevant to them are backness, height (aperture) and roundness.

3.2.2.1 Backness

Backness refers to the position of the tongue during articulation relative to the back of the mouth: to some extent, it is comparable with the place of articulation of consonants (and indeed in some languages the two might interact). Nevertheless, unlike consonants where the place of articulation is discrete (it has a small number of fixed values), backness is a more continuous parameter. On a very basic level, vowels are classified into FRONT, CENTRAL, and BACK, but they can also be subdivided further: between the front and central vowels there are NEAR-FRONT vowels and between central and back there are NEAR-BACK vowels. For simplicity, we will only consider the three basic values of backness, as follows:

- FRONT vowels: [i] (*ee* in *free*), [e] (*e* in Spanish), [ø] (*ö* in German or Turkish or *eu* in the French word *peu*), [y] (*ü* in German or Turkish, or *u* in the French *pu*), [ɛ] (*e* in *hen*), [a] (a front *a* is traditional in French words like *patte*);

- CENTRAL vowels: [ɐ] (*u* in *nut*), [ə] (*a* in *above*), [ɨ] (*ы* in Russian). There is no dedicated symbol in the IPA for a central low vowel (as in Spanish *a*): technically, IPA [a] is front, but in a linguistics problem <a> will often have the "European" value of the central low vowel;
- BACK vowels: [ɔ] (*a* in *call*),[o] (*oa* in *boat*, in Scottish English), [u] (*oo* in *boot*);

3.2.2.2 Height

Height (or aperture) refers to the vertical position of the tongue and the jaw when articulating the sound. Similar to backness, it is a continuous parameter. Generally, we talk about HIGH (CLOSE), MID, and LOW (OPEN) vowels. These can be further subdivided into NEAR-CLOSE (NEAR-HIGH), CLOSE-MID (HIGH-MID), OPEN-MID (LOW-MID), NEAR-OPEN (NEAR-LOW). Again, for simplicity, we will only consider the three main values, as follows:

- LOW vowels: [a], [ɐ], [æ];
- MID vowels: [e], [ɛ], [ɔ], [ø], [ə], [o];
  - HIGH-MID vowels: [e], [o], [ø];
  - LOW-MID vowels: [ɛ], [ɔ];
  - [ə] (called *schwa*) is somewhat special: it is central in backness and mid in height (it represents the resting state of the vocal tract). It can pattern in different ways in different languages and is often (but not always) found only in unstressed syllables.
- HIGH vowels: [i], [y], [ɨ], [u].

3.2.2.3 Roundness

The last main characteristic of vowels is roundness. This, similar to the voicing of consonants, is a binary parameter. We can differentiate:

- ROUNDED vowels – for which the lips are rounded during articulation: [o], [u], [y], [ø], [ɔ];
- UNROUNDED vowels – for which the lips are not rounded during articulation: [ɐ], [æ], [ə], [ɨ], [e], [ɛ], [i].

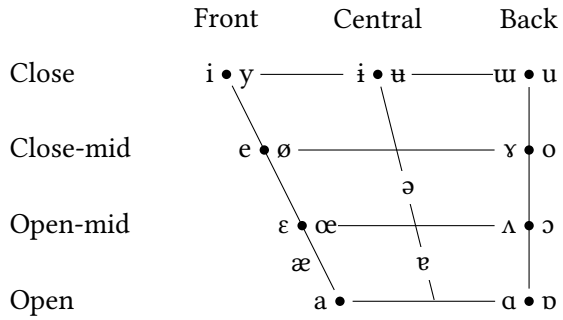


Figure 3.1: Vowel space

We can present all these characteristics as in Figure 3.1. This time, since both the backness and height are continuous parameters, it is not a table, but a diagram showing the *vowel space*.

In this diagram, the two dimensions are height (y-axis) and backness (x-axis), while roundness is determined by the vowel position relative to the reference point: symbols to the left of the dot refer to unrounded vowels and those on the right are rounded.

3.2.2.4 Other characteristics of vowels

There are some other features of vowels, usually marked by diacritics or superscript symbols. For example:

- Tongue root position – can have three values:
  - NEUTRAL = the default tongue root position;
  - ADVANCED = when pronouncing the vowel, the tongue root is slightly shifted towards the front. These vowels are called ATR (advanced tongue root) and are marked in the IPA by the diacritic ◌̟;
  - RETRACTED = the tongue root is shifted towards the back (retracted). These vowels are called RTR (retracted tongue root) and are marked by ◌̠;

  Generally, for languages in which the tongue root position is relevant for the vowel articulation, neutral and retracted positions are represented similarly: vowels are treated as either [+ATR] (the tongue root position is advanced) or [−ATR] (the tongue root position is neutral or retracted). Importantly, this feature of vowels is not the same as backness; there can

be front [+ATR] vowels, back [+ATR] vowels, front [–ATR] vowels and back [–ATR] vowels. Moreover, this feature is often relevant to languages that display vowel harmony processes (see Section 4.5), such as many languages of Africa. In some languages, the following pairs of vowels often pattern as if they were distinguished by the feature [±ATR] (Table 3.2).

Table 3.2: Vowels and the feature ±ATR

[+ATR]	[–ATR]
[u]	[ʊ]
[i]	[ɪ]
[o]	[ɔ]
[e]	[ɛ]
[a]	[ə]

- Nasalisation is another commonly encountered feature (present, for instance, in French and Portuguese). In the articulation of nasalised vowels, the air escapes through both the mouth and the nose. Nasalised vowels are marked by a tilde (~) above the vowel symbol: thus, the nasalised variant of the vowel [o] is [õ] (as in the French word *bon* [bõ]).

- Length can also play an important role in some languages. It refers to the duration of the vowel and we can differentiate at least between short and long vowels (compare *Ken* and *cairn* in most non-rhotic English dialects or, for some speakers, the French words *mettre* and *maître*).[4] Long vowels are marked in IPA by the symbol [ː] placed after the vowel (this symbol is in fact made up of two triangles pointing towards one another and not by a colon. Nevertheless, in practice, the colon [:] is often used for simplicity). In linguistics problems, a long vowel can also be marked by doubling the vowel (thus, *a* is the vowel [a] and *aa* is [aː]) or by a bar above the vowel (*ā*). The same conventions can apply to long consonants, if they exist.

A final important characteristic of sounds is syllabicity. This denotes the ability of a sound to be the nucleus of a syllable (see below the description of syllable structure). In some languages, vowels are the only possible syllabic sounds. Nevertheless, some consonants can be syllabic in certain languages (e.g., Cantonese)

[4]Some languages may have a three-way length distinction.

and this feature is marked by a vertical line below the respective consonant (thus, syllabic *m* is written as [m̩]: consider the English interjection *mmmkay* for 'OK', which we can transcribe as [m̩.keɪ]). We also find syllabic consonants in English words like *even* [iː.vn̩] or *rhythm* [ɹɪ.ðm̩]. Conversely, glides such as [w] and [j] are occasionally treated as non-syllabic versions of their respective vowels, which is signalled by an arch diacritic under the symbol: thus, [i̯] and [u̯] are broadly equivalent to [j] and [w].

3.3 Syllable

A syllable is defined as a group of sounds (phonemes) which are in some way pronounced "together". In phonetic transcription, the boundary between syllables is marked by a full stop [.]: the word *conclusion* can be transcribed as [kən.kluː.ʒən]. A syllable is comprised of three parts:

- The NUCLEUS is the "core" of the syllable: the sound occupying the nucleus is always syllabic. (This is generally what is meant when you hear that a syllable can have one and only one vowel: this is true, but as we have seen, a syllabic sound does not have to be a vowel);
- The ONSET represents the beginning of the syllable (everything before the nucleus);
- The CODA represents the end of the syllable (everything after the nucleus).

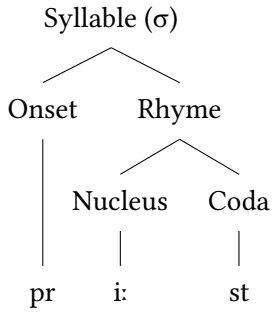


Figure 3.2: Segmentation of the word *priest* [priːst].

The nucleus is the only mandatory component of a syllable: the other two are usually optional. Thus, in English, some syllables consist only of the nucleus (for

example the first syllable of *upon* [ə.pɒn]), some have an onset and a nucleus but no coda (*me* [mi]), some have a nucleus and a coda but no onset (*at* [æt]), and yet others have all three (*cat* [kæt]). A further useful concept is the RHYME, which represents the combination of the nucleus and the coda. For example, a possible segmentation of the word *priest* [pri:st] is given in Figure 3.2.

3.3.1 Syllabification rules

In this subsection, we will consider the rules that we can follow in placing the syllable boundary. (For simplicity, we will not consider syllabic consonants and will only treat vowels as syllable nuclei.) A syllable generally contains only one vowel (or diphthong):[5] therefore, two consecutive vowels will always be part of two different syllables (VV ⟶ V.V).[6] If there is a single consonant between two vowels, this will belong to the second syllable (VCV ⟶ V.CV). This is because the basic principle of syllabification is that *a consonant prefers to be an onset rather than a coda.*

Things are somewhat more complicated when there are two consonants between a vowel. One possibility is that the syllable boundary goes in the middle of the consonant "cluster" so that one consonant becomes the coda of the first syllable and the other becomes the onset of the second syllable (VCCV ⟶ VC.CV). In this case, the absence of consonant clusters within the syllable overrides the preference for consonants to be in an onset. Alternatively, the entire cluster can act as an onset, avoiding the coda (VCCV ⟶ V.CCV). This can also happen, but is usually the more complicated case, in that not all clusters are equally suitable for such a syllabification, so it is perhaps more prudent to assume that VCCV ⟶ VC.CV is the default pattern.

3.4 Versification

Versification problems are a special type of linguistics problem that involves a set of series of lines (verses) in a particular language, often left even without a translation. Their main purpose is determining the rules governing the structure of the verse (*versification*, occasionally also *prosody*, although the latter term has many other meanings). This type of problem can usually be solved using the following method:

[5] A diphthong is a group consisting of two vowel-like sounds that pattern together as if they were a single vowel, such as in *blind* [blaɪnd] or *lie* [laɪ].

[6] However, remember that sequences like *aa* could also represent single instances of long vowels rather than two vowels in a row: check the notes for each problem!

Step 1. Syllabify all the words in each verse. A couple of issues might arise here. First, if you see two vowel symbols next to each other, make sure you know if it is a long vowel, a diphthong (a complex nucleus containing two vowel-like sounds), or a hiatus (a sequence of two vowels belonging to separate syllables). This might be explained in the footnote at the end of the problem. Second, you may need to consider whether the word boundary is relevant: sometimes each word should be syllabified on its own, but sometimes the entire verse should be treated as a single entity.

Step 2. Determine the type of syllable. With these problems, there are two kinds of criterion that can be used. Sometimes the relevant distinction is between syllables that contain short vowels (V) and those that contain long vowels or a diphthong (VV); sometimes, syllables without a coda contrast with those that have one; sometimes both of these criteria apply. Importantly, the onset is generally irrelevant for this type of problem.

Step 3. Determine the metre, that is to say, the rules regulating the kinds of syllables a verse can contain, and any restriction on their ordering. Prosodic systems generally use the criteria listed under step 2 to classify syllables as HEAVY or LIGHT syllables, but languages can differ in the details of this process:

- Sometimes, the distinction is solely based on the length of the vowel: light syllables have the structure (C)V(C), while heavy ones have the structure (C)VV(C);
- In other languages, the distinction is solely based on the existence of a coda: a syllable counts as light if it is OPEN, i.e., lacks a coda and thus has the structure (C)V(V), and as heavy if it is CLOSED, having the structure (C)V(V)C;
- In many languages, a syllable counts as heavy if it either contains a long vowel –(C)VV– or if it is closed – (C)V(C)C; in other words if its rhyme contains more than one element;
- Finally, we cannot exclude the possibility that some languages have three types of weight – light syllables (like (C)V), heavy syllables (like (C)VV or (C)VC), and superheavy syllables (like (C)VVC).

All this information is represented schematically in Figure 3.3.

Therefore, we can certainly tell that (C)V syllables will always be light and (C)VVC will always be heavy, while the other two types can belong to either

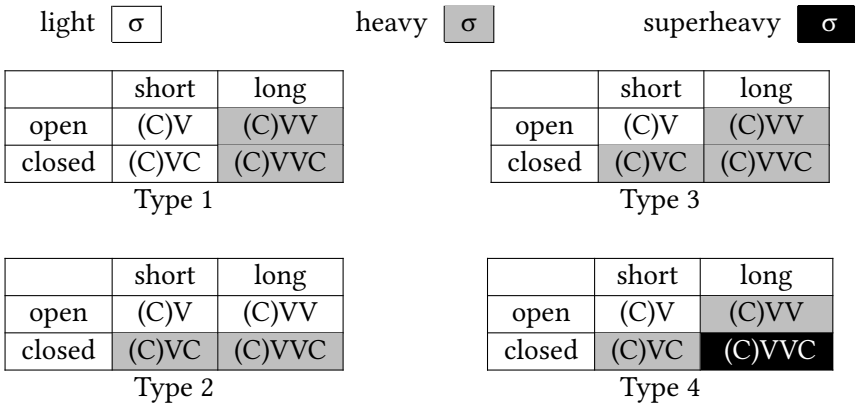


Figure 3.3: Versification schema.

category. The final aim of the problem is to determine the structure of the verse (represented, in general, by a sequence of light and heavy syllables in a particular order). One very important thing to note is that, in many cases, there is an equivalence between one heavy syllable and two light syllables. A good indicator of this phenomenon is if the verses have a variable number of syllables.

Problem 3.1

Somali (Alexander Piperski, IOL 2015)

Here are 25 half-lines of Somali poetry written in a metre known as masafo:

1. *ogaadeen ha ii dirin*
2. *duul haad amxaaraa*
3. *kaa dooni maayee*
4. *amba waa ku daba geli*
5. *dakanka iyo qaankee*
6. *anaa been dabaadee*
7. *galbeed uga dareershaan*
8. *dalkaad adigu joogtiyo*
9. *dar alliyo heshiis iyo*
10. *mase waa dayoobeen*
11. *dacalkaaga kuma shuban*
12. *miyaan duudsiyaayaa*
13. *doodaye maxaad oran*
14. *daliilkii ku siiyaye*
15. *miyaad iigu duurxuli*
16. *dorraad adigu kama dhigin*
17. *ma deldelin raggoodii*
18. *deelqaadkan aad tiri*
19. *diigaanyo ciidana*
20. *wax ma kala dillaallaa*
21. *duunyada ka qaadoo*
22. *diinkiyo dugaaggiyo*
23. *dildillaaca waaberi*
24. *dibnahaaga kama qiran*
25. *hobyo wixii ka soo degey*

To help you understand the structure of masafo, here are ten half-lines which were constructed from genuine masafo half-lines by random rearrangement of words within the half-line. Some of them might conform to the rule of versification, but the majority do not:

26. *u anigaa lehe diin*
27. *waad nimankaad ma diidi*
28. *qoran daftarkaaga kuma*
29. *fuushaan kaama dusha*
30. *helo dabacayuun kulaan*
31. *kuu miyuu tari wax dafir*
32. *kuu daalasaayee nin*
33. *shareecada dikrigiyo*
34. *dumarkii furayaan ma*
35. *ogaadee diyaar kuu*

Problem 3.1a Describe the structure of a masafo half-line.

Problem 3.1b Here are ten more masafo half-lines. Five of them are genuine, and five of them have been obtained by random rearrangement. Which is which?

36. *war ismaaciil daarood*
37. *dir miyaad wadaagtaan*
38. *labadaad ka duudiye*
39. *ka jannadaad daahiye*
40. *adiga iyo deriskaa*
41. *digaxaarka mariyoo*
42. *ciid iyo doolo diraac*
43. *nooma keeneen darka*
44. *kala deyaayaa miyaan*
45. *wuxuun kaa danqaabaan*

Solution

Firstly, we notice that there is no footnote about any of the sounds, so we can consider that there are no diphthongs and that two consecutive identical vowels (*aa*, *ii*, etc.) most likely denote a long vowel.

Next, we need to syllabify the structures. We can do that using the rules above (VV → V.V, VCV → V.CV, VCCV → VC.CV). We use a dash to mark those places where the word boundary could make a difference for the syllabification (for example, if we want to syllabify the phrase *come inside* [kʌm ɪnsaɪd], if we are to take into account the word boundary, we would get [kʌm ɪn.saɪd]. However, if the word boundary is ignored, which happens quite often in rapid speech, we would get [*kʌ.mɪn.saɪd*]).

The verses become:

1. *o.gaa.deen.ha.ii.di.rin*
2. *duul.haad-am.xaa.raa*
3. *kaa.doo.ni.maa.yee*
4. *am.ba.waa.ku.da.ba.ge.li*
5. *da.kan.ka.i.yo.qaan.kee*
6. *a.naa.been.da.baa.dee*
7. *gal.beed-u.ga.da.reer.shaan*
8. *dal.kaad-a.di.gu.joog.ti.yo*
9. *dar-al.li.yo.hes.hii.s-i.yo*
10. *ma.se.waa.da.yoo.been*
11. *da.cal.kaa.ga.ku.ma-shu.ban*
12. *mi.yaan.duud.si.yaa.yaa*
13. *doo.da.ye.ma.xaad-o.ran*
14. *da.liil.kii.ku.sii.ya.ye*
15. *mi.yaad-ii.gu.duur.xu.li*
16. *dor.raad-a.di.gu.ka.ma-dhi.gin*
17. *ma.del.de.lin.rag.goo.dii*
18. *deel.qaad.kan-aad.ti.ri*
19. *dii.gaan.yo.cii.da.na*
20. *wax.ma.ka.la.dil.laal.laa*
21. *duun.ya.da.ka.qaa.doo*
22. *diin.ki.yo.du.gaag.gi.yo*
23. *dil.dil.laa.ca.waa.be.ri*
24. *dib.na.haa.ga.ka.ma.qi.ran*
25. *hob.yo.wi.xii.ka.soo.de.gey*

In the second verse, *haad-am* means that there is a word boundary between *haad* and *am*. If we ignore it, the syllables become *haa.dam*.

In order to simplify the problem, we can replace each syllable with the following notations (based on the syllable typology described above):

1 = syllable with short vowel and no coda = (C)V

2 = syllable with long vowel and no coda = (C)VV

3 = syllable with short vowel and coda = (C)VC

4 = syllable with long vowel and coda = (C)VVC

The corpus becomes:

1. *1.2.4.1.2.1.3*
2. *4.haad-am.2.2*
3. *2.2.1.2.2*
4. *3.1.2.1.1.1.1.1*
5. *1.3.1.1.1.4.2*
6. *1.2.4.1.2.2*
7. *3.beed-u.1.1.4.4*
8. *3.kaad-a.1.1.4.1.1*
9. *dar-al.1.1.3.hiis-i.1*
10. *1.1.2.1.2.4*
11. *1.3.2.1.1.ma-shu.3*
12. *1.4.4.1.2.2*
13. *2.1.1.1.xaad-o.3*
14. *1.4.2.1.2.1.1*
15. *1.yaad-ii.1.4.1.1*
16. *3.raad-a.1.1.1.ma-dhi.3*
17. *1.3.1.3.3.2.2*
18. *4.4.kan-aad.1.1*
19. *2.4.1.2.1.1*
20. *3.1.1.1.3.4.2*
21. *4.1.1.1.2.2*
22. *4.1.1.1.4.1.1*
23. *3.3.2.1.2.1.1*
24. *3.1.2.1.1.1.1.3*
25. *3.1.1.2.1.2.1.3*

It's time to analyse how the word boundary affects the syllable type. Let us take, for example, the structure *haad-am* from half-line 2.

- if we take into account the word boundary, we get: *haad am* ⟶ *haad.am (4.3)*
- otherwise, we get: *haadam* ⟶ *haa.dam (2.3)*

Therefore, we notice that the only difference a word boundary makes is that a coda becomes the onset of the following syllable, so the first syllable can change from type *1* to *3* or from type *2* to *4*. The second syllable, on the other hand, is not affected in any way, because it can only receive (or lose) an onset, which, as mentioned above, is irrelevant for syllable type. If we mark the syllable before the word boundary with *(2/4)* or *(1/3)*, meaning that it can be either type *2* or *4* (or *1* or *3*, respectively), depending on whether we take into account the word boundary, and knowing that the following syllable does not change its type, we get:

1. *1.2.4.1.2.1.3*
2. *4.(2/4).3.2.2*
3. *2.2.1.2.2*
4. *3.1.2.1.1.1.1.1*
5. *1.3.1.1.1.4.2*
6. *1.2.4.1.2.2*
7. *3.(2/4).1.1.1.4.4*
8. *3.(2/4).1.1.1.4.1.1*
9. *(1/3).3.1.1.3.(2/4).1.1*
10. *1.1.2.1.2.4*
11. *1.3.2.1.1.(1/3)/1.3*
12. *1.4.4.1.2.2*
13. *2.1.1.1.(2/4).1.3*
14. *1.4.2.1.2.1.1*
15. *1.(2/4).2.1.4.1.1*
16. *3.(2/4).1.1.1.1.(1/3).1.3*
17. *1.3.1.3.3.2.2*
18. *4.4.(1/3).4.1.1*
19. *2.4.1.2.1.1*
20. *3.1.1.1.3.4.2*
21. *4.1.1.1.2.2*
22. *4.1.1.1.4.1.1*
23. *3.3.2.1.2.1.1*
24. *3.1.2.1.1.1.1.3*
25. *3.1.1.2.1.2.1.3*

Next, we notice that the number of syllables in each half-line varies between five and nine. Now recall that some versification systems make it possible to treat heavy syllables and two or more light syllables as equivalent. Nevertheless, we still need to uncover how the light and heavy syllables are defined (either light = 1, 3 and heavy = 2, 4; or light = 1, 2 and heavy = 3, 4). In order to do that, we look at the shortest (five-syllable) half-lines:

2.2.1.2.2 *4.(2/4).3.2.2*

Since they are the shortest, we expect that they have the same structure (since there is more scope in longer verses for light syllable sequences). We can see that type *2* heavy syllables ((C)VV) occur in the same positions in the verse as type *4* syllables ((C)VVC), whilst the type *1* light syllable ((C)V) matches a type *3* syllable ((C)VC). Therefore, we deduce that this metre features a distinction based on vowel length: light syllables are those which have a short vowel, while heavy syllables are those which have a long vowel; the presence of a coda does not make a syllable heavy. We can now rewrite all the half-lines as a sequence of light and heavy syllables, replacing *1* and *3* by L and *2* and *4* by H. We obtain the following:

1. L.H.H.L.H.L.L
2. H.H.L.H.H
3. H.H.L.H.H
4. L.L.H.L.L.L.L.L
5. L.L.L.L.L.H.H
6. L.H.H.L.H.H
7. L.H.L.L.L.H.H
8. L.H.L.L.L.H.L.L
9. L.L.L.L.L.H.L.L
10. L.L.H.L.H.H
11. L.L.H.L.L.L.L.L
12. L.H.H.L.H.H
13. H.L.L.L.H.L.L
14. L.H.H.L.H.L.L
15. L.H.H.L.H.L.L
16. L.H.L.L.L.L.L.L.L
17. L.L.L.L.L.H.H
18. H.H.L.H.L.L
19. H.H.L.H.L.L
20. L.L.L.L.L.H.H
21. H.L.L.L.H.H
22. H.L.L.L.H.L.L
23. L.L.H.L.H.L.L
24. L.L.H.L.L.L.L.L
25. L.L.L.H.L.H.L.L

Note that the question regarding word boundary has turned out to be irrelevant: whether we count the ambiguous consonant as a coda or not does not influence the weight of the preceding syllable.

Next, we can group the half-lines based on their number of syllables:

5 syllables	6 syllables	7 syllables	8 syllables
H.H.L.H.H	L.H.H.L.H.H	L.H.H.L.H.L.L	L.L.H.L.L.L.L.L
H.H.L.H.H	L.L.H.L.H.H	L.L.L.L.L.H.H	L.H.L.L.L.H.L.L
	L.H.H.L.H.H	L.H.L.L.L.H.H	L.L.L.L.L.H.L.L
	H.H.L.H.L.L	H.L.L.L.H.L.L	L.L.H.L.L.L.L.L
	H.H.L.H.L.L	L.H.H.L.H.L.L	L.L.H.L.L.L.L.L
	H.L.L.L.H.H	L.H.H.L.H.L.L	L.L.L.H.L.H.L.L
		L.L.L.L.L.H.H	
		L.L.L.L.L.H.H	
		H.L.L.L.H.L.L	
		L.L.H.L.H.L.L	

Only one five-syllable half-line (H.H.L.H.H) seems to be possible. So, based on the above, we expect that the six-syllable half-lines can be formed by replacing one heavy syllable (H) with two light syllables (LL). Indeed, this is how most of the six-syllable half-lines can be derived. Nevertheless, two half-lines do not follow this pattern: instead, they have the structure (L.H.H.L.H.H), which is identical to that of the five-syllable half-line, but with an additional light syllable at the beginning.

Starting from the five-syllable half-line and replacing any of the heavy syllables with two light syllables (H ⟶ LL), we obtain the following possible combinations:

5 syllables	6 syllables	7 syllables	8 syllables
H.H.L.H.H	L.L.H.L.H.H	L.L.L.L.L.H.H	L.L.L.L.L.L.L.H
	H.L.L.L.H.H	L.L.H.L.L.L.H	L.L.L.L.L.H.L.L
	H.H.L.L.L.H	L.L.H.L.H.L.L	L.L.H.L.L.L.L.L
	H.H.L.H.L.L	H.L.L.L.L.L.H	H.L.L.L.L.L.L.L
		H.L.L.L.H.L.L	
		H.H.L.L.L.L.L	

Comparing our data with the types shown in the table above, we are left with the following lines still not accounted for by the rule:

6 syllables	7 syllables	8 syllables
L.H.H.L.H.H	L.H.H.L.H.L.L	L.L.L.H.L.H.L.L
L.H.H.L.H.H	L.H.L.L.L.H.H	L.H.L.L.L.H.L.L
	L.H.H.L.H.L.L	
	L.H.H.L.H.L.L	

As previously, all the remaining six-syllable half-lines in masafo have the structure L.H.H.L.H.H, equivalent to a five-syllable half-line with an extra light syllable in the beginning. Furthermore, all the remaining seven- and eight-syllable half-lines are also formed by replacing one or two heavy syllables in this type of six-syllable structure with light syllables (H ⟶ LL).

The only half-line we have not analysed so far is the nine-syllable one, which has the structure L.H.L.L.L.L.L.L.L. We notice that this one is also derived from the same six-syllable half-line:

$$\text{L.H.L.L.L.L.L.L.L} \Longleftrightarrow \text{L.H.(L.L).L.(L.L).(L.L)} \Longleftrightarrow \text{L.H.H.L.H.H.}$$

We can therefore deduce that all half-lines in the masafo metre derive from either H.H.L.H.H or L.H.H.L.H.H. This can also be written as (L).H.H.L.H.H, in which any of the heavy syllables (H) can be replaced by two light syllables (LL).

Of course, there are other approaches which yield the same result. For example, we can notice that most half-lines end in two heavy syllables; where this is not the case, we always have a heavy syllable followed either by two light syllables or just by four light syllables. We discover that all half-lines end with H.H (where each H can be replaced by L.L). Excluding these syllables, we are left with:

5 syllables	6 syllables	7 syllables	8 syllables	9 syllables
H.H.L	L.H.H.L	L.H.H.L	L.L.H.L	L.H.L.L.L
H.H.L	L.L.H.L	L.L.L.L.L	L.H.L.L.L	
	L.H.H.L	L.H.L.L.L	L.L.L.L.L	
	H.H.L	H.L.L.L	L.L.H.L	
	H.H.L	L.H.H.L	L.L.H.L	
	H.L.L.L	L.H.H.L	L.L.L.H.L	
		L.L.L.L.L		
		L.L.L.L.L		
		H.L.L.L		
		L.L.H.L		

Next, we notice that all half-lines end in a light syllable. Excluding this syllable as well, we get the following table. Note that we removed all duplicate structures (i.e., if, for example, all five-syllable verses remained with the structure H.H, we only included it once in the table:

5 syllables	6 syllables	7 syllables	8 syllables	9 syllables
H.H	L.H.H	L.H.H	L.L.H	L.H.L.L
	L.L.H	L.L.L.L	L.H.L.L	
	H.H	L.H.L.L	L.L.L.L	
	H.L.L	H.L.L	L.L.L.H	
		L.L.H		

Repeating the same thought process we notice, again, that most half-lines end in two heavy syllables or their equivalent (that is, L.L.L.L or H.L.L or L.L.H). If we also exclude these, we get:

6 syllables	7 syllables	8 syllables	9 syllables
L	L	L	L

Now we are left (only in some half-lines) with a light syllable in the beginning; we infer that this syllable is optional.

Both approaches yielded the same result: the structure of a masafo half-line is (L).H.H.L.H.H, where H can be replaced by L.L; H represents a heavy syllable (with a long vowel), while L represents a light syllable (with a short vowel).

In order to solve task (b), we need to follow the same algorithm: syllabification and marking the syllables as L or H. We get:

36. L.L.H.H.H.H
37. L.L.H.L.H.H
38. L.L.H.L.H.L.L
39. L.L.L.H.H.L.L
40. L.L.L.L.L.L.H
41. L.L.H.L.L.L.H
42. H.L.L.H.L.L.H
43. H.L.H.H.L.L
44. L.L.L.H.H.L.H
45. L.H.H.L.H.H

We already know that each half-line needs to end in two heavy syllables (or the equivalent with light syllables), and this needs to be preceded by a light syllable (L.H.H /L.L.L.H/L.H.L.L/ L.L.L.L.L), so we can easily figure out that the half-lines 36, 39, 42, 43, and 44 are not genuine. Since we are told that exactly five of them are not genuine, we know for sure that the rest are genuine.

3.5 Stress

In another type of linguistics problem based solely on phonetic concepts, we are given some words with the stress marked and asked to identify the pattern of stress placement and apply it to other words. Roughly, a stressed syllable is pronounced with more emphasis, allowing us to differentiate between, e.g., ***con****trast* (stress on *con*, as in the noun) and *con****trast*** (stress on *trast*, as in the verb). The stress can also be marked by a prime symbol before the relevant syllable.[7] This

[7]For example, in the IPA notation, the phonetic transcription of the two words would be [ˈkɒn.trɑːst] and [kən.ˈtrɑːst], respectively.

stress is also called *primary* stress and, by definition, each word only contains one syllable with a primary stress.

From a typological perspective, some languages have fixed stress and others have mobile (or variable) stress. For those which have fixed stress, the stress is usually placed towards the edge of the word (either the beginning or the end). Therefore, there are languages in which the primary stress always falls on the last syllable and languages in which stress always falls on the first syllable. Alternatively, there are languages in which stress always falls on the second or penultimate syllable or even the third or antepenultimate syllable. Generally, the first and last three syllables of a word are the most likely stress positions.

Languages which have variable stress can be further subdivided into two classes: languages with "free" stress and languages with "predictable" stress. By free stress is meant that on a word-by-word basis you cannot predict the stress pattern (but each word will have a fixed pattern). In languages with "predictable" stress, the stress placement follows some well-defined rules, usually related to the weight of syllables. In linguistics problems, we usually find languages with predictable stress, because free stress cannot be predicted, while fixed stress is extremely easy to analyse.

In the case of languages with predictable stress, stress is usually placed within a *window* (i.e., a group of syllables, with the accent always falling on one of these syllables, based on certain rules). The most common stress windows are the first and the last three syllables and, within these windows, certain rules dictate stress placement, which are often similar to the rules in versification problems. For instance, stress could always fall on the syllable that contains a long vowel (within a certain stress window), and, if there is no such syllable, it will be placed on the first syllable within that window. Generally, all these rules are based on syllable weight (i.e., vowel length and/or the presence of a coda). It is possible, but rare, for stress placement to also be determined by factors such as the quality of the vowel (front/back, open/close, etc.).

To give you an idea of how complex such systems can be, let us consider the case of the Pirahã language.[8] In this language, the stress window is represented by the final three syllables (thus, the primary stress will always be placed on one of the last three syllables). The placement of the stress is determined by a hierarchy of syllable types that takes into account both the type of vowel (long or short) and the TYPE OF ONSET (an extremely rare thing). Specifically, syllables with a long vowel are always higher in the hierarchy ("heavier") than those with a short vowel, while syllables that feature a voiceless onset are higher than those

[8]This phenomenon was featured in a problem by Artūrs Semeņuks (IOL 2013).

that have a voiced onset (which are themselves higher than those which have no onset). If we denote a voiced consonant as G and a voiceless consonant as C (and, as previously, use V for a short vowel and VV for a long vowel or diphthong), the syllable hierarchy in Pirahã is: CVV > GVV > VV > CV > GV > V. Furthermore, if there are several syllables of the same type within the window, stress will be placed on the rightmost syllable (last > penultimate > antepenultimate). For example, in the word [ka.gi.hi] 'wasp' we have two CV syllables ([ka] and [hi]) and one GV syllable ([gi]]. We know that CV syllables have priority, so, certainly, the stress will not be placed on [gi]. Amongst the remaining two syllables (which are of the same type), the stress will be placed on the rightmost syllable, so the word is stressed [ka.gi.ˈhi].

By contrast in [ʔi.soo.bai] 'otter' stress goes on [soo]: it has priority over the first syllable since it has a long vowel and it also has priority over the last syllable since it has a voiceless onset.

Apart from primary stress, some languages can also have *secondary* stresses. These are also marked with a prime symbol, but placed below the line (ˌ). There is no limit as to how many secondary stresses a word can have and usually the rules for placing the secondary stress are based on the type of syllable or vowel (e.g., all syllables with a long vowel that do not bear primary stress will receive secondary stress) or related to the primary stress (e.g., starting from the primary stress towards the right, every other syllable receives secondary stress).

Problem 3.2

Manobo (Saujas Vaduguru, PLO 2020)

Given below are some words in Sarangani Manobo. In each word, the location of the stress is marked with ˈ at the beginning of the stressed syllable:

> *ˈbaso, deˈitek, meneˈnoo, beˈgas, leˈkat, ˈotaw, eteˈbay, binaleˈsan, ˈdeget, ˈbenget, miˈneles, ˈikan, ˈdoen*

Problem 3.2a Mark the stress in the following words:

mengabat	*iselem*
tadon	*mola*
migbasa	*benal*
belegkong	*medaet*

Problem 3.2b Given a new word in Sarangani Manobo, how would you identify the syllables in the word? Once you have identified the syllables, how would you determine which syllable is stressed?

Solution

Syllabification has already been discussed above; we can apply the usual rules (VV → V.V, VCV → V.CV, VCCV → VC.CV).

The first step is to notice where the stress is usually placed. We know it is most likely placed in a window either at the beginning or at the end of the word. Looking at the word *binele'san*, we infer that, most likely, the stress window is at the end of the word; if it were in the beginning, it would be four syllables long, which is unlikely. Assuming the stress window is at the end of the word, we notice that the stress can only fall on the last or penultimate syllable.

The next step is splitting the words into two groups, based on the stress position:

Penultimate syllable	Final syllable
'ba.so	*bi.ne.le.'san*
'de.get	*e.te.'bay*
'ben.get	*le.'kat*
mi.'ne.les	*be.'gas*
'i.kan	*me.ne.'noo*
'do.en	
'o.taw	
de.'i.tek	

Since most of the words have the stress placed on the penultimate syllable, we can assume that this is the default position and, in some cases, the stress shifts to the last syllable. The first hypothesis is that the stress moves to the last syllable if it contains a long vowel (*me.ne.'noo*), but this is the only word that contains a long vowel: such a rule is unlikely to help us with the task. Another idea to consider is that all words in which the stress is on the last syllable end in a closed syllable (with a coda) and furthermore in all these cases the penultimate syllable is open (lacks a coda). Therefore, we could hypothesise that the stress falls, in general, on the penultimate syllable, and moves to the last syllable if the latter is closed whilst the penultimate syllable is open. This hypothesis is also quickly rejected

since we have words such as *de.'i.tek* and *'o.taw* in which the penultimate syllable is open and the last syllable is closed, but the stress still falls on the penultimate syllable.

Therefore, vowel length and syllable type seem to not play an important role in stress placement. Looking at the type of vowel, we notice that all words in which the stress falls on the last syllable contain the vowels *a* or *o*. Perhaps stress prefers a syllable that contains these two vowels? However, this also proves to be wrong since we have words such as *'i.kan*.

Up until now, we have tried finding a reason for stress to move to the last syllable rather than the penultimate (that is, stress is *attracted* to the final syllable by virtue of some property of that syllable, such as a long vowel, coda, or the vowels *a* or *o*). Perhaps, instead, it is the opposite: could stress actually be *repulsed* from the penultimate syllable to the last one in some contexts? We can observe that in all words with stress on the last syllable, the penultimate syllable contains the vowel *e*. Therefore, we need to consider the hypothesis that stress prefers to be on a syllable that does not contain the vowel *e*.

We can try and classify the words based on the presence of the vowel *e* in the last or penultimate syllable:

	Penultimate syll.	
Last syll.	V=*e*	V≠*e*
V=*e*	*'deget* *'benget* *mi'neles*	*de'itek* *'doen*
V≠*e*	*mene'noo* *be'gas* *le'kat* *ete'bay* *binele'san*	*'baso* *'otaw* *'ikan*

We notice that if only one of the last two syllables contains the vowel *e*, stress is placed on the other syllable. If both syllables contain the vowel *e*, stress falls on the penultimate syllable; the same happens if none of them contains the vowel *e*. Alternatively, we can describe the situation thus: stress falls on the last syllable only if it does not contain the vowel *e* and the penultimate syllable does. An even simpler way of formulating the rule is that stress is placed on the penultimate syllable, but is repulsed to a final syllable by the vowel *e*. Therefore:

- if none of the vowels is *e*, the stress remains in the default position, the penultimate syllable;
- if the last vowel is *e*, the stress stays in the default position, on the penultimate syllable;
- if the penultimate vowel is *e*, the stress is forced to move to the last syllable. We have two cases:
  - if the last vowel is not *e*, the stress will favour this position and will remain on the last syllable;
  - if the last vowel is also *e*, the stress cannot fall on it either, so it will stay on the penultimate syllable since this is the default position.

Finally, if we return to the only example featuring a long vowel, *me.ne.ˈnoo*, we notice that there is an alternative explanation: *oo* is not a long vowel, but rather two short, independent vowels and the syllabification of the word is, in fact, *me.ne.ˈno.o*. Either way, the stress rule is followed. Since there is only one example of a double vowel and there is no footnote explaining its role, we can simply assume that there are two short vowels.

Thus, we can solve the two tasks:

Solution 3.2a

men.ˈga.bat	*i.ˈse.lem*	*ˈta.don*	*ˈmo.la*
mig.ˈba.sa	*be.ˈnal*	*be.leg.ˈkong*	*me.ˈda.et*

Solution 3.2b For syllabification, we use the rules:

VV → V.V VCV → V.CV VCCV → VC.CV

Stress falls on the penultimate syllable. If it contains the vowel *e* (and the last syllable does not), stress will move to the last syllable.

3.6 Tone

Another phonetic phenomenon is *tone*. This term covers the use of *pitch* to make distinctions in the meaning of words, and/or grammatical distinctions.

Technically, pitch is the frequency with which the vocal folds vibrate in the production of speech, also known as the *fundamental frequency* (for this reason, only voiced sounds are characterised by pitch). Most, if not all, languages make some use of pitch. For example, in many languages syllables that carry stress are characterised by a higher pitch than unstressed ones: most varieties of English

belong to this type. In some languages, such as for example Russian, the pitch trajectory is the only difference between a statement and a general question, in contrast to, say, English, which uses a special construction to distinguish between *You solved the problem* and *Did you solve the problem?*

Such uses of pitch are found in many languages of the world, and are covered by the label *intonation*. The term tone is narrower: we say that a language is a *tone language* if tone can make lexical or grammatical distinctions between words. For example, the syllable *ma* in Mandarin can have different meanings depending on the tone it receives (as we have seen in the previous chapter). As such, the word *mā*, with first tone, means 'mother', while the word *má* (second tone) means 'hemp', *mǎ* (third tone) means 'horse', and *mà* (fourth tone) means 'scold'.

In such cases, tonal differences are not really different from distinctions made using, say, two different consonants. There are also languages where changes in tone can express grammatical meanings, similar to how such meaning can be expressed by a morpheme (see Chapter 5 for more information about morphemes).

The terminology used for tone languages can differ quite a lot, so you may need to read the pronunciation notes carefully to make sure you understand the notation in the problem.

Tone, like stress, is an example of a *suprasegmental* phenomenon. Whilst properties such as, say, backness or place of articulation belong only to the vowel or consonant they are associated with, suprasegmental features characterise groups of consonants and vowels (collectively called *segments*). With stress, we saw that it is mostly a property of syllables, and this is often the case also with tone.

3.7 Practice problems

Problem 3.3

Old Javanese (Michael Salter, UKLO 2021)

The Kakawin poems of Old Javanese were long narrative tales made up of four-line stanzas. In the tradition of early Sanskrit poetry, each line was made up of a precise pattern of heavy and light syllables. The first section of the 11th-century CE Kakawin poem *Arjunawiwāha* ('The Marriage of Arjuna') consisted of lines which followed the *śārdūlawikrīdita* metre. In this metre, all lines have the same pattern: each consists of 19 syllables in an exact pattern of heavy and light syllables, except for the last syllable which can be either heavy or light. The first three syllables of a *śārdūlawikrīdita* line are all heavy.

Below are two stanzas from the opening section of the *Arjunawiwāha*:

1. *lakṣmī niŋ suraloka sampun ayaśâŋrĕñcĕm tapa mwaŋ brata*
   akweh saŋ pinilih pituŋ siki tikāŋ antuk niŋ okir mulat
   rwêkāŋ ādi Tilottamā paměkas iŋ kocap lawan Suprabā
   tapwan marma tuhun lĕhĕŋ lĕhĕŋa saŋkê rūpa saŋ hyaŋ Ratih

2. *tambenyân liniŋir kětêkin inaměr deniŋ watĕk dewata*
   sampūrna pwa ya mapradakṣina ta yâmūjâmiděr pintiga
   hyaŋ Brahmā dumadak caturmuka batārêndrâmahâkweh mata
   eraŋ miŋgĕka kociwâmbĕk ira yan kālanyan uŋgw iŋ wuri

Problem 3.3a Here are four more lines from the first section of the *Arjunawiwāha* (in the *śārdūlawikrīdita* metre), with their component parts (labelled A–D) *in random order*:

Verse 1:
- Part A: *yêkā rakwa*
- Part B: *kapwa tâmurṣita*
- Part C: *Indra sĕděŋ amwit*
- Part D: *kinon hyaŋ*

Verse 2:
- Part A: *daśagunan*
- Part B: *tan sora*
- Part C: *pwa tĕkap nikā*
- Part D: *rūpanya dentânaku*

Verse 3:
- Part A: *widyādarī mūr tĕhĕr*
- Part B: *sinambahakĕn iŋ*
- Part C: *liŋ hyaŋ*
- Part D: *Śakra nahan*

Verse 4:
- Part A: *lokika*
- Part B: *tan sangkêŋ*
- Part C: *lwir saŋgrahêŋ*
- Part D: *wiṣaya prayojñananira*

For each verse, place the parts A-D in the right order to obtain the original verse.

Problem 3.3b Below are given four more words which appear in different lines from the first section of the *Arjunawiwāha*:

paramārthapandita	*santosâhĕlĕtan*
ametmetâśrayā	*candanâpāyunan*

For each of them, write the number (from 1 to 19) of the syllable in their respective lines of which these words begin (e.g., 1 if the word is at the start of the line, 3 if the word's first syllable is the third in the line).

ā, *â*, *ê*, *ĕ*, *ī*, *ū* are vowels. *ŋ*, *ñ*, *ş*, *y* are consonants. You should assume that long vowels *ā*, *ī*, *ū* (and possibly others) occur only in heavy syllables.

Problem 3.4

Chuvash (Artūrs Semeņuks, LLO 2013)

Here are some Chuvash words, transcribed in Latin script. The stress is marked by a prime symbol before the respective syllable.

a'vallăh	'antiquity'	*malašne'hi*	'future'
asărha'nullă	'sensitive'	*měskěn'len*	'to respect'
ănsăr'tran	'unexpected'	*nušalan'tar*	'to make suffer'
'ăšăn	'to warm up'	*'pělětlě*	'cloudy'
'vărlăh	'seed'	*'pitěrěnčěk*	'closed'
'ěměrlěh	'for life'	*su'narşă*	'hunter'
jü'şek	'sour, bitter'	*ču'ralăh*	'slavery'
kansěr'le	'to trip'	*čuhăn'lan*	'to become poor'
kěrkunie'hi	'autumn'		

Problem 3.4a Mark the stress in the following words:

vĕltrentărri	'tit (bird)'	*jyvărlăh*	'difficulty'
vişmine	'overmorrow'	*măkărălčăk*	'convex'
ilĕrtüllĕ	'tempting'		

ă and *ĕ* are extra-short vowels, which are pronounced shorter than the other vowels in the language. *ü* and *y* are vowels; *ş*, *š* and *č* are consonants.

Problem 3.5

Ancient Greek (Dan-Mircea Mirea, RoLO 2019)

In Ancient Greece, poetry was the most celebrated art. There were several poetic forms, each with its own rules of rhythm, metre and rhyme. In order to describe these forms, the Greeks established the art of prosody, regulating the structure of verses. In this problem, you will decipher the structure of some poems of Ancient Europe.

A metrical foot is a sequence of short (*S*) and long (*L*) syllables repeated in a verse. A line consists of several feet, as in the following example:

L-S-X | L-S-X | L-S-X

The verse above contains nine syllables and three metrical feet. The metrical foot, in this case, is formed by three syllables *L-S-X*, where *X* is a syllable that can be either short or long. All the verses of a poem written in this metre will have the same structure. In Ancient Greek, a syllable counts as long if it contains a long vowel or a diphthong, or if it ends in a consonant. Syllabification does not necessarily take into account word boundaries.

Below are eight verses, transliterated into Latin script, from *Oedipus Tyrannus* (or *Oidipous Tūrannos*, in Ancient Greek), Sophocles' tragedy. These are written in the most common metre at that time. The number after the verse represents the verse's number in the original work.

ō tekna, Kadmou tou palai neā trop^{h}ē,	(1)
tinas pot^{h}' edrās tāsde moi t^{h}oazdetē	(2)
^{h}iktēriois kladoisin eksestemmenoi;	(3)
polis d' ^{h}omou men t^{h}ūmiāmatōn gemei,	(4)
^{h}omou de paianōn te kai stenagmatōn.	(5)
ēn ^{h}ēmin, ōnaks, Lāïos pot^{h}' ^{h}ēgemōn	(103)
poion logon; leg' aut^{h}is, ōs māllon mat^{h}ō	(358)
ouk^{h}i ksunēkas prost^{h}en; ^{h}ē 'kpeirā legōn;	(359)

Problem 3.5a Describe the syllabification rules and the structure of the metre in *Oedipus Tyrannus*. How many metrical feet does a verse have?

Problem 3.5b The following four verses were written by other important writers of Ancient Greek. Only one of them is written in the same metre as that of Sophocles. Which one is it?

(Aeschylus, *Prometheus bound*, 543)
t^{h}eit' ema gnōmā kratos antipalon Zeus

(Aristophanes, *The clouds*, 609)
prōta men k^{h}airein At^{h}ēnaioisi kai tois...

(Euripides, *Hippolytus*, 1054)
ei pōs dunaimēn, ōs son ek^{h}t^{h}airō karā.

(Herodas, *Mimiamb*, 14)
o pēlos ak^{h}ris ignuōn prosestēken:

Problem 3.5c Latin civilization perpetuated many aspects of Greek culture. In particular, the metre used by Sophocles and other Greek writers was later taken over and adapted by Latin writers. The following excerpt is from *Medea*, the tragedy by Seneca. The verse structure is identical to that of Sophocles, but one of the feet presents a slight modification. What is the modification and in which foot is it?

Lūcīna, custos, quaeque domitūram fretī (2)
Tiphyn nouam frēnāre docuistī ratem... (3)

In Ancient Greek, *ai*, *oi*, *ei*, *au*, *eu*, *ou* are diphthongs; *p^h*, *t^h* and *k^h* are consonants. *^{h}* before a vowel shows that it is preceded by a puff of air similar to 'h'. Treat sequences of a vowel and *i* as diphthongs, unless the *i* appears as *ï*, in which case it forms its own syllable. The apostrophe *'* marks a vowel that has been deleted (or elided).

In Latin, *qu* is a single consonant pronounced like 'c' followed by 'w'; *ph* and *f* are pronounced the same as each other, *y* is a vowel, *ae* is a diphthong. *u* between two vowels behaves like a consonant.

The mark *ō* above a vowel denotes length.

Problem 3.6

Fijian (Roxana Dincă, RoLO 2015)

Here are some words in Fijian, with the primary and secondary stresses marked:

láko, *paràimarí:*, *tálo*, *βináka*, *kilá:*, *nrè:nré:*, *atómi*, *perèsiténdi*, *mìnisìterí:*, *mbàsikètepólo*, *mbè:léti*, *Seŋái*, *taràusése*, *parò:karámu*, *mì:sìniŋgáni*, *ndàirèkitá:*

Problem 3.6a Mark the primary and secondary stresses in the following words:

mbelembo:tomu, *mbasa:*, *ndikonesi*, *ndoketa:*, *palasita:*, *terenisisita:*

The marks ◌̀ and ◌́ above a vowel mark the primary and secondary stress, respectively. Two consecutive vowels form a diphthong. The mark *:* after a vowel denotes length.

Problem 3.7

Old Norse (Peter Arkadiev & Elena Gurevich, MSK 2006)

A professor of Old Norse philology, while explaining to his students the principles of Old Icelandic versification, invited them to analyse several lines from an Old Icelandic manual on versification written in the 13th century, containing lists of the names of mythological characters and objects, sometimes connected with each other by the word *ok* ('and'). Below are these few lines and the professor's instructions to the students. There are some gaps in the instructions.

A1. *Randverr, Rökkvi, || Reifnir, Leifnir*

A2. *Gaurekr ok Húnn, || Gjúki, Buðli*

A3. *Þórr ok Hildolfr, || Hermóðr, Sigi*

A4. *Byrvill, Kílmundr, || Beimi, Ƒórekr*

A5. *Svalinn ok Randi, || Saurnir, Borði*

A6. *Hildr ok Skeggöld, || Hrund, Geirdriful*

A7. *Viðarr ok Baldr, || Váli ok Heimdallr*

A8. *Frigg ok Freyja, || Fulla ok Snotra*

A9. *Skávær, Skáviðr, || Skirfir, Virfir*

The professor's instructions:

> In Old Icelandic versification, the main poetic technique is alliteration, i.e., the repetition of the initial sounds of words. You need to know the following about Old Icelandic alliteration:
>
> 1. It affects only content words;
> 2. There are four positions in each line that are significant for alliteration, although the number of content words in a line can, in principle, exceed four; two such positions must be in the first half-line and two in the second (the boundary between the half-lines is indicated by the sign /);
> 3. Position __(1)__always alliterates, while position __(2)__never alliterates;
> 4. Out of the remaining positions (__(3)__and __(4)__), one must alliterate (it does not matter which one); sometimes they both alliterate.

Problem 3.7a Fill in the blanks in the professor's instructions.

After the students figured out the above lines, the professor said:

> Now look at another text, which lists the names of various parts of the ship, intended for use in certain kinds of poetic diction. At first glance, it will seem to you that the text contains gross violations of the rules, but in fact, everything is in order. Although all content words in this text __(5)__, for the purposes of alliteration, __(6)__behave as if they were __(7)__ and never alliterate with each other, with __(8)__ or with __(9)__. By the way, pay attention to line A__(10)__, which I showed you earlier.

After some thought, the professor added:

And keep in mind that in line B __(11)__ you are dealing with a rare example of "extra" alliteration, whilst lines B __(12)__ and B __(13)__, despite having more than four content words, do not show any "extra" alliterations.

B1. *Segl, skör, sigla, || sviðvís, stýri*

B2. *sýjur, saumför, || súð ok skautreip*

B3. *stag, stafn, stjórnvið, || stuðill ok sikulgjörð*

B4. *snotra ok sólborð, || sess, skutr ok strengr*

B5. *Söx, stœðingar, || sviptingr ok skaut*

Problem 3.7b Fill in the blanks in the second part of the professor's instructions.

Problem 3.7c Here is one more line from the same text, which the professor absent-mindedly forgot to show the students. Mark the positions of the alliteration in it and explain your choice.

B6. *spíkr, siglutré, || saumr, lokstólpar*

j = 'y' in 'year', *h* = 'h' in 'hat', *þ* = 'th' in 'thin', *ð* = 'th' in 'that'; *ö*, *y*, *æ*, *œ* are vowels. The mark ◌́ above a vowel denotes length.

Problem 3.8

Chickasaw (Saujas Vaduguru, PLO 2019)

Here are some words in Chickasaw and their meanings. The words are given in IPA notation. Both primary and secondary stresses are marked.

tʃoˈkaːˌno	'fly'	*ˌmaʃkoːˌkiʔ*	'(name of a tribe)'
ˌlokˈtʃok	'mud'	*ˌokˌfokˈkol*	'(type of snail)'
ˈaːˌtʃomˌpaʔ	'local store'	*taˈlaːˌnomˌpaʔ*	'telephone'
iˌbiɬˈkan	'snot, mucus'	*ˈnaːɬtoˌkaʔ*	'policeman'
ˌʃanˈtiʔ	'rat'	*aˈboːkoˌʃiʔ*	'river'
ˈsaːɬkoˌna	'earthworm'	*noˌtakˈfa*	'jaw'
ˌtʃonˈkaʃ	'heart'	*tʃiˌkaʃʃaʔ*	'Chickasaw'
faˈlaːt	'crow'	*ˌokˈtʃaːˌlinˌtʃiʔ*	'saviour'

Problem 3.8a If you are given a new Chickasaw word, how would you identify the syllables and determine which syllables get primary stress and which ones get secondary stress?

Problem 3.8b Here are some more words in Chickasaw:

taʔossa:pontaʔ	‘finance company’
ʃimmano:liʔ	‘(name of a tribe)’
kanannak	‘(type of lizard)’
intikba:t	‘sibling’
okta:k	‘prairie’

Mark the primary and secondary stress(es).

The mark *:* after a vowel denotes length. *ʃ*, *ɬ* and *ʔ* are consonants. The marks ◌́ and ◌̩ before a syllable mark the primary and secondary stress, respectively.

Problem 3.9

Ligurian (Kevin Liang, UKLO 2020)

Below are some words written in Ligurian along with their English translations. The stress is indicated by the mark ◌́ above the vowel. *Two of the words have their stress marked on the wrong syllable.*

ɔ:ʒél:i	‘birds’	*sité:*	‘city’
pásta	‘pasta’	*skwád:ra*	‘team’
vjú:vet:a	‘purple’	*damáskina*	‘pear’
nɔ́stru	‘our’	*venín*	‘poison’
du:seméɲte	‘sweetly’	*kutél:u*	‘knife’
kúm:e	‘how’	*mejʒíŋ:a*	‘medicine’
dát:ɔw	‘date (fruit)’	*pe:tená:*	‘to comb’
ba:ʒú	‘kiss’	*májstra*	‘teacher’
pú:vje	‘dust’	*rám:u*	‘copper’
taramɔ́t:u	‘earthquake’	*agýs:u*	‘sharp’
agysá:	‘to sharpen’	*béstja*	‘beast’

Problem 3.9a Identify the two words from the list above in which stress is marked on the wrong syllable and write them with their stress on the correct syllable.

Problem 3.9b Mark the stress in the following words:

bulak:u	'bucket'	*abityd:ine*	'habit'
rystegu	'rustic'	*akɔrdju*	'agreement'
fyrmine	'lightning'	*ɛ:gwa*	'water'

In Ligurian, both consonants and vowels can be long (these are marked with the sign *:* placed after the sound); *ʒ*, *j*, *ŋ* and *w* are consonants; *ɔ*, *y* and *ɛ* are vowels.

3.8 Solutions of practice problems

Solution for practice problem 3.3. Old Javanese

Solution 3.3a

V1: *yêkā rakwa kinon hyaŋ Indra sĕdĕŋ amwit kapwa tâmurṣita*

V2: *tan sora pwa tĕkap nikā daśagunan rūpanya dentânaku*

V3: *liŋ hyaŋ Śakra nahan sinambahakĕn iŋ widyādarī mūr tĕhĕr*

V4: *tan sangkêŋ wiṣaya prayojñananira lwir saŋgrahêŋ lokika*

Solution 3.3b

paramārthapandita ⟶ 4th syllable

ametmetâśrayā ⟶ 11th syllable

santosâhĕlĕtan ⟶ 1st syllable

candanâpāyunan ⟶ 14th syllable

Rules: The structure of the line is as follows:

H.H.H.L.L. H.L.H.L.L. L.H.H.H.L. H.H.L.(H/L)

Heavy syllables are those which contain one of the vowels *â*, *ê*, *e*, *o*, *ā*, *ī*, *ū*, or those that have a coda. Syllabification does not respect word boundaries.

Solution for practice problem 3.4. Chuvash

Solution 3.4a

vĕltrentăr'ri	'tit (bird)'	*'jyvărlăh*	'difficulty'
vişmi'ne	'overmorrow'	*'măkărălčăk*	'convex'
ilĕr'tüllĕ	'tempting'		

Rules: The stress is generally placed on the last syllable. If this contains an extra-short vowel, the stress instead falls on the previous syllable; this is repeated until the syllable does not contain an extra-short vowel. If all vowels are extra-short, the stress is placed on the first syllable.

This rule can be rephrased as follows: stress falls on the rightmost syllable that does not contain an extra-short vowel. If all vowels are extra-short, it falls on the first syllable.

Solution for practice problem 3.5. Ancient Greek

Solution 3.5a Syllabification follows the usual rules: ((VV → V.V, VCV → V.CV, VCCV → VC.CV, VCCCV → VC.CCV). Moreover, based on the footnotes, we know which vowels can form diphthongs.

- A light syllable (L) contains a short vowel and does not have a coda.
- A heavy syllable (H): contains a long vowel OR a diphthong OR has a coda.
- Each verse contains 12 syllables and has the structure:

  X H L H | X H L H | X H L H

  where X represents a syllable which can be either light or heavy.

  Thus, the verse has three metrical feet of the form X H L H.

Solution 3.5b The third verse (Euripides) features the same metre, with the extra condition that syllable X is always heavy.

Solution 3.5c Two consecutive light syllables are equivalent to a heavy one. Thus, the metre of this poem is:

X H L H | X H L L L | X H L H

Moreover, X is always heavy, so the metre can be written as:

H H L H | H H L L L | H H L H

Solution for practice problem 3.6. Fijian

Solution 3.6a

mbèlembò:tómu	*ndìkonési*	*palàsitá:*
mbasá:	*ndòketá:*	*terènisìsitá:*

Solution 3.6b Primary stress is generally placed on the penultimate vowel. However, if the final syllable contains a long vowel (or a diphthong), i.e., if it's a heavy syllable, then the stress falls on it.
For the secondary stresses, the following algorithm is used:

1. Assign secondary stress to all other syllables that contain either a long vowel or a diphthong;
2. Once this has been done, count syllables from right to left. If this procedure reveals a sequence of two unstressed syllables, assign secondary stress to the leftmost of these.

For example, in the word *mbelembo:tomu*, we have the following step-by-step results:

- marking primary stress → *mbelembo:tómu*
- marking secondary stress on all other heavy syllables: *mbelembò:tómu*
- among the remaining stressed syllables, we look for the first group of two consecutive syllables from right to left: ***mbele****mbò:tómu*. The leftmost syllable in this group then receives secondary stress → *mbèlembò:tómu*. This last step is repeated until there are no more pairs of consecutive unstressed syllables.

The rule for secondary stress can be rephrased as follows: all heavy syllables which do not bear primary stress receive secondary stress. For the rest of the syllables, every other syllable from right to left receives secondary stress. The counting is reset when encountering a stress mark.

Solution for practice problem 3.7. Old Norse

Solution 3.7a

(1) third	(3) first
(2) last/fourth	(4) second

Solution 3.7b
(5) start with the same sound (*s*)
(6) the combination of *s* with a voiceless consonant
(7) a single sound
(8) *s*
(9) the combination of *s* with a voiced consonant
(10) 9
(11) 3
(12) 1
(13) 4

Solution 3.7c B6. *spíkr, siglutré, || saumr, lokstólpar*
sp represents a single sound and does not alliterate.

Rules: The positions that alliterate in the verses given are marked in a box. The word *ok* is not a content word (it is written in grey):

A1. *R andverr, R ökkvi, || R eifnir, Leifnir*
A2. *G aurekr ok Húnn, || G júki, Buðli*
A3. *Þórr ok H ildolfr, || H ermóðr, Sigi*
A4. *B yrvill, Kílmundr, || B eimi, Þórekr*
A5. *S valinn ok Randi, || S aurnir, Borði*
A6. *H ildr ok Skeggöld, || H rund, Geirdriful*
A7. *V iðarr ok Baldr, || V áli ok Heimdallr*
A8. *F rigg ok F reyja, || F ulla ok Snotra*
A9. *Sk ávær, Sk áviðr, || Sk irfir, Virfir*

B1. *S egl, skör, s igla, || s viðvís, stýri*
B2. *s ýjur, s aumför, || s úð ok skautreip*
B3. *st ag, st afn, st jórnvið, || st uðill ok sikulgjörð*
B4. *s notra ok s ólborð, || s ess, skutr ok strengr*
B5. *S öx, stœðingar, || s viptingr ok skaut*

Solution for practice problem 3.8. Chickasaw

Solution 3.8a

taˌʔosˈsaːˌponˌtaʔ	‘finance company’
ʃimmaˈnoːˌliʔ	‘(name of a tribe)’
kaˌnanˈnak	‘(type of lizard)’
ˌinˌtikˈbaːt	‘sibling’
ˌokˈtaːk	‘prairie’

Rules: Syllabification follows the usual rules: (VCV → V.CV, VCCV → VC.CV). Note that *tʃ* is a single sound (voiceless post-alveolar affricate).

- Primary stress:
  - is always placed on one of the last three syllables;
  - if one of the syllables has a long vowel, that syllable receives primary stress;
  - otherwise, primary stress is placed on the last syllable.
- Secondary stress:
  - is placed on all syllables that have a coda;
  - if the last syllable does not bear primary stress, it will receive secondary stress, even if it does not have a coda.

Solution for practice problem 3.9. Ligurian

Solution 3.9a *vju:vét:a* and *bá:zu*

Solution 3.9b *bulák:u* *abitýd:ine* *rýstegu* *akɔ́rdju* *fýrmine* *é:gwa*

Rules:

1. Stress always falls before a long consonant.
2. If there are no long consonants, the stress falls on the last heavy syllable (a syllable counts as heavy if it contains either a coda or a long vowel).

The rule can be simplified if we treat long consonants as sequences of identical consonants broken up by a syllable boundary.

Further reading

Ladefoged, Peter & Keith Johnson. 2014. *A course in phonetics*. Wadsworth: Cengage Learning.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Setter, Jane. 2021. *Your voice speaks volumes*. Oxford: Oxford University Press.

4 Phonology

4.1 Introduction

Whilst phonetics, the subject of the previous chapter, is concerned with the physical aspects of sound, phonology is the study of their functions and the ways in which sounds can pattern within the system of the language. Phonology is connected to phonetics in the sense that, once the basic phonetic concepts are understood (characterisation of the sounds), we can begin unveiling the ways in which these sounds interact with one another. It will therefore be useful to introduce a way to succinctly describe the phonological changes that can take place.

4.2 Phonological notation

Note

These notations are meant to make rule writing easier and faster, thus saving time. However, their use is not necessary; if you are unsure of how to use them, it is better to avoid using them in an actual competition, rather than risk using them incorrectly.

(1) $[\ \pm A\ ]$

This shows that the sound we refer to has (+) or does not have (–) the feature A. Thus, we can characterise the sound [m] (bilabial nasal) as

$$\begin{bmatrix} +\text{BILABIAL} \\ +\text{NASAL} \end{bmatrix},$$

and the sound [s] (voiceless alveolar fricative) as

$$\begin{bmatrix} +\text{ALVEOLAR} \\ -\text{VOICE} \\ +\text{FRICATIVE} \end{bmatrix}.$$

Note

In reality, the notation is more complex in the sense that each feature *A* must be binary (i.e., have only two possible values). A feature like voice is binary in that all consonants are either voiced (+VOICE) or voiceless (-VOICE). The parameter fricative is somewhat different: the sounds that are not fricatives (-FRICATIVE) do not necessarily share a common property, e.g., they can be stops, affricates, nasals, and so on. Nevertheless, for the purpose of linguistics problems, there is no need to get into any more details and we can simply write the place and manner of articulation.

(2) $A \longrightarrow B / C \_ V$

This is a phonological rule and is to be read as follows: "the sound A becomes (⟶) B if it appears in a certain environment (/)..." (in this case, if it is between a consonant (C) and a vowel (V)). When writing phonological rules, the underscore (_) shows the position of the sound which is affected. Therefore, if we want to write the rule "*k* becomes *g* before *a*", we write $k \longrightarrow g / \_\ a$.

There are other symbols that can be used to explain the environment in which the transformation takes place. For example, the hash sign (#) marks the word boundary (so _ # represents the end of the word – the sound that changes is before the word boundary, so at the end of the word –, while # _ represents the beginning of the word). As mentioned in the previous chapter, the syllable boundary is marked with a full stop (.). The notation C_0 represents a sequence of consonants (or an absence of a consonant): thus, the string *u*C_0 matches *u*, *u*C, *u*CC, *u*CCC, etc.

In order to denote multiple possibilities, we can use curly brackets. If we want to write the rule "*k* becomes *g* if at the beginning of the word or after a vowel", we can write:

$$\text{k} \rightarrow \text{g} / \left\{ \begin{array}{c} \#\_\_ \\ V\_\_ \end{array} \right\}$$

When writing these rules, we can also use the (square) bracket notation shown above, therefore, if we want to write "*k* becomes *g* after a voiced consonant", we can write:

$$\text{k} \rightarrow \text{g} \,/ \begin{bmatrix} -\text{SYLLABIC} \\ +\text{VOICE} \end{bmatrix} \_\_$$

We need to mention the parameter [-SYLLABIC] (specifying that it is a non-syllabic sound) in order to exclude the vowels.

Moreover, in certain cases, we can use Greek letters (α, β, γ, etc.) instead of $\pm$. They are used in order to show that a certain feature has the same value in multiple parts of the rule, but it could be either a plus or a minus but, importantly, the values must match.

One common use of this notation is to describe assimilation. For example, the following rule:

(3) $$\text{n} \rightarrow \begin{bmatrix} \alpha\ \text{PLACE} \end{bmatrix} / \_\_ \begin{bmatrix} +\text{STOP} \\ \alpha\ \text{PLACE} \end{bmatrix}$$

can be read as: "The consonant *n* before a stop will change its place of articulation to match the place of articulation of the stop." Basically, this rule combines all the following into one:

- $\text{n} \rightarrow \begin{bmatrix} +\text{BILABIAL} \end{bmatrix} / \_\_ \begin{bmatrix} +\text{STOP} \\ +\text{BILABIAL} \end{bmatrix}$
- $\text{n} \rightarrow \begin{bmatrix} +\text{PALATAL} \end{bmatrix} / \_\_ \begin{bmatrix} +\text{STOP} \\ +\text{PALATAL} \end{bmatrix}$
- $\text{n} \rightarrow \begin{bmatrix} +\text{VELAR} \end{bmatrix} / \_\_ \begin{bmatrix} +\text{STOP} \\ +\text{VELAR} \end{bmatrix}$ and so on.

For more examples of how to use alpha notation, see problems 4.3 and 4.4.

4.3 Complementary distribution

Complementary distribution is the relation between two or more sounds, in which each sound can only be found in certain environments (under certain conditions). Some languages can use the same symbol for two or more sounds which have a complementary distribution because, depending on the environment where they are found, there will be no ambiguity regarding the sound it

represents. For example, in Korean, the character ㄹ represents both the sounds *l* and *r*, but there are specific pronunciation rules. This character is pronounced *r* if it is between two vowels, and *l* otherwise. As one can notice, there is a predictable environment (which can be described) and another default environment (for all the other cases). We can write a phonological rule to explain the transformation of the sound in the given context. For example, for Korean, we can write the rule: l ⟶ r / V _ V (reading "the sound *l* becomes *r* if it is between two vowels"). In all cases, the transformation is made starting from the default sound (whose environment is not predictable) towards the predictable sound (so we can then specify the exact environment in which the transformation takes place).

Often, complementary distribution environments depend on the position before/after a vowel or the beginning/end of the word. Thus, these are the first things to check. Of course, there can be more complex cases in which the complementary distribution depends on the features of the previous or following sounds or even sounds further away in the word.

Let us consider the following examples from Spanish. We focus on the sounds *d* and *ð*, which, in these data, are in a complementary distribution:

> *aban[d]onar*, *alcal[d]e*, *[d]ecir*, *[d]oncel*, *[d]on[d]e*, *entra[ð]a*, *la[ð]o*, *me[ð]ir*, *na[ð]a*, *senti[ð]o*

The first step is to split the words into two groups, depending on which sound they contain:

d sound	*ð* sound
aban[d]onar	*entra[ð]a*
alcal[d]e	*la[ð]o*
[d]ecir	*me[ð]ir*
[d]oncel	*na[ð]a*
[d]on[d]e	*senti[ð]o*

We notice that neither of the two sounds appears exclusively at the end or beginning of the word, so we can exclude the possibility that this is the relevant factor. Next, we check if any of the two prefer to be before/after/in between vowels. We notice that both *d* and *ð* appear before vowels, so this is not helpful, but we notice that *ð* is found only after a vowel (while *d* is always either at the beginning of the word, or after a consonant). Thus, since *ð* is the one with

a predictable environment (after a vowel), we can write the phonological rule: d ⟶ ð / V_.

Let us analyse the following examples (for simplicity, they are already split into two columns):

In Kimatuumbi (spoken in Tanzania), sounds *g* and *ɠ* are in a complementary distribution.

g sound	*ɠ* sound
lisɛɛŋgɛlɛ	*ɠɔlɔja*
kjaaŋgi	*nuɠa*
likɔɔŋgwa	*ɠɔlɔja*
ŋgaambalɛ	*ɠɔlɔka*

In this case, we can easily observe that *g* only appears after *ŋ*. Thus, we can write the rule: ɠ ⟶ g / ŋ _.

In a variety of Luganda, the sounds *l* and *r* are in a complementary distribution:

l sound	*r* sound
kola	*beera*
lwana	*jjukira*
lja	*erjato*
luula	*effirimbi*
omugole	*emmeeri*
lumonde	*eraddu*
oluganda	*wawaabira*

This time, we notice that both *l* and *r* appear solely next to vowels (or *j*). Moreover, their environment does not seem to be connected to the beginning or end of the word, so the next step is to look at the type of sounds next to which they appear. We notice that *l* appears only at the beginning of the word or after the vowel *o*, while *r* appears only after the vowels *i* and *e*. Therefore, based on this observation, we can write two possible phonological rules: l ⟶ r / {e, i} _ or r ⟶ l / {# _, o _}. This is the moment when it is important to understand the phonetic concepts. Generally, phonological rules are conditioned by a particular feature or features of sounds in the environment where the alternating sounds

appear. Thus, we notice that both *i* and *e* are front vowels (and the data do not contain any other front vowels), so the rule becomes: l ⟶ r / [+ FRONT] _. Generally, if an environment depends on more sounds (in this case *i* and *e*), there will be a connection between these sounds (they will share a certain feature) and, in most cases, in linguistics problems, there will be a task asking you to extend this rule to other sounds. For example, for this problem, it is possible that one of the tasks would feature another front vowel, and you will need to deduce that after that vowel the sound *r* will be used.

Problem 4.1

Irish (Anton Kukhto, Elementy)

Here are some Irish verb forms for the imperative and present indicative, as well as their English translations:

Imperative		Present indicative	
fan	'Stay!'	*fanaim*	'I stay'
cuir	'Put!'	*cuireann sé*	'he puts'
ceannaigh	'Buy!'	*ceannaíonn tú*	'you$_{SG}$ buy'
creid	'Believe!'	*creidim*	'I believe'
críochnaigh	'End!'	*críochnaíonn sé*	'he ends'
déan	'Do!'	*déanann sí*	'she does'
smaoinigh	'Think!'	*smaoiníonn sibh*	'you$_{PL}$ think'
ól	'Drink!'	*ólann sé*	'he drinks'
oibrigh	'Work!'	*oibríonn siad*	'they work'
fág	'Leave!'	*fágann muid*	'we leave'
éirigh	'Raise!'	*éiríonn sí*	'she raises'
lig	'Let!'	*ligeann tú*	'you$_{SG}$ let'
tosaigh	'Start!'	*tosaím*	'I start'
ith	'Eat!'	*itheann sé*	'he eats'

Problem 4.1a Translate into Irish:

'you$_{SG}$ believe'	'I work'	'I think'
'you$_{SG}$ stay'	'I put'	'you$_{SG}$ start'
'I end'	'I drink'	

Solution

The first step in this type of problems is to classify the different forms based on their characteristics. In this case, we can classify the forms in the right column based on person. Nevertheless, we notice that all forms, except for those in the 1st person singular (1SG), end in *nn* and are followed by the subject pronoun, while forms in the 1SG are composed of one word (do not include the subject pronoun).

We can start with the 1SG forms.

Imperative		Present indicative	
fan	'Stay!'	*fanaim*	'I stay'
creid	'Believe!'	*creidim*	'I believe'
tosaigh	'Start!'	*tosaím*	'I start'

We observe that there are three ways in which we can construct the 1SG form: adding the suffix *-aim*, adding the suffix *-im* or replacing the ending *-igh* with *-ím*.

Similarly, we will want to observe how the other verb forms are formed. We ignore the subject pronouns added in the end but focus on the verb form. We get:

- 2SG: add the suffix *-eann* or replace *-igh* ⟶ *-íonn*;
- 3SG, masc.: add suffixes *-eann*, *-ann* or replace *-igh* ⟶ *-íonn*;
- 3SG, fem.: add suffix *-ann* or replace *-igh* ⟶ *-íonn*;
- 1PL: add suffix *-ann*;
- 2PL: replace *-igh* ⟶ *-íonn*;
- 3PL: replace *-igh* ⟶ *-íonn*;

We notice that all these verb forms are obtained through the same three transformations, independent of person: adding the suffixes *-eann*, *-ann* or replacing *-igh* ⟶ *-íonn*. Therefore, most likely, in this problem, there are only two verb forms: the form for 1SG and the form for all the others (in which case, to avoid ambiguity, the subject pronoun is added after the verb). This is also supported by the way in which the task is phrased since all the verb forms that we need to translate are only 1SG or 2SG.

Moreover, just like in the case of 1SG, there are three possible transformations. Therefore, we can assume that Irish verbs can be classified into three different categories, each of them having its own way of expressing these forms. We can easily figure out that the verbs which end in *-igh* in their imperative form the 1SG present indicative by replacing *-igh* with *-ím* and the other forms by replacing *-igh* with *-íonn*. We are left to discover the environment for the other two transformations. For this, we classify the verbs into two categories, based on the way they form their non-1SG form:

-eann	*-ann*
cuir 'to put'	*déan* 'to do'
lig 'to eat'	*ól* 'to drink'
ith 'to let'	*fág* 'to leave'

The classification of these forms seems to not take into account any semantic features (related to meaning) since there seems to be nothing in common between the verbs {'to put', 'to eat', 'to let'} compared with {'to do', 'to drink', 'to leave'}. Therefore, most likely there are some phonetic characteristics (related to the form of the word) that are relevant.

Probably, at first glance, we would be tempted to consider that the ending *-ann* is used for verbs that contain a vowel marked with the acute accent (in the Irish spelling system, the accent marks vowel length, although you cannot know this from the problem). Nevertheless, this rule would not apply to the 1SG form, since the form *-im* appears with the verb *creid*, and the form *-aim* appears with the verb *fan*, none of them having an accent. Looking closely at the verbs and knowing that this distinction must also be present between the verbs *fan* and *creid*, we notice that in all verbs that receive the suffix *-eann*, the last vowel before the consonant is *i*. Therefore, the verbs that have the last vowel *i* (and which do not end in *-igh*) form their 1SG by adding the suffix *-im* and their non-1SG by adding the suffix *-eann*, while the other verbs use the suffixes *-aim* and *-ann*, respectively. Therefore, we have all the information needed to solve the task.

Problem 4.1a

'you$_{SG}$ believe' = *creideann tú*	'you$_{SG}$ stay' = *fanann tú*
'I end' = *críochnaím*	'I work' = *oibrím*
'I put' = *cuirim*	'I drink' = *ólaim*
'I think' = *smaoiním*	'you$_{SG}$ start' = *tosaíonn tú*

Rules:

1. If the imperative form ends in *-igh*:
   - 1SG: *-igh* ⟶ *-ím*;
   - non-1SG: *-igh* ⟶ *-íonn*;
2. else:
   a) If last vowel is *i*:
      - 1SG: *-im*;
      - non-1SG: *-eann*;
   b) Else:
      - 1SG: *-aim*;
      - non-1SG: *-ann*;

For non-1SG, the verb is followed by the corresponding subject pronoun: 2SG = *tú*, 3SG, masc. = *sé*, 3SG, fem. = *sí*, 1PL = *muid*, 2PL = *sibh*, 3PL, masc. = *siad*.

Problem 4.2

Roro (Vladimir I. Belikov, MSK 1991)

Here are 11 words in the four dialects of the Roro language and their English translations:

Dialect				
Hisiu	Delena	Kivori	Paitana	Translation
__(1)__	*aitau*	__(2)__	__(3)__	‘three’
aihi	*aisi*	*aihi*	*aisi*	‘crab’
cici	*sisi*	*čiči*	*cici*	‘meat’
ebeoahi	*ebeoasi*	*ebeoahi*	*ebeoaci*	‘he ran’
hiabu	*siabu*	*hiabu*	*ciabu*	‘smoke’
nihe	*nite*	*nihe*	*nite*	‘tooth’
icu	__(4)__	__(5)__	*icu*	‘nose’
maciu	__(6)__	__(7)__	__(8)__	‘tree’
moihana	*moitana*	*moihana*	__(9)__	‘look!’
mahi	__(10)__	__(11)__	*maci*	‘beast’
cubu	*subu*	*čubu*	*cubu*	‘grass’

Problem 4.2a Fill in the blanks.

Solution

In linguistics problems featuring related languages or dialects, the first step is usually figuring out which sounds are different across the data and which remain the same. We start by writing the words for which all four forms are given:

Dialect				
Hisiu	Delena	Kivori	Paitana	Translation
aihi	*aisi*	*aihi*	*aisi*	‘crab’
cici	*sisi*	*čiči*	*cici*	‘meat’
ebeoahi	*ebeoasi*	*ebeoahi*	*ebeoaci*	‘he ran’
hiabu	*siabu*	*hiabu*	*ciabu*	‘smoke’
nihe	*nite*	*nihe*	*nite*	‘tooth’
cubu	*subu*	*čubu*	*cubu*	‘grass’

We should notice that there are some words, like ‘crab’ and ‘he ran’, that have *h* in Hisiu and Kivori, *s* in Delena, and *c* in Paitana. In others, like ‘tooth’, Hisiu and Kivori *h* correspond to *t* in Delena and Paitana.

Summing up, we can write the following correspondences, numbered for convenience:

	Hisiu	Delena	Kivori	Paitana
(1)	*h*	*s*	*h*	*c*
(2)	*c*	*s*	*č*	*c*
(3)	*h*	*t*	*h*	*t*

In other problems of this type, we would need to find the environment in which *h* becomes *s* or *t* in Delena (and *c* or *t* in Paitana), but, before doing this, it is important to check the tasks and see whether we need this type of generalisation.

Dialect				
Hisiu	Delena	Kivori	Paitana	Translation
__(1)__	*aitau*	__(2)__	__(3)__	‘three’
icu	__(4)__	__(5)__	*icu*	‘nose’
maciu	__(6)__	__(7)__	__(8)__	‘tree’
moihana	*moitana*	*moihana*	__(9)__	‘look!’
mahi	__(10)__	__(11)__	*maci*	‘beast’

On the first row, the only Delena consonant that undergoes any transformation is *t* and, according to the rules above, there is only one transformation that yields the sound *t* in Delena (rule 3). Similarly, for tasks 4–8 there is only one possible transformation of the sound *c* in Hisiu (rule 2). The issue emerges for tasks 9–11 where we are given Hisiu words that contain the sound *h*. Nevertheless, for task 9 we know that the sound *h* in Hisiu becomes *t* in Delena (thus, we know that this is rule 3). As for tasks 10–11, we notice that the Hisiu *h* becomes *c* in Paitana (thus following rule 1). Now we have enough information to solve the tasks and there is no need to identify the environments in which each transformation takes place.

Solution 4.2a

Dialect				
Hisiu	Delena	Kivori	Paitana	Translation
aihau	*aitau*	*aihau*	*aitau*	‘three’
icu	*isu*	*iču*	*icu*	‘nose’
maciu	*masiu*	*mačiu*	*maciu*	‘tree’
moihana	*moitana*	*moihana*	*moitana*	‘look!’
mahi	*masi*	*mahi*	*maci*	‘beast’

Rules:

Hisiu	Delena	Kivori	Paitana
h	*s*	*h*	*c*
c	*s*	*č*	*c*
h	*t*	*h*	*t*

4.4 Phonological processes

Below we will present and discuss some of the most common phonological changes. These can be classified into four groups, depending on the effect they have:

1. Deletion of a sound. When the deleted sound is a vowel, you might encounter some more specific terms, such as:
   a) Aphaeresis: deletion at the beginning of the word;
   b) Syncope: deletion in the middle of the word, especially when another syllable follows the deleted vowel;

c) Apocope: deletion at the end of the word.

In linguistics problems, we can use, for simplicity, the term deletion, as long as we specify where it takes place and in which environment. One common kind of deletion occurs when two identical sounds come in contact with one another. For example, in Ainu, the suffix *-re* becomes *-e* if it is added to a word that already ends in *-r*. Thus, an *r* is deleted in order to avoid two consecutive identical sounds. We can write: $r \longrightarrow \emptyset \ / \ r \_ \ e$.

2. Insertion of a sound, also known as epenthesis.

   In linguistics problems, epenthesis often occurs between two consonants or two vowels, as well as word-initially (this particular type is sometimes called prosthesis). Thus, if a stem that ends in a consonant adds a suffix that starts with a consonant, it is possible to add an epenthetic vowel (and avoid two consecutive consonants). For example, most English nouns form the plural with the addition of a single consonant (*cats*, *dogs*), but in words like *horses* there is an extra vowel before the plural marker.

   The same thing may happen between two vowels, as in British English varieties that have so-called linking and intrusive *r* (an example of intrusive *r* is *drawing* ⟶ *draw-**r**-ing*).

3. Metathesis = process that causes the transposition of two or more sounds.

   For example, the word *third* suffered a metathesis from Old English (*th**ri**dda*), in which the sounds *i* and *r* switched places. Although in general, the sounds that switch places are next to one another, there are also cases when metathesis can occur at a considerable distance: for instance, in the Tertenia dialect of Sardinian the word 'belly' occurs as *br**ɛ**nti* in isolation but as *(b)ɛnt**r**i* after the definite article *sa*.

   Another example is the Romanian word *înt**reg*** ('whole'). It comes from the Latin word *int**eg**rum* in which the sound *r* and the sequence *eg* switched places.

4. The last and most important category is that of processes in which one sound is transformed because of another sound in its vicinity. These processes are generally called assimilations. Assimilations can be classified based on different criteria as follows:

   a) assimilation can affect both vowels and consonants;

   b) based on the degree of assimilation we can have total assimilation (in which the target sound becomes completely identical to the trigger) or partial assimilation (in which only certain features are changed);

c) based on the position of the sound: contact assimilation (the two sounds are next to one another) or assimilation at a distance. Assimilation at a distance is often referred to as *harmony*; vowel harmony is quite common, while consonant harmony exists, but is relatively rare;

d) based on the direction of the assimilation: progressive (in which the sound after is changed as a result of its interaction with a previous sound) or regressive (in which the sound that changes occurs before the sound that triggers the assimilation).

One of the most common examples of assimilation is the partial assimilation of nasals in consonant clusters, as exemplified before. Nasal consonants are extremely prone to assimilation, and they often assimilate to the place of articulation of the neighbouring consonant. This phenomenon also occurs in English where, for example, *can be* is pronounced [kæmbi] in fast speech, in which case the nasal *n* "borrows" the place of articulation from the following consonant (*b*, bilabial) and becomes *m*.

Problem 4.3

Behaviour of nasal consonants (Tom McCoy, NACLO 2018)

Problem 4.3a Below are some Indonesian verbs in their active and passive forms and their English translations. Fill in the blanks.

Active	Passive	Translation
meŋuji	*diuji*	'to test'
meɲeja	*dieja*	'to spell'
meŋgaruk	*digaruk*	'to scratch'
mendapat	*didapat*	'to obtain'
memberi	*diberi*	'to give'
menulis	*ditulis*	'to write'
memutus	*diputus*	'to cut off'
__(1)__	*dibuat*	'to make'
__(2)__	*dipilih*	'to choose'

ŋ = 'ng' in 'king'.

Problem 4.3b Below are some Mandar words in their active and passive forms, and their English translations. Fill in the blanks.

Active	Passive	Translation
mambatta	*dibatta*	'to split'
mandeŋŋeq	*dideŋŋeq*	'to carry on the back'
maŋidaŋ	*diidaŋ*	'to crave'
mappasuŋ	*dipasuŋ*	'to send out'
mattunu	*ditunu*	'to burn'
massiraq	*disiraq*	'to tie'
__(3)__	*ditimbe*	'to throw'
__(4)__	*dipande*	'to feed'

ŋ = 'ng' in 'king'.

Problem 4.3c Below are some words in Quechua in their nominative, genitive and locative (preposition 'in') form, as well as their English translations. Fill in the blanks.

Nominative	Genitive	Locative	Translation
kam	*kamba*		'you$_{SG}$'
atam		*atambi*	'frog'
hatum	__(5)__	__(6)__	'the big one'
sinik	*sinikpa*		'porcupine'
čilis	*čilispa*		'streamless region'
sača		*sačapi*	'jungle'
punǰa		*punǰapi*	'day'

č = 'ch' in 'chop'.

Problem 4.3d Given below are some words in Zoque in their base forms and their 1SG possessive ('my'), as well as their English translations. Fill in the blanks.

Base	Possessed	Translation
burru	*mburru*	‘donkey’
pama	*mbama*	‘clothing’
tatah	*ndatah*	‘father’
faha	*faha*	‘belt’
sis	*sis*	‘meat’
flawta	__(7)__	‘harmonica’
šapun	*šapun*	‘soap’
disko	__(8)__	‘phonograph record’
kayu	*ŋgayu*	‘horse’
kopak	__(9)__	‘head’

ŋ = ‘ng’ in ‘king’, *š* = ‘sh’ in ‘shop’.

Problem 4.3e Below are given some nouns in Lunyole in their singular and plural forms, as well as their English translations. Fill in the blanks.

Singular	Plural	Translation
oludaalo	*endaalo*	‘day’
oluboyooboyo	*emboyooboyo*	‘hullabaloo’
olufudu	*efudu*	‘rainbow’
olukalala	*ekalala*	‘list’
olusosi	__(10)__	‘mountain’
olubafu	__(11)__	‘rib’
olupagi	__(12)__	‘spoke (of a bike)’
olutambi	__(13)__	‘candle’

Problem 4.3f All five of the languages in this problem display processes that avoid a specific type of sound combination. Fill in the blanks to describe this generalisation. The blanks should be chosen from: *vowel*, *consonant*, *nasal*, *voiced consonant*, *voiceless consonant*.

Avoid having a __(14)__directly followed by a __(15)__.

Solution

Solution 4.3a We notice that the active form is formed from the passive form by replacing the prefix *di* with one of the prefixes: *meŋ*, *men*, or *mem*. Moreover, we notice that, in certain situations, the first vowel of the root is dropped. We split the given words based on these transformations:

	meŋ	*men*	*mem*
Unchanged stem	*-uji* *-eja* *-garuk*	*-dapat*	*-beri*
Stem drops first consonant		*-tulis*	*-putus*

Looking at the three types of prefix (which only differ by the type of nasal consonant), we expect a nasal assimilation. Thus we notice, indeed, that the place of articulation of the final nasal of the prefix assimilates to the first consonant of the stem. If the stem begins with a vowel, the nasal used is *ŋ*. Moreover, we notice that if the stem begins with a voiceless stop, it gets dropped. Hence, the blanks are (1) *membuat* and (2) *memilih*.

We can write the rules in different ways:

- using words: *di* ⟶ *meN* (*N* is a nasal): *N* assimilates to the place of articulation of the following consonant. If the following sound is a vowel, use *ŋ*. If the root starts with a voiceless stop, the stop is deleted.
- using phonological notation:
  - di → meŋ / #__V , and
  - $\text{di} \rightarrow \text{me}\begin{bmatrix} \alpha\ \text{PLACE} \\ +\text{NASAL} \end{bmatrix} / \#\_\_\begin{bmatrix} \alpha\ \text{PLACE} \\ +\text{VOICE} \\ \text{STOP} \end{bmatrix}$, and
  - $\text{di}\begin{bmatrix} \alpha\ \text{PLACE} \\ -\text{VOICE} \\ +\text{STOP} \end{bmatrix} \rightarrow \text{me}\begin{bmatrix} \alpha\ \text{PLACE} \\ +\text{NASAL} \end{bmatrix} / \#\_\_$

These rules can also be written as a process which takes place in three steps:

Step 1: di → meŋ

Step 2: $\text{ŋ} \rightarrow \begin{bmatrix} \alpha \text{ PLACE} \\ +\text{NASAL} \end{bmatrix} / \text{me}\_\_ \begin{bmatrix} \alpha \text{ PLACE} \\ +\text{STOP} \end{bmatrix}$

Step 3: $\text{me}[\ +\text{NASAL}\ ][\ -\text{VOICE}\ ] \rightarrow \text{me}[\ +\text{NASAL}\ ] / \#\_\_$

In this case, the three steps are: (1) replacing the prefix *di* with *meŋ*, (2) the assimilation of the nasal *ŋ* to the following consonant, and (3) elision of this consonant, if it is voiceless.

Solution 4.3b We notice this is a similar process in which the prefix *di* is replaced by one of the prefixes *mam*, *man*, *maŋ*, *map*, *mat*, *mas*. Thus, we notice that we have two types of prefix: *maN* (where *N* is a nasal and hence, we expect it to partially assimilate to the place of articulation, as happened above) and *maX* (where *X* represents the consonant that follows, thus being a total assimilation).

mam	*man*	*maŋ*	*maX*
			-pasuŋ
-batta	*-deŋŋeq*	*-idaŋ*	*-tunu*
			-siraq

Therefore, as above, there is an assimilation of the nasal consonant if the stem begins with a voiced stop or with a vowel (if it starts with a vowel, the nasal used is *ŋ*). Moreover, we notice that the total assimilation happens if the stem begins with a voiceless stop or with a non-stop consonant (e.g., fricative).

Note: Based on the data given, we can generalise that the prefix *maX* occurs if the stem begins with a voiceless consonant. Therefore, the answers are (3) = *mattimbe* and (4) = *mappande*, and the rules are:

- using words: *di* ⟶ *maX*
  - if the stem starts with a voiceless consonant, *X* is identical to the first consonant;
  - if stem starts with a vowel, *X* = *ŋ*;
  - if stem starts with a voiced consonant, *X* is a nasal with the same place of articulation as the first consonant.

- using phonological rules:
  - di → maŋ / #__V
  - $\text{di} \rightarrow \text{ma}\left[\begin{array}{c}\alpha\ \text{PLACE}\\ +\text{NASAL}\end{array}\right] / \#\_\_\left[\begin{array}{c}\alpha\ \text{PLACE}\\ +\text{VOICE}\\ +\text{STOP}\end{array}\right]$
  - $\text{di} \rightarrow \text{maC} / \#\_\_\text{C}$ if $C = \left[\ -\text{VOICE}\ \right]$

Solution 4.3c The first observation is that the genitive is formed by adding the suffixes *-pa* or *-ba*, and the locative is formed with the suffixes *-pi* or *-bi*. Therefore, we are only interested in the choice of the consonant (*p* or *b*).

p	*b*
sinik	
čilis	*kam*
sača	*atam*
punǰa	

Based on these examples, there are multiple possible rules, such as:

- use *b* if the stem ends in *m*;
- use *b* if the stem ends in a nasal;
- use *b* if the stem ends in a voiced consonant (remembering that nasals are voiced);

Since until now the main phenomenon of this problem has been the behaviour of nasals, we choose the second rule, so the answers are (5) = *hatumba* and (6) = *hatumbi* and the rules are:

locative: *-pi*, genitive: *-pa* $\qquad \text{p} \rightarrow \text{b} / \left[\ +\text{NASAL}\ \right]\_\_$

Solution 4.3d This time we notice that the possessive can be marked by a nasal added as a prefix (*m*, *n*, *ŋ*) or it can be unmarked (∅). Moreover, the first consonant in the stem can change. For choosing the prefix, we have:

m	*n*	*ŋ*	∅
burru *pama*	*tatah*	*kayu*	*faha* *sis* *šapun*

Thus, we notice that we add a nasal (which assimilates to the place of articulation) if the stem begins with a stop; the possessive is unmarked if the stem starts with a fricative.

Regarding the change in the root, we notice: ***p**ama* → *m**b**ama* and ***t**atah* → *n**d**atah*. Therefore, if the nasal is added as a prefix, the following consonant will become voiced. Thus, the answers are: (7) = *flawta*, (8) = *ndisko*, (9) = *ŋgopak*. The rules are:

- if the stem begins with a stop, add a nasal with the same place of articulation before it. Moreover, if the stop at the beginning of the stem is voiceless, it will become voiced.

  This can also be written using phonological notation as:

$$\begin{bmatrix} \alpha \text{ PLACE} \\ +\text{STOP} \end{bmatrix} \rightarrow \begin{bmatrix} \alpha \text{ PLACE} \\ +\text{NASAL} \end{bmatrix} \begin{bmatrix} \alpha \text{ PLACE} \\ +\text{STOP} \\ +\text{VOICE} \end{bmatrix} / \#\_\_$$

- if the stem starts with a fricative, no prefix is added.

Solution 4.3e For the last part of the problem, we notice that the singular always begins with *olu-* and the plural prefix can be: *e*, *en*, *em*.

e	*em*	*en*
-fudu *-kalala*	*-boyooboyo*	*-daalo*

We notice that this is a process similar to the one in task (b), where the prefix gets a nasal consonant (which assimilates to the place of articulation) if the stem begins with a voiced stop; otherwise, it only gets the prefix *e-*. Thus, the answers are: (10) = *esosi*, (11) = *embafu*, (12) = *epagi*, (13) = *etambi* and the rules are:

$$\text{olu} \rightarrow \text{e}\left[\begin{array}{c}\alpha\ \text{PLACE}\\ +\text{NASAL}\end{array}\right] / \_\_\left[\begin{array}{c}\alpha\ \text{PLACE}\\ +\text{STOP}\\ +\text{VOICE}\end{array}\right], \text{and}$$

$$\text{olu} \rightarrow \text{e} / \_\_\left[\ -\text{VOICE}\ \right]$$

Solution 4.3f Finally, this task helps us understand the core reason for these transformations. The rule is: "Avoid having a nasal directly followed by a voiceless consonant." Each of the five languages has its own way to deal with that. As such, in task (a) the voiceless stop is deleted; in task (b) the nasal consonant is fully assimilated; in tasks (c) and (d) the voiceless consonant becomes voiced (it assimilates to the voicing of the nasal); in task (e) the nasal consonant is deleted.

4.5 Vowel harmony

A special type of assimilation is VOWEL HARMONY. This process is common to all Turkic languages and dictates the way in which the affixes change form depending on the word to which they are added. Let us consider the case of Turkish. The Turkish language has eight vowels: *a*, *e*, *i*, *o*, *u*, *ö*, *ü*, *ı* (they correspond to [a], [e], [i], [o], [u], [ø], [y], and [ɨ], respectively in IPA), which can be classified based on the three core features (backness, height, roundness):

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
Close	*i*	*ü*	*ı*	*u*
Open	*e*	*ö*	*a*	*o*

Turkish has two types of vowel harmony, involving two different vowel features, i.e., backness and rounding.

The first dictates the assimilation of backness to the added suffix. For example, the plural suffix in Turkish is *-lar* or *-ler*. The suffix *-lar* is used when the word ends in a back vowel, while *-ler* is used if the word ends in a front vowel (*-lar* and *-ler* are called allomorphs, which will be further discussed in the next chapter). Therefore, we have the following pairs of words: *baba* – *baba**lar***, *okul* – *okul**lar***, but *kedi* – *kedi**ler***, *ev* – *ev**ler***. Another suffix that follows this type of vowel harmony is the locative suffix ('at'/'in'): *-da* / *-de*.

The second type of vowel harmony dictates the assimilation of roundness but also takes into account the closeness of the vowels. The possessive suffix for 1SG ('my') in Turkish has four allomorphs:[1] *-im*, *-ım*, *-um*, *-üm*. Similarly, the choice of suffix depends on the last vowel of the word, as follows:

- if the last vowel is front unrounded, use the form *-im*;
- if the last vowel is back unrounded, use the form *-ım*;
- if the last vowel is front rounded, use the form *-üm*;
- if the last vowel is back rounded, use the form *-um*.

Briefly, we can also represent this harmony as follows:

$$\{a, ı\} \longrightarrow ı \qquad \{e, i\} \longrightarrow i \qquad \{o, u\} \longrightarrow u \qquad \{ö, ü\} \longrightarrow ü$$

Therefore, if the last vowel is *a* or *ı*, use the vowel *ı*; if the last vowel is *o* or *u*, use the vowel *u*, and so on. We have the following pairs: *ev – ev**im***, *hortum – hortum**um***, *raf – raf**ım***, *göz – göz**üm***.

Moreover, notice that in Turkish all the vowels in a word have the same backness (the word only contains either front or back vowels). While this is a general trend, there are also exceptions to this rule, especially for words borrowed into Turkish from other languages.

Problem 4.4

Valley Yokuts (Paul Helmer, RoLO 2019)

Here are some verbal forms in Valley Yokuts in four different forms and their English translations:

Dubitative	Passive voice	Non-future	Imperative	Translation
do:sol	__(1)__	*doshin*	__(2)__	'to report'
__(3)__	*dubut*	*dubhun*	*dubk'a*	'to conduct'
yawa:lal	*yawa:lit*	*yawalhin*	*yawalk'a*	'to follow'
logwol	*logwit*	*logiwhin*	*logiwk'a*	'to pulverise'
wo:nol	*wo:nit*	*wonhin*	__(4)__	'to hide'

[1] In reality, there is also a fifth form, *-m*, for words ending in a vowel.

Dubitative	Passive voice	Non-future	Imperative	Translation
xatal	*xatit*	*xathin*	__(5)__	‘to eat’
__(6)__	__(7)__	__(8)__	*t’oyixk’a*	‘to treat’
__(9)__	*ʔopo:tit*	*ʔopothin*	*ʔopotk’o*	‘to get out of bed’
ʔugnal	__(10)__	*ʔugunhun*	__(11)__	‘to drink’
__(12)__	__(13)__	__(14)__	*ʔilikk’a*	‘to sing’
__(15)__	*lihmit*	*lihimhin*	__(16)__	‘to run’
__(17)__	*luk’lut*	__(18)__	__(19)__	‘to bury’
__(20)__	*k’oʔit*	__(21)__	*k’oʔk’o*	‘to throw’
me:k’al	__(22)__	__(23)__	__(24)__	‘to swallow’

Problem 4.4a Fill in the blanks. If you believe that some blanks could allow multiple answers, write them all.

k’, *t’*, *x*, *y*, *ʔ* are consonants. The mark *:* after a vowel denotes length.

Solution

We begin by segmenting the forms in order to figure out which part is the stem and which are the morphemes corresponding to the four verb forms. We notice that, for two of the verbs, we are given all four forms:

Dubitative	Passive voice	Non-future	Imperative	Translation
yawa:lal	*yawa:lit*	*yawalhin*	*yawalk’a*	‘to follow’
logwol	*logwit*	*logiwhin*	*logiwk’a*	‘to pulverise’

Comparing these forms, we deduce that the dubitative is formed by adding the suffixes *-al* or *-ol*, the passive voice is formed using the suffix *-it*, the non-future using *-hin* and the imperative using *-k’a*. Moreover, we notice that the stems can undergo some changes (for the verb ‘to follow’ there are two possible stems *yawa:l* and *yawal*, while for the verb ‘to pulverise’ there is *logw* and *logiw*). Moreover, we notice that, in both cases, the dubitative and passive voice use the same stem, while the other two forms use the “modified” stem.

Looking at the other words in the corpus, we notice that each of the four forms has two possible suffixes: *-al* and *-ol* for dubitative, *-it* and *-ut* for passive voice, *-hin* and *-hun* for non-future, and *-k'a* and *-k'o* for imperative.

Since the passive voice is the one with the most examples, we can start with it. We make a table in which we split the words into two classes, based on the choice of the suffix used in forming the passive voice.

-it		*-ut*	
yawa:lit	'to follow'	*dubut*	'to conduct'
logwit	'to pulverise'	*luk'lut*	'to bury'
wo:nit	'to hide'		
xatit	'to eat'		
ʔopo:tit	'to get out of bed'		
lihmit	'to run'		
k'oʔit	'to throw'		

We can easily see that the suffix *-ut* is used for the verbs whose stem contains *u*, so this can be viewed as vowel harmony, being a total assimilation of *i* to *u*. This phenomenon can also be written as a phonological rule as follows:

$$\text{it} \rightarrow \text{ut} \;/\; \text{u}\text{C}_0\_\_\#$$

Since the suffixes corresponding to non-future also feature the vowels *i* and *u*, we expect them to be chosen based on the same rule (or a very similar rule). Indeed, analysing the given examples, we notice that *-hun* is used if the last vowel of the stem is *u*, while *-hin* is used otherwise.

We use the same process for the other two verb forms, the dubitative and the imperative, by making a table for each of them. Moreover, considering that for passive and non-future, the rule was purely phonological, not semantic (it does not depend on the meaning of the word, but rather on the form of the word), in the following table we do not include the translation of the verbs.

Dubitative		Imperative	
-al	*-ol*	*-k'a*	*-k'o*
yawa:lal	*do:sol*	*dubk'a*	*ʔopotk'o*
xatal	*logwol*	*yawalk'a*	*k'oʔk'o*
ʔugnal	*wo:nol*	*logiwk'a*	
me:k'al		*t'oyixk'a*	
		ʔilikk'a	

We notice this is a similar phenomenon. The allomorph containing *o* of the dubitative and imperative suffixes (*-ol* and *-k'o*, respectively) is used if the last vowel of the stem is *o*, so it is a total assimilation of the vowel *a* (or a vowel harmony).

Thus, we can sum up the information we have so far in the following table:

Last vowel	Dubitative	Passive voice	Non-future	Imperative
o	*-ol*	*-it*	*-hin*	*-k'o*
u	*-al*	*-ut*	*-hun*	*-k'a*
else	*-al*	*-it*	*-hin*	*-k'a*

Alternatively, we can consider the four basic suffixes: *-al*, *-it*, *-hin*, *-k'a* and the following phonological rules:

• a → o / oC$_0$__ • i → u / uC$_0$__

We are left to discover the way in which the stem changes. We noticed at the beginning of the solution that we have three situations: (1) the stem is unchanged for all forms, (2) the stem for dubitative and passive has a long vowel, which becomes short for non-future and imperative, (3) there is a vowel added (epenthesis) for non-future and imperative.

We make a new table with these three situations:

Unchanged	*V:* ⟶ *V*	Epenthesis
dub	*do:s*	*logw* ⟶ *logiw*
xat	*yawa:l*	*ʔugn* ⟶ *ʔugun*
k'oʔ	*wo:n*	*lihm* ⟶ *lihim*
	ʔopo:t	

We notice that we have three outcomes, based on the last syllable of the stem:

- If the last syllable contains a short vowel and ends in a consonant (CVC), then the stem is left unchanged;
- If the last syllable contains a long vowel and ends in a consonant (CV:C), then the vowel becomes short for non-future and imperative (CVC);
- If the last syllable contains a short vowel and ends in a consonant cluster (CVCC), then in non-future and imperative an (epenthetic) vowel is added

between the two consonants. From the examples above, we notice that the epenthetic vowel is either *i* or *u*, so we can expect that the choice of vowel will be the same as above (*u* if the last vowel in the stem is *u*, otherwise *i*).

Therefore, we have discovered all the rules and we can complete all the tasks.

Solution 4.4a

1. *do:sit*	2. *dosk'o*	3. *dubal*	4. *wonk'o*
5. *xatk'a*			8. *t'oyixhin*
9. *ʔopo:tol*	10. *ʔugnut*	11. *ʔugunk'a*	
	14. *ʔilikhin*	15. *lihmal*	16. *lihimk'a*
17. *luk'lal*	18. *luk'ulhun*	19. *luk'ulk'a*	20. *k'oʔol*
21. *k'oʔhin*	22. *me:k'it*	23. *mek'hin*	24. *mek'k'a*

For the blanks 6–7 and 12–13, we have multiple options since we do not know the underlying form of the stem. The stems of the imperative form (*t'oyix* and *ʔilik*) could be the result of three different processes:

1. We could consider an epenthetic process which causes the insertion of the vowel *i*, therefore the underlying stems are *t'oyx* and *ʔilk*. In this case, the answers are:

   6. *t'oyxol* 7. *t'oyxit* 12. *ʔilkal* 13. *ʔilki*

2. We could also consider the case in which the stem undergoes no change, therefore the underlying stems are *t'oyix* and *ʔilik*. The answers would then be:

   6. *t'oyixal* 7. *t'oyixit* 12. *ʔilikal* 13. *ʔiliki*

3. Lastly, the stem might have undergone vowel shortening. We can infer based on the given data that the long vowel is always the last vowel of the stem. Thus, the underlying stems are *t'oyi:x* and *ʔili:k* and the answers are:

   6. *t'oyi:xal* 7. *t'oyi:xit* 12. *ʔili:kal* 13. *ʔili:ki*

Rules: There are two types of rule: stem changes (depending on the type of the last syllable of the stem) and suffix changes (depending on the last vowel of the stem).

1. Stem changes. Take place only in non-future and imperative:

   a) *CV:C* ⟶ *CVC* (vowel shortening);
   b) *CVCC* ⟶ *CVCV*$_e$*C* (epenthesis); *V*$_e$ follows the vowel harmony (see below);

2. Suffixes: dubitative = *-al*, passive voice = *-it*, non-future = *hin*, imperative = *k'a*.

3. Vowel harmony. Applied both to the suffixes and to the epenthetic vowel:
   a) If last vowel of the root is *o*: *a* ⟶ *o*;
   b) If last vowel of the root is *u*: *i* ⟶ *u*;

Discussion (not part of the solution)

It is interesting to consider the reason why the stem changes occur. These changes only happen for non-future and imperative, and these are the only forms for which the corresponding suffixes begin with a consonant. Thus, the epenthesis (rule 1b) is rather common, and it is used to avoid a three-consonant cluster, since this would impede the pronunciation.

The vowel shortening, on the other hand, highlights a much more interesting phenomenon which can be connected to syllable weight. Remember that in the previous chapter, we classified syllables into four types, depending on vowel length and on the existence of a coda. In this language, it would appear that syllables which have both a long vowel and a consonant are not allowed; moreover, the onset and the coda cannot be complex (cannot be formed by consonant clusters). Let us analyse each transformation.

Vowel shortening

Let us consider a *CV:C* stem. For the dubitative and passive (whose suffixes begin with a vowel), we would find words of the shape *CV:CVC*. Therefore, based on the basic syllabification rules, we would obtain *CV:.CVC*, the first syllable having a long vowel (but no coda), while the second has a coda (but only a short vowel).

For the other two forms, whose suffixes begin with a consonant, we would get words like **CV:CCVC*, and, by syllabifying them, we would get **CV:C.CVC*. Since the language does not allow *CV:C* syllables, the vowel shortening emerges, resulting in the form *CVC.CVC*.

Note

An asterisk (*) before a word/sentence shows that it is not grammatically correct or is not attested in the language.

Epenthesis

For *CVCC* type stems, in the case of dubitative and passive, whose suffixes begin with a vowel, we get words like *CVCCVC*, which, after syllabification, results in *CVC.CVC*. On the other hand, for the other two forms, we obtain **CVCCCVC*, which, after syllabification, would result in **CVC.CCVC* or **CVCC.CVC*. In both cases, either the onset or the coda would have a consonant cluster, which is not allowed in the language. Therefore, an epenthetic vowel is added, resulting in the word $CVCV_eCCVC \longrightarrow CV.CV_eC.CVC$.

Therefore, we can posit that the two root change phenomena take place in order to avoid consonant clusters or super-heavy syllables (having both a long vowel and a coda).

These observations also help us understand why for blanks 6–7 and 12–13 the first vowel of the stem cannot be long (i.e., we cannot consider answers such *t'o:yixal* and *t'o:yi:xal* to be correct). If the first vowel of the stem were long, there would be no reason for it to be shortened in the imperative form since the environment in which it appears does not change.

Moreover, if this is true, it means that the verbal stems (*CV:C* and *CVCC*) are not single words, since their structure would not be allowed.

Problem 4.5

Evenki (Vlad A. Neacșu, original)

Given below are some words in Evenki in five different cases: nominative singular (e.g., 'the dog'), nominative plural (e.g., 'the dogs'), directional-locative singular (e.g., 'to the dog'), possessive 1SG (e.g., 'my dog'), possessive 1PL (e.g., 'our dog'), as well as their English translations:

Nom. SG	Nom. PL	Dir-loc. SG	Pos. 1SG	Pos. 1PL	Transl.
bitəg	*bitəgsəl*	*bitəglə*	*bitəgwi*	*bitəgmʉn*	‘book’
bɵ:s	*bɵ:ssɵl*	*bɵ:slɵ*	*bɵ:swi*	__(1)__	‘cloth’
udun	*udunsul*	__(2)__	*udunbi*	*udunmun*	‘rain’
igga	*iggasal*	*iggala*	*iggawi*	*iggamun*	‘flower’
ixʉldʉ:r	*ixʉldʉ:rsʉl*	*ixʉldʉ:rlɵ*	*ixʉldʉ:rwi*	*ixʉldʉ:rmʉn*	‘shovel’
nʉxʉn	*nʉxʉnsʉl*	*nʉxʉnlɵ*	*nʉxʉnbi*	*nʉxʉnmʉn*	‘brother’
oron	*oronsol*	*oronlo*	*oronbi*	*oronmun*	‘place’
satan	*satansal*	*satanla*	*satanbi*	*satanmun*	‘candy’
təggəŋ	*təggəŋsəl*	*təggəŋlə*	*təggəŋbi*	*təggəŋmʉn*	‘car’
ʉ:ŋkʉ	*ʉ:ŋkʉsʉl*	*ʉ:ŋkʉlɵ*	*ʉ:ŋkʉwi*	*ʉ:ŋkʉmʉn*	‘towel’
xocco	*xoccosol*	*xoccolo*	*xoccowi*	*xoccomun*	‘shop’
iggə	*iggəsəl*	__(3)__	__(4)__	*iggəmʉn*	‘tail’
jʉ:	__(5)__	*jʉ:lɵ*	__(6)__	*jʉ:mʉn*	‘house’
xə:m	__(7)__	__(8)__	__(9)__	__(10)__	‘meal’
do:son	__(11)__	__(12)__	__(13)__	__(14)__	‘salt’

Problem 4.5a Fill in the blanks.

Problem 4.5b You are given some more Evenki words in the same five forms:

Nom. SG	Nom. PL	Dir-loc. SG	Pos. 1SG	Pos. 1PL	Translation
umatta	*umattasul*	*umattalo*	*umattawi*	*umattamun*	‘egg’
a:gun	*a:gunsal*	*a:gunla*	*a:gunbi*	*a:gunmun*	‘hat’
ʉrəl	*ʉrəlsʉl*	*ʉrəllɵ*	*ʉrəlwi*	*ʉrəlmʉn*	‘child’
moriŋ	*moriŋsol*	*moriŋlo*	*moriŋbi*	*moriŋmun*	‘horse’
xɵ:ggʉ	__(15)__	__(16)__	__(17)__	__(18)__	‘leg’
oʃitta	__(19)__	__(20)__	__(21)__	__(22)__	‘star’
xərʉ:ldi:	__(23)__	__(24)__	__(25)__	__(26)__	‘quarrel’

Fill in the blanks.

ʉ and *ɵ* are vowels pronounced like *u* and *o* respectively, but with the tongue placed centrally (central vowels); *a* is similar to 'u' in 'cut', but the tongue is placed more towards the back (back vowel); *ə* = 'ea' in 'pearl', *ŋ* = 'ng' in 'king', *ʃ* = 'sh' in 'shop'.

The mark *ː* after a vowel denotes length.

Solution

We start by noticing that the nominative singular can be considered the base form. From this word, all other forms are derived. Moreover, we notice that there are no changes (alternations) in this form, so we can consider it as the stem of all other forms.

The next step is separating the segments that are added to obtain the other forms. So for the nominative plural we have the suffixes *-sal*, *-səl*, *-sol*, *-sul*, *-sɵl*, *-sʉl*. Such a variety of suffixes which only differ by the vowel is a strong indicator towards some process of vowel harmony. Therefore, we can classify the words based on the suffix they select for the nominative plural form.

-sal	*-səl*	*-sol*	*-sul*	*-sɵl*	*-sʉl*
igga	*bitəg*	*oron*	*udun*	*bɵːs*	*ixʉldʉːr*
satan	*təggəŋ*	*xocco*			*nʉxʉn*
	iggə				*ʉːŋkʉ*

Indeed, the supposition of vowel harmony is easily confirmed. We can observe that the vowel in the suffix is identical to the last vowel of the stem. Therefore, we deduce that the nominative plural suffix is *-sVl*, where *V* is the last vowel of the stem. Since the suffix repeats the final stem vowel in all cases, we can talk here about a total assimilation rather than vowel harmony.

We can do the same thing for the other three forms. For the directional-locative singular form, we find the suffixes *-la*, *-lə*, *-lo*, *-lɵ*.

-la	*-lə*	*-lo*	*-lɵ*
igga	*bitəg*	*oron*	*bɵ:s*
satan	*təggəŋ*	*xocco*	*ixʉldʉ:r*
			nʉxʉn
			ʉ:ŋkʉ
			jʉ:

In this case, we can assume that, similar to the previous situation, the choice of the suffix is based on the last vowel of the stem. We notice that the suffixes *-la* and *-lə* are used if the last vowel of the root is *a* or *ə*, respectively, so again we can talk about total assimilation. Nevertheless, the suffix *-lɵ* is used both for words whose last vowel is *ɵ*, as well as *ʉ*. Since we have no example of words whose last vowel is *u*, although these are found as a task, we can assume that they will receive the suffix *-lo*, since we can expect that the pairs {*ɵ*, *ʉ*} and {*o*, *u*} behave similarly. Therefore, in this case we can talk about vowel harmony as follows: *a* ← *{a}*, *ə* ← *{ə}*, *o* ← *{o, u}*, and *ɵ* ← *{ɵ, ʉ}*. We can write the rules for this harmony in different ways:

- using words: vowel harmony based on backness and roundness – rounded vowels with the same level of backness will use a suffix containing the corresponding high vowel.
- schematically, showing the vowel in the suffix and the group of vowels for which it is used: *a* ← *{a}*, *ə* ← *{ə}*, *o* ← *{o, u}*, and *ɵ* ← *{ɵ, ʉ}*.
- using phonological rules:

$$\overset{\text{V}}{\left[\,{+}\textsc{round}\,\right]} \rightarrow \overset{\text{V}}{\left[\begin{array}{c}{+}\textsc{high}\\ \alpha\ \textsc{backness}\end{array}\right]} / \overset{\text{V}}{\left[\begin{array}{c}{+}\textsc{round}\\ \alpha\ \textsc{backness}\end{array}\right]} \text{C}_0\text{—}$$

This rule can be interpreted as: “a round vowel will become high and will have the same level of backness as the vowel before it if it is round”. Generally, it is better not to use phonological notation when it comes to vowel harmony, because they can become extremely complex and there is a lot of room for error.

The 1SG possessive has only two forms: *-bi* and *-wi*. This time, we do not expect vowel harmony since the vowel in both forms is the same, and it is the consonant that changes. We can make a table to see the distribution of the two suffixes:

-bi	*-wi*
udun	*bitəg*
nʉxʉn	*bɵ:s*
oron	*igga*
satan	*ixʉldʉ:r*
təggəŋ	*ʉ:ŋkʉ*
	xocco

We can easily observe that the suffix *-bi* is used if the last consonant of the stem is *n* or *ŋ*. Therefore, we can consider the base form of the suffix to be *-wi*, and w → b / {n, ŋ}__ or w → b / [+NASAL]__

The last form is the 1PL possessive, where we only have only two forms: *-mun* and *-mʉn*.

-mun	*-mʉn*
udun	*bitəg*
igga	*ixʉldʉ:r*
oron	*nʉxʉn*
satan	*təggəŋ*
xocco	*ʉ:ŋkʉ*
	iggə
	jʉ:

We can again assume it is vowel harmony, and we can try and test if this hypothesis makes sense from a phonological perspective. Based on the given examples, we deduce that *u* ← *{a, o, u}*, *ʉ* ← *{ə, ʉ}*. Based on the footnote at the end of the problem, we know that *a* is a back vowel (similar to *o* and *u*), while *ʉ* is a central vowel (just like *ə*). Therefore, this harmony is solely based on backness. Moreover, we deduce that *ɵ* will go into the same class as *ʉ*, since it is also a central vowel. Therefore, the 1PL possessive suffix is *-mun* if the last vowel of the root is back or *-mʉn* if the last vowel is central.

Based on these, we can write all the rules and solve task (a):

Solution 4.5a

1. *bɵ:smʉn*	2. *udunlo*	3. *iggələ*	4. *iggəwi*
5. *jʉ:sʉl*	6. *jʉ:wi*	7. *xə:msəl*	8. *xə:mlə*
9. *xə:mbi*	10. *xə:mmʉn*	11. *do:sonsol*	12. *do:sonlo*
13. *do:sonbi*	14. *do:sonmun*		

Rules:

1. *Suffixes*:

   Nom. PL = sV_1l

   Dir-loc. PL = lV_2

   Pos. 1SG = wi and $w \longrightarrow b$ / [+NASAL] _

   Pos. 1PL = mV_3n

2. *Vowel harmony*:

	Last vowel of the stem					
	ə	*ɵ*	*ʉ*	*a*	*o*	*u*
V_1	*ə*	*ɵ*	*ʉ*	*a*	*o*	*u*
V_2	*ə*	*ɵ*	*ɵ*	*a*	*o*	*o*
V_3	*ʉ*	*ʉ*	*ʉ*	*u*	*u*	*u*

It is time to analyse the examples given in task (b). We expect that the rules are mostly similar and perhaps undergo only small changes or some exceptions. Indeed, looking at the given forms, we notice that the suffixes do not change and the nominative singular form is the base form. Nevertheless, it seems that the vowel harmony does not apply here anymore.

Since we know the nominative plural results from a total assimilation, we can start here and notice the changes that took place, since each suffix corresponds to only one vowel.

Nom. SG	Nom. PL
umatta	*umattasul*
a:gun	*a:gunsal*
ʉrəl	*ʉrəlsʉl*
moriŋ	*moriŋsol*

In the case of the first word, since the nominative plural suffix contains the vowel *u*, we expect the last vowel of the stem to also be *u*. Nevertheless, the only vowel *u* in the stem is at the beginning, Therefore, we might consider that vowel harmony is, in fact, triggered by the first vowel of the stem, not the last one. Indeed, this easily checks out for all examples in task (b). The formation of

1SG possessive is not affected since the choice of the suffix is independent of the vowel.

On the other hand, since we changed the rule, and we noticed that vowel harmony is triggered by the first vowel, we need to double-check whether the examples in task (a) still follow this rule. Fortunately, we notice that most of the words in task (a) are monosyllabic (hence have only one vowel) or, if they have more vowels, they contain the same vowel. There are only four exceptions: *bitəg*, *igga*, *ixʉldʉ:r*, and *iggə*. We notice that all of these words have *i* as their first vowel. Moreover, the vowel *i* does not appear in any of the three vowel harmony patterns from before. Therefore, we can deduce that the vowel *i* is neutral and does not trigger (nor affect) vowel harmony. We can therefore rephrase the rules above and claim that vowel harmony is triggered by the first vowel in the stem *which is not i*.

Therefore, we can solve task (b).

Solution 4.5b

15. *xɵ:ggʉsɵl*	16. *xɵ:ggʉlə*	17. *xɵ:ggʉwi*
18. *xɵ:ggʉmʉn*	19. *ofittasol*	20. *ofittalo*
21. *ofittawi*	22. *ofittamun*	23. *xərʉ:ldi:səl*
24. *xərʉ:ldi:lə*	25. *xərʉ:ldi:wi*	26. *xərʉ:ldi:mʉn*

Discussion (not part of the solution)

The Evenki language is an extremely interesting example since it features three rare characteristics:

1. Vowel harmony is triggered by the first vowel and not the last vowel (as is common in Turkic and Mongolic languages).
2. The language has three different types of vowel harmony, each of them occurring in different contexts:
    a) total vowel harmony (total assimilation) – used, for example, to form the plural;
    b) harmony that only affects round vowels, based on backness;
    c) harmony based on backness that affects all vowels, independent of roundness.
3. There is a neutral vowel (*i*) which neither triggers nor affects vowel harmony.

4.6 Initial consonant mutation

Initial consonant mutation is a phenomenon occurring, for example, in Celtic languages (Irish, Breton, Manx, etc.). In these languages, the first consonant of words can undergo certain transformations (called mutations) based on the grammatical context. The context in which the first consonant mutates is variable; it could be depending on possession (e.g., 'my' vs. 'his'), on the numeral that follows (e.g., there might be one form for numerals 1-4 and another form for numerals 5-9), etc.

In linguistics problems, if the initial consonant mutation is featured, some mutation examples will be given (in specific contexts), and you will be asked to deduce the mutation other consonants undergo. Let us consider the following example from Welsh (it is known that *c* is pronounced like 'c' in 'car'):

Context 1	Context 2	Context 3
p	*b*	*mh*
c	*g*	*ngh*
t		

We can see that in the first context, there are only voiceless stops, which, in the second context, become voiced (therefore, based on the pairs $p - b$ and $c - g$, we can infer that the consonant *t* in the second context will mutate to *d*). In the third context, we notice that there are only nasal consonants (followed by *h*). Therefore, we can deduce that *t* will become *nh* in the third context.

Therefore, in all contexts, the place of articulation is preserved, and it is only the manner of articulation and voicing that change (voiceless stop – voiced stop – voiceless nasal).

An example of a problem which features initial consonant mutation is 5.16.

4.7 Practice problems

Problem 4.6

La-Mi (Evgeniya Korovina, MSK 2013)

Here are some words and expressions in Guoyu, the Taiwanese dialect of Mandarin Chinese, as well as their correspondences in the secret language La-Mi (both transcribed into Latin script):

Guoyu	La-Mi	Translation
e hiau	*le i liau hi*	‘capable’
be ts’ai	__(1)__	‘to go shopping’
__(2)__	*lat t’it*	‘to hit’
ts’in t’iam	__(3)__	‘very tired’
__(4)__	*laŋ gin*	‘human’
gi	__(5)__	‘justice’
pia?	*lia? pi?*	‘wall’
kam tsia	*lam kin lia tsi*	‘sugarcane’
p’ɔŋ hɔŋ	*lɔŋ p’in lɔŋ hin*	‘gust (of wind)’
ho k’e?	__(6)__	‘guest of honour’
pak k’ak	*lak pit lak k’it*	‘to clean’
tsap ap	__(7)__	‘ten boxes’

Problem 4.6a Fill in the blanks.

All vowel combinations are pronounced as a single syllable (syllables are separated by blanks); *p’*, *t’*, *k’*, *ts’*, *?* are consonants; *ɔ* is a vowel; *ŋ* = ‘ng’ in ‘king’.

Problem 4.7

Tolaki (Peter Arkadiev, MSK 2016)

Here are some verbal forms in Tolaki in their active and passive voice and their English translations:

Active	Passive	Translation	Active	Passive	Translation
alo	*inalo*	‘take’	*wala*	*niwala*	‘enclose’
daga	*nidaga*	‘guard’	*baho*	__(1)__	‘bathe’
ehe	*inehe*	‘want’	*inu*	__(2)__	‘drink’
geru	*nigeru*	‘scrape’	*kulisi*	__(3)__	‘dig’
hunu	*hinunu*	‘burn’	*mala*	__(4)__	‘shorten’

Active	Passive	Translation	Active	Passive	Translation
luarako	*niluarako*	'grab'	*paho*	__(5)__	'plant'
oli	*inoli*	'fly'	*ruru*	__(6)__	'collect'
saru	*sinaru*	'borrow'	*solongako*	__(7)__	'empty'
tena	*tinena*	'order'	*usa*	__(8)__	'crush'

Problem 4.7a Fill in the blanks.

Problem 4.7b At first, the author wanted to include the following example, but they changed their mind believing it can be confusing. Why might this be?

nahu *ninahu* 'cook'

w = 'v' in 'van'

Problem 4.8

Sorba (John Henderson, UKLO 2010)

Minangkabau is a language of Indonesia that features a number of "play languages" that people use for fun, like Pig Latin in English. One of these "play languages" is Sorba. Here are some examples of standard Minangkabau words and their Sorba play language equivalents:[2]

Minang-kabau	Sorba	English	Minang-kabau	Sorba	English
raso	*sora*	'taste'	*mangecek*	*cermange*	'talk'
rokok	*koro*	'cigarette'	*bakilek*	*lerbaki*	'lightning'
rayo	*yora*	'celebrate'	*sawah*	*warsa*	'rice field'
susu	*sursu*	'milk'	*pitih*	*tirpi*	'money'

[2] *Source:* John Henderson, University of Western Australia, with the assistance of Sophie Crouch. Based on Crouch (2008, 2009) and data from the MPI EVA Minangkabau corpus.

Minang-kabau	Sorba	English	Minang-kabau	Sorba	English
baso	*sorba*	‘language’	*manangih*	*ngirmana*	‘cry’
lamo	*morla*	‘long time’	*urang*	*raru*	‘person’
mati	*tirma*	‘dead’	*cubadak*	*darcuba*	‘jackfruit’
bulan	*larbu*	‘month’	*iko*	*kori*	‘this’
minum	*nurmi*	‘drink’	*gata-gata*	*targa-targa*	‘flirtatious’
lilin	*lirli*	‘wax’	*maha-maha*	*harma-harma*	‘expensive’
mintak	*tarmin*	‘request’	*campua*	*purcam*	‘mix’
apa	*para*	‘father’			

Problem 4.8a Write the Sorba equivalents of the following words:

rancak (‘nice’)
makan (‘eat’)
ampek (‘hundred’)
dapua (‘kitchen’)
jadi (‘happen’)
marokok (‘smoking’)
limpik-limpik (‘stuck together’)

Problem 4.8b If you know a Sorba word, can you work backwards to a single standard Minangkabau word? Demonstrate with the Sorba word *lore* (‘good’).

Problem 4.8c Another “play language” is Solabar. The rules for converting a standard Minangkabau word to Solabar can be worked out from the following examples:

Minangkabau	Solabar	English
baso	*solabar*	‘language’
campua	*pulacar*	‘mix’
makan	*kalamar*	‘eat’

What is the Solabar equivalent of the Sorba word *tirpi* (‘money’)?

Problem 4.8d In writing Minangkabau, does the sequence *ng* represent one sound or two sounds? Provide evidence that supports your answer.

Problem 4.9

Arabic (Anton Somin, Elementy)

Here are some Arabic nouns in their definite and indefinite form, as well as their English translations:

Indefinite	Definite	Translation	Indefinite	Definite	Translation
šams	*aššams*	'sun'	*ɣayma*	*alɣayma*	'cloud'
qamar	*alqamar*	'moon'	*maṭar*	*almaṭar*	'rain'
naǯm	*annaǯm*	'star'	*ṭaqs*	*aṭṭaqs*	'weather'
faǯr	*alfaǯr*	'dawn'	*ǯafāf*	*alǯafāf*	'draught'
yawm	*alyawm*	'day'	*bard*	*albard*	'coldness'
ðalām	*aððalām*	'darkness'	*taham*	*attaham*	'heat'
samāʼ	*assamāʼ*	'sky'			

Arabic linguists classify the consonants into "lunar" and "solar" consonants. It is known that *š* and *t* are solar consonants, while *q* and *m* are lunar consonants.

Problem 4.9a Write the definite form of the following nouns:

muðannab ('comet')
barq ('lightning')
θalǯ ('ice')
nār ('fire')
ḍawʼ ('light')
layla ('night')
ɣurūb ('sunrise')
šitāʼ ('winter')
rabīʕ ('spring')
ṣayf ('summer')
xarīf ('autumn')

Problem 4.9b Classify the consonants *b*, *ḍ*, *f*, *ɣ*, *n*, *r*, *θ*, *y* and *z* into the two categories proposed by Arabic linguists (solar and lunar).

Problem 4.9c Arab language historians know that the sound represented by one of the Arabic letters was, in time, replaced by another one (in this problem the modern variant is used). Determine which letter it is.

š = ‘sh’ in ‘shop’, *ǯ* = ‘j’ in ‘judge’, *y* = ‘y’ in ‘year’, *ð* = ‘th’ in ‘that’, *θ* = ‘th’ in ‘thin’, *q* = ‘c’ in ‘car’, *x* = ‘ch’ in ‘loch’, *ɣ* is similar to *x*, but voiced. *ʕ* and *ʾ* are consonants.

A dot below a consonant denotes its special pronunciation (so-called emphatic). A bar above the vowel denotes length.

Problem 4.10

Sesotho (Tamila Krashtan, UkrLO 2021)

Sesotho is primarily spoken in two countries: South Africa and Lesotho. As a result, two different orthographies are used for this language. For example, the word ‘ostrich’ is written *mpjhe* in South Africa, but *mpshe* in Lesotho.

Below are given some words in Sesotho. Some of them are written in one of the orthographies (whether South Africa or Lesotho), while others are written in both orthographies. Two of the words are the same in both orthographies.

oache	*yohle*	*titjhere*	*nngwaya*
Kholu	*ʾnete*	*phela*	*chelete*
kalima	*Kgodu*	*nkoe*	*ntate*
kutloisiso	*lula*	*tjhelete*	*ʾme*
kadima	*hlompshoa*	*cha*	*ea*
kgwedi	*nkwe*	*nnete*	
nwa	*ya*	*Mokgatjhane*	

Problem 4.10a For each of the words, determine in which orthography it is written. For the words written in only one of the two orthographies, provide their equivalent in the other one.

Problem 4.10b In which orthography are the words *joang* and *shwa* written?

Problem 4.11

Dutch (Ksenia Gilyarova, MSK 2003)

The Dutch language uses suffixes to form the diminutives of nouns. Any Dutch noun can be transposed to its diminutive form. Below are some Dutch words together with their diminutive forms and their English translation:

Word	Diminutive	Transl.	Word	Diminutive	Transl.
leeuwerik	*leeuwerikje*	'lark'	*dag*	*dagje*	'day'
trom	*trommetje*	'drum'	*stro*	*strootje*	'straw'
peer	*peertje*	'pear'	*vrucht*	*vruchtje*	'fruit'
snor	*snorretje*	'moustache'	*steeg*	__(1)__	'alley'
huis	*huisje*	'house'	*steg*	__(2)__	'way'
potlood	*potloodje*	'pencil'	*bioscoop*	__(3)__	'cinema'
paraplu	*parapluutje*	'umbrella'	*deur*	__(4)__	'door'
viool	*viooltje*	'violin'	*auto*	__(5)__	'car'
tuin	*tuintje*	'garden'	*zoon*	__(6)__	'son'
ster	*sterretje*	'star'	*zon*	__(7)__	'sun'
komkommer	*komkommertje*	'cucumber'	*mus*	__(8)__	'sparrow'
sla	*slaatje*	'salad'	*winkel*	__(9)__	'store'
kam	*kammetje*	'comb'	*bal*	__(10)__	'ball'
web	*webje*	'internet'	*ballet*	__(11)__	'ballet'
pin	*pinnetje*	'pin'	*pyjama*	__(12)__	'pyjamas'
verhaal	*verhaaltje*	'story'	*schim*	__(13)__	'ghost'
wodka	*wodkaatje*	'vodka'	__(14)__	*petje*	'beret'

Problem 4.11a Fill in the blanks.

Problem 4.11b The word *vlootje* is a homonym, representing the diminutive of two different words. Which are these words?

Problem 4.12

Finnish – Estonian (Vlad A. Neacșu, HKLO 2021)

Here are some words in Finnish (F) and Estonian (E), declined for the nominative, genitive, and illative cases:[3]

[3] *Source:* Adapted after a problem by Andrei Zaliznyak (published in *Задачи лингвистических олимпиад 1965–1975* (*Problems for the Linguistics Olympiad 1965–1975*), Moscow, 2007.

	Nominative		Genitive		Illative	
English	F	E	F	E	F	E
‘people’	*rahvas*	*rahvas*	__(1)__	*rahvas*	__(2)__	__(3)__
‘naked’	__(4)__	__(5)__	*paljaan*	__(6)__	*paljaaseen*	*paljasse*
‘row’	__(7)__	*toores*	*tuoreen*	*toore*	*tuoreeseen*	*tooresse*
‘axe’	*kirves*	*kirves*	__(8)__	__(9)__	*kirveeseen*	__(10)__
‘ready’	*valmis*	__(11)__	*valmiin*	*valmi*	__(12)__	*valmisse*
‘part’	*osa*	__(13)__	*osan*	*osa*	*osaan*	*ossa*
‘city’	*linna*	__(14)__	*linnan*	*linna*	__(15)__	*linna*
‘village’	*külä*	*külä*	*külän*	*küla*	*külään*	__(16)__
‘shelter’	*maja*	*maja*	*majan*	*maja*	*majaan*	*majja*
‘ace’	__(17)__	*äss*	__(18)__	__(19)__	*ässään*	*ässa*
‘wheel’	*püörä*	__(20)__	__(21)__	__(22)__	__(23)__	__(24)__
‘snow’	*lumi*	*lumi*	*lumen*	__(25)__	*lumeen*	*lumme*
‘horn’	*sarvi*	*sarv*	__(26)__	*sarve*	*sarveen*	*sarve*
‘cape’	*niemi*	*neem*	*niemen*	*neeme*	__(27)__	*neeme*
‘hackberry’	__(28)__	*toom*	__(29)__	*toome*	__(30)__	__(31)__
‘sea’	__(32)__	__(33)__	__(34)__	__(35)__	__(36)__	*merre*

Problem 4.12a Fill in the blanks.

For this problem the orthography of Finnish has been slightly changed. In reality, the character denoted here by *ü* is written as *y*.

Problem 4.13

Bari (Jan Petr, ČLO 2019)

Given below are some verbal roots in Bari, together with two different forms, as well as their English translations. The shaded cells represent forms that are not essential for solving the problem (they may exist).

Root	Form 1	Form 2	Tone	Translation
dɛ́ʔ	*dɪ́lɪ́kɪ́n*			‘to bend’
kʊ́r	*kʊ́rə̀kɪ́n*	*kʊ́rə̀rə̂ʔ*		‘to borrow’
ʼdɔ́k	*ʼdʊ́kʊ́kɪ́n*			‘to carry’
mók	*mòkákìn*	*mòkáràʔ*		‘to catch’
__(1)__	*tʊ́kʊ́kɪ́n*	*tɔ́kɔ́rɔ̂ʔ*		‘to cut with an axe’
__(2)__	__(3)__	*tòjúpùrùʔ*		‘to dress’
yúk	*yùkúkìn*	*yùkúrùʔ*		‘to shepherd’
ʼdép	*ʼdépákín*	__(4)__		‘to hold’
gáʔ	__(5)__		*g*	‘to seek’
lʊ́sə̀k	__(6)__			‘to defrost’
sə̀pʊ̂k	__(7)__	__(8)__		‘to return’
ʼyút	*ʼyùtúkìn*	__(9)__		‘to seed’
tɔ̀kʊ̂	*tɔ̀kʊ́kɪ̀n*	*tɔ̀kʊ́ə̀rə̀ʔ*	*t*	‘to preach’
bʊ́dʊ́	*bʊ́dʊ́kɪ́n*			‘to reach the top’
báʔ	*bàlákìn*			‘to punish’
sɔ́n	*sʊ́nyʊ́kɪ́n*	*sɔ́nyɔ́rɔ̂ʔ*		‘to send (something)’
yə̀kɪ̂	*yə̀kɪ́kɪ̀n*	*yə̀kɪ́ə̀rə̀ʔ*		‘to send (someone)’
dòdôngʼ	*dòdóngʼàkìn*	*dòdóngʼàràʔ*		‘to shake’
__(10)__	*ʼbórókín*			‘to smear’
lɪ̀lɪ̂ngʼ	*lɪ̀lɪ́ngʼə̀kɪ̀n*	__(11)__		‘to exterminate’
rɛ́m	*rɪ́mɪ́kɪ́n*			‘to inject’
bérén	__(12)__			‘to poison’
__(13)__	*lókín*			‘to dry in the sun’
dwán	*dwànyákìn*			‘to open’
lák	__(14)__	*lákárâʔ*		‘to untie’
dɔ́k	__(15)__		*g*	‘to unpack’

Problem 4.13a Bari verbs can be classified into two groups, based on their tone. Fill in the column “Tone”, specifying whether the verb has the tone *g* (it behaves like *gáʔ* and *dɔ́k*) or *t* (it behaves like *tɔ̀kʊ̂*).

Problem 4.13b Fill in the blanks.

'b, *'d*, *'y*, *ng'*, *ny*, *y*, *ʔ* are consonants. A line below a vowel denotes that the vowel is pronounced with an advanced tongue root (+ATR). The marks ◌́, ◌̀ and ◌̂ above a vowel denote high, low and falling tones, respectively.

Problem 4.14

Cushillococa Ticuna (Tsuyoshi Kobayashi, APLO 2021)

Here are some words and phrases in Cushillococa Ticuna and their English translations:

'kɑ̲^{1} a^{5} tɨ^{3}	'kɑ̲^{1} tree leaves'
'ku^{43} te^{4} e^{3} ɟɑ̲^{1}	'your husband's sister'
'ku^{43} ʔɑ̲^{1}	'your mouth'
'ku^{43} ʔu^{4} ne^{1}	'your entire body'
na^{4} 'me^{43} ʔe^{5} tʃi^{1}	'it is really good'
na^{4} 'bu^{3} ʔu^{1} ra^{1}	'it is sort of immature'
'to^{5} ne^{1}	'owl monkey's tree trunk'
'tɨ^{2} ʔe^{1} a^{1} ne^{1}	'cassava garden'
'tɑ̲^{1} ʔtʃi^{5} ru^{1}	'owl monkey's clothes'
'to^{1} ʔɑ̲^{1}	'other one's mouth'
'tɑ̲^{1} ʔo^{5} tʃi^{1}	'really an owl monkey'
'tʃau^{1} ʔtʃi^{5} ru^{1}	'my clothes'
'tʃo^{1} ʔmɑ̲^{1} ne^{1}	'my wife's tree trunk'
'tʃo^{1} me^{4} na^{2} ʔã^{2}	'my stick'
'to^{1} bɨ^{2}	'other one's high-starch food'
'tʃo^{1} pa^{3} tɨ^{4}	'my fingernail'
'tʃau^{1} e^{3} ɟɑ̲^{1} te^{4}	'my sister's husband'
na^{4} 'tʃɑ̲^{1} bɨ^{2}	'its high-starch food is delicious'
na^{4} 'tʃo^{5} o^{1} ne^{1} ʔɨ^{1} ra^{1}	'its garden is sort of white'
'ŋo^{3} ʔɑ̲^{1} a^{1} ne^{1}	'place where there are lots of ŋo^{3} ʔɑ̲^{1}'

Problem 4.14a What is the *literal* translation of *'ŋo^{3} ʔɑ̲^{1} a^{1} ne^{1}*?

Problem 4.14b Translate into English:

1. *'ka^{5} ne^{1}*
2. *na^{4} 'tʃo̱^{1} o^{5} tɨ^{3}*
3. *'ŋo^{3} ʔo̱^{1} ʔɨ^{5} tʃi^{1}*
4. *'to^{1} o^{1} ne^{1}*
5. *'to̱^{1} ʔo^{4} ne^{1}*
6. *'tʃau^{1} ne^{1}*

Problem 4.14c Translate into Cushillococa Ticuna:

7. 'it is sort of delicious'
8. 'its clothes are really white'
9. 'my husband's entire body'
10. 'my high-starch food'

tʃ, *ɟ*, *ŋ*, and *ʔ* are consonants; *ɨ* is a vowel; *au* is a diphthong: consider it as one vowel. The mark ' indicates that the following syllable is stressed. ○^{1}, ○^{2}, ○^{3}, ○^{4}, ○^{5}, and ○^{43} denote tones of the preceding syllable. Pitches of the tones:

low = ○^{1} < ○^{2} < ○^{3} < ○^{4} < ○^{5} = high; ○^{43} = ○^{4} ↘ ○^{3}

A tilde below a vowel (e.g., *a̱*) denotes creaky voice (a type of phonation that is often perceived as low-pitched and "rough"). A tilde over a vowel (e.g., *ã*) denotes a nasal sound.

An 'owl monkey' is a type of monkey. 'Cassava' is a woody plant native to South America. A 'ka̱^{1} tree' is a kind of fruit tree. 'ŋo^{3} ʔo̱^{1}' is a kind of fish.

4.8 Solutions of practice problems

Solution for practice problem 4.6. La-Mi

Solution 4.6a

Guoyu	La-Mi	Translation
be ts'ai	*le bi lai ts'i*	'to go shopping'
t'at	*lat t'it*	'to hit'
ts'in t'iam	*lin ts'in liam t'in*	'very tired'
gaŋ	*laŋ gin*	'human'
gi	*li gi*	'justice'
ho k'eʔ	*lo hi leʔ k'iʔ*	'guest of honour'
tsap ap	*lap tsit lap it*	'ten boxes'

Rules: We note with C_1, V and C_2 the onset, nucleus and coda of the syllable, respectively. Now each syllable can be written as $(C_1)V(C_2)$. The transformation is: $C_1VC_2 \longrightarrow lVC_2\ C_1iX$, where X depends on C_2, as follows:

C_2	X
∅	∅
ʔ	*ʔ*
m, *n*, *ŋ*	*n*
p, *t*, *k*	*t*

We deduce that X is identical with C_2, but with an assimilated place of articulation (alveolar). Exception: *ʔ* (which remains unchanged). Another explanation is:

- ʔ→ʔ
- $\left[\ \text{+NASAL}\ \right] \rightarrow$ n
- $\left[\begin{array}{c}\text{+STOP}\\ \text{−VOICE}\end{array}\right]\rightarrow$t

Solution for practice problem 4.7. Tolaki

Solution 4.7a

baho	*nibaho*	'bathe'
inu	*ininu*	'drink'
kulisi	*kinulisi*	'dig'
mala	*nimala*	'shorten'
paho	*pinaho*	'plant'
ruru	*niruru*	'collect'
solongako	*sinolongako*	'empty'
usa	*inusa*	'crush'

Solution 4.7b Because it is not known which *n* is part of the stem and which one is part of the affix. Thus, it can either be the prefix *ni-* added to the word *nahu*, or the infix *-in-*, added after the first consonant of the word *nahu*.

Rules:

- If the lexeme starts with a vowel, add *in-* at the beginning;
- If the lexeme starts with a voiced consonant, add *ni-* at the beginning.
- If the lexeme starts with a voiceless consonant, add *-in-* after the first consonant.

Solution for practice problem 4.8. Sorba

Solution 4.8a *caran*, *dirja*, *karma*, *kormaro*, *peram*, *kormaro*, *peram*, *pirlim-pirlim*, *purda*

Solution 4.8b The word cannot be uniquely determined since the coda of the last syllable disappears. Moreover, it is not known if the *r* in *lore* is part of the root or if it is added, thus *lore* can result from any of the following words: *relo*, *elo*, *reloC*, *eloC* (where *C* can be any consonant).

Solution 4.8c *tilapir*

Solution 4.8d A single sound, since the word *manangih* becomes *ngirmana* in Sorba. If *ng* was two sounds, the word would have become *girmanan*.

Rules: In order to form the Sorba word, we need to take the last syllable of the word, remove its coda (only keep the first consonant – if any – and the first vowel) and add it to the beginning of the word, separated by an *r*. If the word already begins with an *r*, no extra *r* is added. Thus, a word such as *...(C)V(V)(C)* becomes *(C)Vr...*. It is important to notice that two consecutive vowels will form a diphthong, according to the example *cam.pua* ⟶ *pu-r.cam*. If it were a hiatus, we would get *cam.pu.a* ⟶ *a-r.cam.pu*.

In order to form the Solabar word, the same process as in Sorba is applied, but the connecting *r* is replaced by *la* (resulting in *(C)Vla...*).

Solution for practice problem 4.9. Arabic

Solution 4.9a

almuðannab	*albarq*	*aθθalǯ*
annār	*aḍḍaw´*	*allayla*
alɣurūb	*aššitā´*	*arrabīʕ*
aṣṣayf	*alxarīf*	

Solution 4.9b Lunar consonants: *b*, *f*, *ɣ*, *y*

Solar consonants: *ḍ*, *n*, *r*, *θ*, *z*

Solution 4.9c *ǯ*

Rules: Solar consonants include the coronal consonants. In their case, the definite form is constructed by adding the prefix *aX-* (where *X* is the first consonant of the word). Lunar consonants are all the rest (labial and dorsal), and if a word begins with a lunar consonant, the definite form is simply obtained by adding the prefix *al-*. Alternatively, we can write:

$$\text{l} \rightarrow \begin{matrix}\text{C} \\ [+\text{CORONAL}]\end{matrix} \; / \; \begin{matrix}\text{C} \\ [+\text{CORONAL}]\end{matrix} \; \text{—}$$

The consonant *ǯ*, although coronal, does not assimilate the prefix *al-*; therefore, most likely, in the past, it was pronounced as a dorsal (probably as a voiced palatal plosive).

Solution for practice problem 4.10. Sesotho

Solution 4.10a

South Africa	Lesotho
dula	*lula*
hlompjhwa	*hlompshoa*
kadima	*kalima*
Kgodu	*Kholu*
kgwedi	***khoeli***
kutlwisiso	*kutloisiso*
mme	*'me*
Mokgatjhane	***Mokhachane***
nkwe	*nkoe*
nnete	*'nete*

South Africa	Lesotho
nngwaya ntate	***'ngoaea***
nwa phela	***noa***
titjhere	***tichere***
tjha	*cha*
tjhelete	*chelete*
watjhe	*oache*
ya	*ea*
yohle	***eohle***

The words in bold are those that do not appear in the dataset.

Solution 4.10b *joang* – Lesotho (in South Africa it is written as *jwang*)
shwa – South Africa (in Lesotho it is written as *shoa*)

Rules: We have the following sound correspondences:

South Africa	Lesotho
di	*li*
du	*lu*
kg	*kh*
mm	*'m*
nn	*'n*
pjh	*psh*
tjh	*ch*
w + vowel	*o* + vowel
y + vowel	*e* + vowel

Solution for practice problem 4.11. Dutch

Solution 4.11a

(1) *steegje*
(2) *stegje*
(3) *bioscoopje*
(4) *deurtje*
(5) *autootje*
(6) *zoontje*
(7) *zonnetje*
(8) *musje*
(9) *winkeltje*
(10) *balletje*
(11) *balletje*
(12) *pyjamaatje*
(13) *schimmetje*
(14) *pet*

Solution 4.11b *vlo* and *vloot*

Rules: The diminutive depends on the last sound of the word. We have the following cases:

- vowel ⟹ add *Vtje* (where *V* is the last vowel of the word);
- obstruent (stop or fricative) ⟹ add *-je*;
- sonorant (nasal or liquid – *m*, *n*, *l*, *r*):
  - if the word has only one syllable and the vowel is short, add the suffix *-Cetje* (where *C* is the last consonant);
  - else, add *-tje* (if (1) the word has only one syllable and contains a long vowel or a diphthong or (2) the word contains more than a syllable).

Solution for practice problem 4.12. Finnish – Estonian

Solution 4.12a

(1) *rahvaan*	(13) *osa*	(25) *lume*
(2) *rahvaaseen*	(14) *linn*	(26) *sarven*
(3) *rahvasse*	(15) *linnaan*	(27) *niemeen*
(4) *paljas*	(16) *külla*	(28) *tuomi*
(5) *paljas*	(17) *ässä*	(29) *tuomen*
(6) *palja*	(18) *ässän*	(30) *tuomeen*
(7) *tuores*	(19) *ässa*	(31) *toome*
(8) *kirveen*	(20) *pöör*	(32) *meri*
(9) *kirve*	(21) *püörän*	(33) *meri*
(10) *kirvesse*	(22) *pööra*	(34) *meren*
(11) *valmis*	(23) *püörään*	(35) *mere*
(12) *valmiiseen*	(24) *pööra*	(36) *mereen*

Rules: We divide the nouns (nominative, Finnish) into three classes: ending in *s* (preceded by a vowel), ending in *i*, and ending in another vowel. We get:

	Nominative		Genitive		Illative	
	F	E	F	E	F	E
Class I	*-Vs*	*-Vs*	*-VVn*	*-V*	*-VVseen*	*-Vsse*
Class II	*-V*	*	*-Vn*	*-V*	*-VVn*	*-V*
Class III	*-i*	*	*-en*	*-e*	*-een*	*-e*

* To form the Estonian nominative, if the Finnish root contains a diphthong or a consonant cluster, the final vowel is dropped in Estonian (*-V* $\longrightarrow \emptyset$). Else, the form is identical to the Finnish one (exception: *ä* $\longrightarrow$ *a* / _ #).

- Diphthongs in Finnish become long vowels in Estonian: $V_1V_2 \longrightarrow V_2V_2$.
- If the Estonian nominative ends in a vowel, the consonant before it is doubled in the illative. (*-CV* $\longrightarrow$ *-CCV* or *Ci* $\longrightarrow$ *-CCe*).

Solution for practice problem 4.13. Bari

Solution 4.13a

Root	Form 1	Form 2	Tone	Translation
dɛ́ʔ	*dɪ̰lɪ̰́kɪ̰́n*		*g*	‘to bend’
kʊ́r	*kʊ́rə̰kɪ̰́n*	*kʊ́rə̰rə̰̂ʔ*	*g*	‘to borrow’
ʼdɔ́k	*ʼdʊ́kʊ́kɪ̰́n*		*g*	‘to carry’
mók	*mòkákìn*	*mòkáràʔ*	*t*	‘to catch’
tɔ́k	*tʊ́kʊ́kɪ̰́n*	*tɔ́kɔ́rɔ̰̂ʔ*	*g*	‘to cut with an axe’
tòjûp	*tòjúpùkìn*	*tòjúpùrùʔ*	*t*	‘to dress’
yúk	*yùkúkìn*	*yùkúrùʔ*	*t*	‘to shepherd’
ʼdép	*ʼdépákín*	*ʼdépárâʔ*	*g*	‘to hold’
gáʔ	*gálákín*		*g*	‘to seek’
lʊ́sə̰k	*lʊ́sə̰kə̰kɪ̰́n*		*g*	‘to defrost’
sə̰pʊ̂k	*sə̰pʊ́kə̰kɪ̀n*	*sə̰pʊ́kə̰rə̰ʔ*	*t*	‘to return’
ʼyút	*ʼyùtúkìn*	*ʼyùtúrùʔ*	*t*	‘to seed’
tɔ̀kʊ̂	*tɔ̀kʊ́kɪ̀n*	*tɔ̀kʊ́ə̰rə̰ʔ*	*t*	‘to preach’
búdʊ́	*búdʊ́kɪ́n*		*g*	‘to reach the top’
báʔ	*bàlákìn*		*t*	‘to punish’
sɔ́n	*sʊ́nyʊ́kɪ́n*	*sɔ́nyɔ́rɔ̰̂ʔ*	*g*	‘to send (something)’
yə̰kɪ̂	*yə̰kɪ́kɪ̀n*	*yə̰kɪ́ə̰rə̰ʔ*	*t*	‘to send (someone)’
dòdôngʼ	*dòdóngʼàkìn*	*dòdóngʼàràʔ*	*t*	‘to shake’
ʼbóró	*ʼbórókín*		*g*	‘to smear’
lɪ̀lɪ̰̂ngʼ	*lɪ̀lɪ́ngʼə̰kɪ̀n*	*lɪ̀lɪ́ngʼə̰rə̰ʔ*	*t*	‘to exterminate’
rɛ́m	*rɪ́mɪ́kɪ́n*		*g*	‘to inject’
bérén	*bérényákín*		*g*	‘to poison’
ló	*lókín*		*g*	‘to sundry’
dwán	*dwànyákìn*		*t*	‘to open’
lák	*lákákín*	*lákárâʔ*	*g*	‘to untie’
dɔ́k	*dbúkʊ́kɪ́n*		*g*	‘to pack’

Rules:

1. Change of the final consonant of the root (applied for both forms): *n* ⟶ *ny*, *ʔ* ⟶ *l*;

2. Tone:
    a) Type *g*: all vowels (root and both forms) have high tone (*á*), except for the last vowel of Form 2 which has falling tone (*â*);
    b) Type *t*: chosen depending on the number of vowels (syllables), independent of the form:
        - 1 syllable: high tone (*á*);
        - 2 syllables: low + falling (*à* + *â*);
        - 3 syllables: *à* + *á* + *à*;
        - 4 syllables: *à* + *á* + *à* + *à*;

3. Form 1:
    a) Added suffix:

	Last vowel of the stem	
Stem ends in	−ATR	+ATR
consonant	*-akin*	*-a̱ki̱n*
vowel	*-kin*	*-ki̱n*

    b) If the last vowel of the root is *u*: *akin* ⟶ *ukin*;
    c) If the last vowel of the root is *o̱*, it becomes *u̱* and *a̱ki̱n* ⟶ *u̱ki̱n*;
    d) If the last vowel of the root is *e̱*, it becomes *i̱* and *a̱ki̱n* ⟶ *i̱ki̱n*;

Note

Rules c) and d) can be combined and rewritten as: if the last vowel of the stem is [+FRONT] and [+ATR], it will become [+BACK] and the epenthetic vowel *a* will fully assimilate to it.

4. Form 2:
    a) Added suffix: *-araʔ* or *-a̱ra̱ʔ* (harmony based on ATR);
    b) If the last vowel of the stem is *u*, *-araʔ* ⟶ *-uruʔ*;
    c) If the last vowel of the stem is *o̱*, *-a̱ra̱ʔ* ⟶ *-u̱ru̱ʔ*;

Note

Vowel changes (rules 3b-d and 4b-c) can be explained in another way: considering the last vowel of the root (V_1) and the two vowels of the added suffixes (*-VkVn* and *-VrVʔ*), we have the following transformations:

	Form 1	Form 2
V_1	V_1-V-V	V_1-V-V
o̱	*u̱-u̱-i̱*	*o̱-o̱-o̱*
u	*u-u-i*	*u-u-u*
e̱	*i-i-i*	

Solution for practice problem 4.14. Cushillococa Ticuna

Solution 4.14a ‘ŋo^3 ʔo̱^{1} (’s) garden’

Solution 4.14b
1. ‘ka̱^{1} tree trunk’
2. ‘its leaves are white’
3. ‘really a ŋo^3 ʔo̱^{1}’
4. ‘other one’s garden’
5. ‘owl monkey’s entire body’
6. ‘my tree trunk’

Solution 4.14c
7. *na^4 ˈʧi^5 ʔi^1 ra^1*
8. *na^4 ˈʧo̱^{1} ʔʧi^5 ru^1 ʔi̱^{5} ʧi^1*
9. *ˈʧau^1 te^4 ʔi̱^{4} ne^1*
10. *ˈʧo^1 bi̱^{2}*

Rules:

- The possessive and the adjective are placed before the noun.
- The possessive is marked by: *to*^{1} (‘other one’s’), *ku*^{43} (‘your’), *tʃau*^{1} (‘his’).
  - *tʃau*^{1} becomes *tʃo*^{1}, if it is before a bilabial consonant (*p*, *b*, *m*). The glottal stop does not block the transformation. Thus:
    $$tʃau^{1} \longrightarrow tʃo^{1} \; / \; \_(ʔ)C_{bilabial}$$
  - The stress falls on the first syllable. If the first syllable is *na*^{4} (‘it is’), the stress shifts onto the second syllable.
  - Phonological processes:
    - *a* ⟶ *o* / *ˈo(ʔ)* _ (*a* becomes *o* if it follows a stressed syllable that contains *o*, if there is no consonant between them (except for glottal stop));
    - *ɨ* ⟶ *V* / *ˈV(ʔ)* _ (*ɨ* fully assimilates to the preceding vowel if this vowel is in a stressed syllable and between them there is no other consonant (except for glottal stop));
    - *ˈV̰*^{1} ⟶ *ˈV*^{5} / _ *(C)V*^{1} (tone dissimilation – a pharyngealised vowel with tone 1 in a stressed syllable will become non-pharyngealised and with tone 5 if it is before another syllable with tone 1).

Supplementary: Dictionary of base forms

Adjectives

bu^{3} = ‘immature’

me^{43} = ‘good’

tʃḭ^{1} = ‘delicious’

tʃo̰^{1} = ‘white’

Nouns

a^{1} *ne*^{1} = ‘garden’

a^{5} *tɨ*^{3} = ‘leaves’

ʔa̰^{1} = ‘mouth’

bɨ^{2} = ‘high-starch food’

e^{3} *ɲa̰*^{1} = ‘sister’

ʔɨ^{4} *ne*^{1} = ‘entire body’

me^{4} *na*^{2} *ʔã*^{2} = ‘stick’

ʔma̰^{1} = ‘wife’

ne^{1} = ‘tree trunk’

pa^{3} *tɨ*^{4} = ‘fingernail’

te^{4} = ‘husband’

tɨ^{2} *ʔe*^{1} = ‘cassava’

to̰^{1} = ‘owl monkey’

ʔtʃi^{5} *ru*^{1} = ‘clothes’

Further reading

Rocca, Iggy & Wyn Johnson. 1999. *Course in phonology*. Oxford: Blackwell.
Zsiga, Elizabeth C. 2013. *The sounds of language: An introduction to phonetics and phonology*. Chichester: Wiley-Blackwell.

5 Noun and noun phrase

5.1 Introduction

Before presenting and analysing the main types of problems, it is important to get accustomed to the following general concepts. Traditional notional definitions of grammatical categories include the following:

Noun: words that name specific objects, places, beings, etc.

Proper noun: refers to names of people or places (*Giovanni*, *London*, etc.)

Common noun: describes a class of entities (*city*, *class*, *girl*, *cat*, etc.).

We can also think about the ways in which words combine to form larger constituents and the roles these larger constituents play in syntactic structures. For instance:

Noun phrase: a phrase that has a noun as its head and performs various grammatical functions such as subject and direct object

Thus, we can define a sentence (S) as a combination of a noun phrase (NP) and a verb phrase (VP). We can write: $S = NP + VP$. For example:

(1) a. *Andrew eats.* (NP = Proper noun = *Andrew*, VP = V = *eats*)
b. *A boy eats those sandwiches.* (NP = *A boy*, VP = V + NP = *eats* + *those sandwiches*).

Other examples of noun phrases (the word in bold represents the head noun) can be: *the good* ***boy***, *the* ***boy*** *inside the house*, *the five green* ***tables***, etc.

We notice that the noun phrase includes the noun (the *head* component) and other words subordinated to it (definite article, possessives, adjectives, etc.). We usually refer to all of these subordinated components as *modifiers*.[1]

[1]Sometimes the term *determiner* is also used in linguistics problems. In current theories, however, the term determiner refers strictly to the definite/indefinite article (the grammatical category of determination). Thus, the term modifier is usually preferred to refer to the elements of the noun phrase (adjectives, demonstratives, etc.).

Note

Some words can belong, depending on the sentence structure, to either the noun phrase or the verb phrase. For example, the sentences *The boy inside the house eats* and *The boy eats inside the house* differ only in terms of word order. In the former, *inside the house* is part of the noun phrase and modifies *boy* (the boy who is inside the house), while in the latter it refers to where the action takes place and modifies the verb, as part of the verb phrase.

Morpheme: the smallest linguistic unit with a meaning or a grammatical function. This is a broad term which includes different subtypes, such as stems (usually, the lexical stem of the word), prefixes, suffixes, etc. For example, the word *undesirability* is made up of four morphemes: the morpheme *desire* (which is also the stem), the morpheme *-able* (forming the adjective *desirable*), the morpheme *un-* (forming the adjective *undesirable*), and finally, the morpheme *-ity* (forming the noun *undesirability*).

Allomorph: a variant of a morpheme. In some situations, a morpheme can have two or more similar forms, which are chosen based on phonological considerations. For example, in the words *impossible* and *intolerant*, the prefixes *im-* and *in-*, although they look different, serve the same function, for which reason we can consider them to be the same morpheme. Moreover, from a phonological point of view, *im-* appears when the nasal assimilates to the place of articulation of the following sound (*p* in ***p****ossible*).

Affix: a morpheme added to the stem. It is also a rather broad term which includes different types of affixes, like suffix or prefix.

Prefix: affix placed at the beginning of the word.

Suffix: affix placed at the end of the word.

Although English has only two types of affixes (prefixes and suffixes), other languages can have a much greater variety, as follows:

Infix: affix placed inside the stem.

Let us consider the following examples from Bontoc:

fikas	'strong'	*fumikas*	'to be strong'
kilad	'red'	*kumilad*	'to be red'
pusi	'poor'	*pumusi*	'to be poor'

We can notice that the derivation of a verb from an adjective (e.g., 'red' → 'to be red') is made by adding the letters *um* after the first letter of the adjective. Thus, *-um-* is an example of an infix.

Circumfix: is a type of affix with two components, one part that is added to the beginning of the word and another that is added to the end of the word.

To some extent, a circumfix can be considered as a combination of a prefix and a suffix. Nevertheless, there is a major difference between a circumfix and a prefix + suffix combination: the circumfix is indivisible (i.e., the part in the beginning has neither meaning nor function without the part at the end; both parts are needed in order to achieve a meaning or function), while a prefix + suffix combination represent two independent entities, each of them having their own function or meaning. An example of a circumfix can be found in Chickasaw, in which the negation is formed using the circumfix *ik- -o* added to the stem, with the additional property that the final vowel of the stem is dropped:

chokma	'he is good'	*ikchokmo*	'he is not good'
tiwwi	'he opens'	*iktiwwo*	'he does not open'
palli	'it is hot'	*ikpallo*	'it is not hot'

Transfix: is a type of discontinuous affix (resembling a combination of different infixes).

Transfixes are generally associated with Semitic languages (Arabic, Hebrew, Maltese, etc.) in which, for example, the verb stems are discontinuous, consisting of two to four consonants. Thus, in Maltese, the stem of the verb 'to write' is *k-t-b*. In order to conjugate it, we need to add different vowels before, between or after the consonants of the stem. For example, *kiteb* = 'he wrote'. So, in order to obtain the 3SG masculine past simple of the verb, we need to use the transfix *-i-e-*.

Sometimes, part of the stem is deleted. We will refer to the part of the word that is deleted as a DISFIX.

Let us consider the following example in Alabama:

tipsali ‘to break’ *tipli* ‘to break up’

In this case, the second word is formed from the first through the elision of the penultimate syllable of the stem (*sa*). Based on this example, we could in fact consider that the second word is the stem and *sa* is merely an infix added to form the verb to break. Let us consider another example from the same language: *batatli – batli*. The formation of the second word can be easily explained using our first theory, in which the penultimate syllable is dropped; nevertheless, if we try to use our second explanation, in which an infix is added, we notice that in the first example, the infix added is *-sa-*, while the second example uses the infix *ta*. Moreover, there seems to be no connection or pattern regarding the choice of the infix. As a result, the disfix is the reasonable explanation, since the deleted morpheme is not predictable (cannot be deduced), but rather it is intrinsic to the stem.

Most of the linguistics problems focused on the noun phrase will include different ways of forming new words (either singular – plural, etc.) or will focus on how nouns interact with different modifiers.

5.2 Basic principle of morphological analysis

1. If a single phonetic form has two distinct meanings (or functions), it must be analysed as representing two different morphemes.

For example, the morpheme *-s* in the words *cats* and *sees* could be considered, at first sight, to be a single morpheme, since it looks the same. Nevertheless, in each of the two words it serves different functions (in the first case it forms the plural of the noun, while in the second it forms the 3SG present tense). Therefore it needs to be treated as two different morphemes (the morpheme *-s* used to form the plural of the nouns, and the morpheme *-s* used for the 3SG present tense).

2. If the same function (or meaning) is associated with two or more phonetic forms, these different forms all represent the same morpheme and the choice of form in each case is usually predictable based on phonological or other considerations.

For example, the prefixes *il-*, *im-*, *in-*, *ir-* (***il****legal*, ***im****possible*, ***in****tolerant*, ***ir****responsible*) serve the same function, so we can consider them to be allomorphs of the same morpheme. Moreover, the choice of the allomorph can be predicted based on phonological considerations: if the stem starts with a liquid, the prefix

will be *iX-*, where *X* is the first consonant of the stem; if the stem starts with a stop, the prefix will be *iN-*, where *N* is a nasal consonant that assimilates to the place of articulation from the first consonant in the stem.

Problem 5.1

Zulu (Vlad A. Neacșu, original)

Here are some words in Zulu and their English translations:

umdwebi	'painter'	*abazingeli*	'hunters'	*umbulali*	'killer'
abadwebi	'painters'	*zingela*	'to hunt'	*ababazi*	'carvers'

Problem 5.1a Translate into Zulu: 'to paint', 'hunter', 'killers', 'to kill', 'carver', 'to carve'.

Solution

We notice that we have three types of words: singular nouns, plural nouns and verbs which are semantically related to the nouns. Therefore, we can create the following table:

	Noun SG	Noun PL	Verb
'painter'	*umdwebi*	*abadwebi*	
'hunter'		*abazingeli*	*zingela*
'killer'	*umbulali*		
'carver'		*ababazi*	

From the table, we can easily deduce that the singular noun is formed using the circumfix *um- -i*, while the plural is formed with the circumfix *aba- -i*. The verb is formed by adding the suffix *-a*.

Another explanation, in order to avoid the idea of a circumfix, is that the singular and plural are formed using the prefixes *um-* and *aba-*, respectively, while the verb is formed by replacing the final vowel (*i*) with the vowel *a*.

Thus, the answers are:

Solution 5.1a

	Noun SG	Noun PL	Verb
‘painter’	*umdwebi*	*abadwebi*	***dweba***
‘hunter’	***umzingeli***	*abazingeli*	*zingela*
‘killer’	*umbulali*	***ababulali***	***bulala***
‘carver’	***umbazi***	*ababazi*	***baza***

Problem 5.2

Swedish (Vlad A. Neacșu, original)

Here are some words in Swedish and their English translations:

en flaska	‘a bottle’	*hunden*	‘the dog’	*hyllor*	‘shelves’
en stol	‘a chair’	*flaskorna*	‘the bottles’	*kattar*	‘cats’
en hund	‘a dog’	*stolarna*	‘the chairs’	*bilen*	‘the car’
flaskor	‘bottles’	*hundarna*	‘the dogs’	*hyllan*	‘the shelf’
stolar	‘chairs’	*en bil*	‘a car’	*katten*	‘the cat’
hundar	‘dogs’	*en hylla*	‘a shelf’	*bilarna*	‘the cars’
flaskan	‘the bottle’	*en katt*	‘a cat’	*hyllorna*	‘the shelves’
stolen	‘the chair’	*bilar*	‘cars’	*kattarna*	‘the cats’

Problem 5.2a Here are some more words in Swedish and their English translations:

en flicka = ‘a girl’ *bussarna* = ‘the buses’

Translate into Swedish: ‘the girl’, ‘girls’, ‘the girls’, ‘a bus’, ‘the bus’, ‘buses’.

Solution

Just as in the previous problem, we first notice the forms of the given nouns. We realise that each noun is given in four different forms: definite singular, indefinite singular, definite plural, and indefinite plural. Therefore, in order to facilitate the analysis of the data and notice the similarities between them, we make the following table:

Indef. SG	Def. SG	Indef. PL	Def. PL	Translation
en flaska	*flaskan*	*flaskor*	*flaskorna*	'bottle'
en stol	*stolen*	*stolar*	*stolarna*	'chair'
en hund	*hunden*	*hundar*	*hundarna*	'dog'
en bil	*bilen*	*bilar*	*bilarna*	'car'
en hylla	*hyllan*	*hyllor*	*hyllorna*	'shelf'
en katt	*katten*	*kattar*	*kattarna*	'cat'

Based on this table, we can consider the indefinite singular form as the stem (excluding the proclitic[2] article *en*). Moreover, we notice that the definite plural is derived from the indefinite plural by adding the suffix *-na*. We are left to discover how to form the definite singular and the indefinite plural.

We notice that the def. SG is formed using the suffixes *-n* or *-en*. Therefore, we need to discover in which context each of them is used. It could be either a semantic context, in which the variation is driven by the meaning of the word, or a phonological context, in which the choice of the allomorph is dictated by the phonological structure of the word. In this case, we can easily notice that the suffix *-en* is used if the base form ends in a consonant, while *-n* is used if the base form ends in a vowel, thus the distinction is purely phonological. Another explanation for the definite singular endings can include a phonological process, an elision, by considering the suffix *-en* as the sole suffix for def. SG, with the additional feature that $en \longrightarrow n \,/\, V\_$.

Applying the same thought process for the indefinite plural, we notice that the two suffixes are *-ar* and *-or*, the first being used if the stem ends in a consonant and the latter if the stem ends in a vowel. Moreover, if the stem ends in a vowel, the vowel is dropped when the suffix *-or* is added (alternatively: the suffix is *-ar* and *-V* + *-ar* ⟶ *-or* – i.e., when the suffix *-ar* is added after a vowel, it merges with it and results in the suffix *-or*). Thus, we can write the rules that govern the formation of these noun forms in Swedish (we used the abbreviations *S* = stem, *C* = consonant, *V* = vowel):

Indef. SG	Def. SG	Indef. PL	Def. PL
en S-C	*S-C-en*	*S-C-ar*	*S-C-arna*
en S-V	*S-V-n*	*S-or*	*S-orna*

[2]The term *proclitic* refers to the fact that the definite article is placed in front of the noun. This contrasts with the term *enclitic*, meaning it is placed after the noun.

Thus, the answers to the task are:

Solution 5.2a

'the girl' = *flickan*	'a bus' = *en buss*
'girls' = *flickor*	'the bus' = *bussen*
'the girls' = *flickorna*	'buses' = *bussar*

Problem 5.3

Māori (Vlad A. Neacșu, original)

Consider the following word forms in Māori:

Form I	Form II	Form III
inu	*inumia*	*inumaŋa*
hopu	*hopukia*	*hopukaŋa*
eke	*ekeŋia*	*ekeŋaŋa*
ɸera	*ɸerahia*	*ɸerahaŋa*
aɸi	*aɸitia*	*aɸitaŋa*
tupu	*tupuria*	*tupuraŋa*

Problem 5.3a Explain how the forms are constructed.

Solution

We can easily observe that, in order to obtain Form III from Form II we just replace the suffix *-ia* with *-aŋa*. Moreover, we notice that Form II derives from Form I by adding the suffix *-Cia*, where *C* is a consonant. The only thing left to do is figure out how the consonant is chosen.

The first thing we notice is that there are no two examples which use the same consonant; furthermore, the consonant does not seem to be related in any way to the structure of the word (to the phonological characteristics of the other sounds). Finally, we are sure that the choice of the consonant cannot be related to the meaning since the translations are not given in the data. Therefore, since the consonant does not seem to follow any pattern, we can consider it as being part of the stem, thus being a disfix.

If we consider that consonant as a disfix, we can easily figure out all the rules that generate the three forms:

Solution 5.3a Form I: elision of the last consonant of the stem;
Form II: add suffix *-ia*;
Form III: add suffix *-aŋa*.

Note

Linguistics problems featuring disfixes are extremely rare. Before considering the possibility of a disfix, make sure there is absolutely no correlation between that sound/morpheme and the shape or meaning of the stem.

5.3 Variables of the noun

When we talk about the *variables*[3] of the noun, we mean those features of the noun which can change in a linguistics problem and which it is best to identify when starting to solve the problem, in order to form an idea regarding what types of morphemes we are looking for. These variables can also be marked on other elements in the phrase/clause, e.g., on adjectives. This phenomenon is known as AGREEMENT.

The three common variables are: NUMBER (singular, dual, plural), GENDER (masculine, neuter, feminine etc.) which is a highly grammaticalised subtype of noun CLASS. The way in which we usually realise that gender is relevant in a linguistics problem is by identifying different allomorphs. Thus, if we notice that a set of allomorphs appears solely with some stems and another set with other stems, we can assume in that language there are two classes of nouns (two genders), each of them having their own characteristic morphemes. Class and classifiers will be further discussed in the next section.

[3] By using the term noun *variables*, we refer to the grammatical categories specific to English (number, case, etc.), as well as to other distinctions which might be relevant in other languages (usually, of a semantic nature).

Although all the grammatical categories of the noun are considered variables of the noun, not all variables are grammatical categories. For example, the distinction ±human or ±animate, to which we will refer in the following sections, is an important variable of the noun, but it cannot be considered a grammatical category.

The *number* variable denotes how many items the noun refers to. The simplest distinction is singular (referring to one item) and plural (referring to more than one item). However, different languages might have other distinctions, among which the most common are dual (two items), trial (three items), and paucal (referring to a relatively small number of items. This would typically be translated into English as 'a few' and usually refers to less than ten items). Of course, there are also languages which do not mark nouns for number.

An interesting case regarding noun plurals is encountered in Dagaare.[4] This language features an *inherent plurality* of the designated noun and, starting from it, it differentiates an unmarked form (a base form) and a marked form. Thus, if a noun represents an entity that is usually found alone ('forehead', 'hat'), the unmarked form will be represented by the singular (while the plural will be the marked form). However, if the noun is usually found in pairs ('leg', 'lung', 'shoe') or groups ('bee'), then the unmarked form is the plural (and the singular is denoted by the marked form).

There is one more variable of the noun that we have not mentioned yet: GRAMMATICAL CASE. Case is an important feature of nouns and indicates the role the noun plays in the sentence. In some languages, nouns can be marked for many cases; for example, Uralic languages, which are known for their large number of cases, can have more than 20 cases – nominative, genitive, partitive, accusative, inessive, elative, illative, adessive, ablative, additive, egressive, comitative, terminative, abessive, translative, allative, essive, instructive, instrumental, dative, causal, sublative, superessive, delative, temporal, sociative. Most of these cases will be translated into English using different prepositions, for example, the instrumental case (which shows the object with which the action is performed) is, usually, translated using the preposition 'with' or 'by' in English (*I write* ***with a pen*** or *I sew* ***by hand***). When solving a linguistics problem, assuming the instrumental is marked by the suffix *-ok* in the target language, we do not necessarily have to write "Instrumental = *-ok*", but rather we can simply write; "'with *X*' = *X-ok*", without using the name of the case.

Moreover, most of the cases of the Uralic languages are locatives (they show the location). The elative case can be translated using 'from $+M$', illative = 'in $+M$', allative = 'on $+M$', adessive = 'on $-M$', etc, where M denotes movement. Thus, the cases $+M$ are cases in which the object is moved in a direction (relative to the noun), while the $-M$ cases are those in which the objects are already in that location.

Let us consider, for example, the allative and adessive cases (both of them can be expressed in English using the preposition 'on', but they differ in terms of the

[4]This phenomenon was featured in a problem by Ethan Chi (NACLO 2021).

parameter $\pm M$). As such, in the sentence *The apple is on the table*, *on* links to the concept of adessive, since it shows a location without any movement involved. However, in the sentence *He put the apple on the table*, the apple is moved towards the table: in the beginning, the apple was not on the table, but at the end of the action it was; therefore, in this case, *on* links to the concept of allative. Notice that, because English has very little in the way of inflectional morphology, the equivalents of nouns that appear in the adessive or allative cases in other languages such as Finnish appear as the complements of prepositions in English. In other words, what Finnish expresses via the morphological structure of the noun, English expresses in its syntactic structure (i.e. as a prepositional phrase).

Because case is relevant in both morphology and syntax, we will discuss specific examples when they arise. For example, nominative, accusative, ergative, and absolutive, which we will discuss in Section 7.4.

Problem 5.4

Bulgarian (Kai Low Rui Hao & Martin Vasev, NACLO 2017)

Here are some sentences in Bulgarian (written in Latin script) and their English translations *in random order*:

1.	*Veshterăt nahrani maymunata.*	A.	'Your son watched you.'
2.	*Kamilata vărvya.*	B.	'The girl hugged the cat.'
3.	*Momicheto pregărna kotkata.*	C.	'You dressed yourself.'
4.	*Veshtitsata prokle kotkata.*	D.	'The cat scratched you.'
5.	*Kotkata prokle tvoya sin.*	E.	'You fed the son.'
6.	*Ti nahrani sina.*	F.	'The witch cursed the cat.'
7.	*Kotkata te odraska.*	G.	'The camel walked.'
8.	*Ti skochi.*	H.	'The cat cursed your son.'
9.	*Tvoyat sin te gleda.*	I.	'The wizard fed the monkey.'
10.	*Veshterăt pregărna edna kamila.*	J.	'The son dressed your baby.'
11.	*Ti se obleche.*	K.	'You jumped.'
12.	*Sinăt obleche tvoeto bebe.*	L.	'The wizard hugged a camel.'

Problem 5.4a Determine the correct correspondences.

Problem 5.4b Translate into English:

13. *Maymunata gleda tvoyata veshtitsa.*
14. *Tvoyata kamila obleche edno momiche.*

15. *Veshterăt se prokle.*
16. *Ti pregărna bebeto.*
17. *Ti vărvya.*
18. *Ti prokle edin veshter.*

Problem 5.4c Translate into Bulgarian:

19. 'The witch dressed you.'
20. 'The baby watched the girl.'
21. 'The monkey jumped.'
22. 'You hugged a son.'
23. 'Your son dressed a baby.'

ă ≈ 'u' in 'but'.

Solution

There are multiple possible starting points when approaching this problem, but one of the most common (and generally applicable) is to notice that there are only two sentences in Bulgarian which have two words (2 and 8). Therefore, we can assume that these are the simplest sentences and, most likely, correspond to the shortest sentences in English, i.e., those that have only a subject and a verb. Among all the English sentences, the only ones that follow this pattern are K and G. We have different ways to figure out which is which: we can use the similarity between the Bulgarian and English words (*kamilata* – 'camel') or we can notice that *ti* appears in two other sentences and *kamilata* does not occur in any other sentence, while in English we have got two other sentences that begin with 'you', but none that begin with 'camel'. Since neither of the two verbs ever occurs again in the data and the first word does, we discover that this one is the subject and the last word is the verb. Hence, we deduce that 2-G and 8-K and the word order is S-V (Subject-Verb).

From here, we can continue with the other two sentences which contain the subject *ti* = 'you', which are 6, 11 and C, E. In order to match them, we can notice that the last word in sentence 6 occurs (in a slightly changed form) as the first

word of sentence 12, so we can deduce it is a noun (since it is a subject in sentence 12), thus it cannot mean 'myself'. Therefore, 6-E and 11-C. Moreover, we understand that *sina* = 'son', *nahrani* = 'to feed', *obleche* = 'to dress'. Since *obleche* and *nahrani* occur one more time in another sentence, it follows that 1-I and 12-J (we can notice again the similar words *bebe* – 'baby'). Based on this information, we can easily match all the other sentences. We get:

1.	*Veshterăt nahrani maymunata.*	I.	'The wizard fed the monkey.'
2.	*Kamilata vărvya.*	G.	'The camel walked.'
3.	*Momicheto pregărna kotkata.*	B.	'The girl hugged the cat.'
4.	*Veshtitsata prokle kotkata.*	F.	'The witch cursed the cat.'
5.	*Kotkata prokle tvoya sin.*	H.	'The cat cursed your son.'
6.	*Ti nahrani sina.*	E.	'You fed the son.'
7.	*Kotkata te odraska.*	D.	'The cat scratched you.'
8.	*Ti skochi.*	K.	'You jumped.'
9.	*Tvoyat sin te gleda.*	A.	'Your son watched you.'
10.	*Veshterăt pregărna edna kamila.*	L.	'The wizard hugged a camel.'
11.	*Ti se obleche.*	C.	'You dressed yourself.'
12.	*Sinăt obleche tvoeto bebe.*	J.	'The son dressed your baby.'

Based on these correspondences, we can establish the structure of Bulgarian sentences. The word order is SVO (Subject-Verb-Object, if the object is a noun) or SOV (Subject-Object-Verb, if the object is a pronoun). Moreover, the determiner always precedes the noun (Det-Noun).

Next, we notice that the verb is invariable in these examples. The only phenomenon that we have not analysed yet is the structure of the noun phrase. We notice from the examples above that the noun can receive a suffix (equivalent to the definite article in the English examples) which can be *-ăt*, *-ta*, *-to* or *-a* or it can get a modifier which precedes it: indefinite article (*edin*, *edna*, *edno*) or 2SG possessive (*tvoyata*, *tvoyat*, *tvoya*, *tvoeto*).

Therefore, we can make a table with the different forms of each noun. Moreover, in order to get more data, we will also use the examples in task (b), which we can easily translate into English solely based on the word order and dictionary (we do not need to know the rules of noun declension). In order to cover all possibilities, we separate the subject and the object.

Noun	Subject	Object
‘wizard’	*-ăt*	*edin*
‘monkey’	*-ta*	*-ta*
‘camel’	*-ta* / *tvoyata*	*edna*
‘girl’	*-to*	*edno*
‘cat’	*-ta*	*-ta*
‘witch’	*-ta*	*tvoyata*
‘son’	*tvoyat* / *-ăt*	*tvoya* / *-a*
‘baby’		*tvoeto* / *-to*

From the table, we notice that the definite suffix (the definite article, which is marked as a suffix) has only three forms, and the nouns meaning ‘camel’ and ‘witch’, which use the suffix *-ta*, use the same form of 2SG possessive (*tvoyata*). Therefore, we can assume that there are three noun classes (in reality, they correspond to three genders: feminine, masculine, and neuter): Class 1 (using the definite suffix *-ăt* for the subject; it contains the nouns ‘wizard’, ‘son’), Class 2 (*-ta*: ‘camel’, ‘cat’, ‘witch’, ‘monkey’) and Class 3 (*-to*: ‘girl’). Based solely on the definite suffix, we cannot classify the noun *bebe* ‘baby’, but we notice that the object uses the same definite suffix. Therefore, we can assume that it also belongs to class 3. Now we can make a new table based on the noun class in order to see how different determiners are formed.

Class		Def. suff.	Indef. art.	2SG poss.
1	S	*-ăt*		*tvoyat*
1	O	*-a*	*edin*	*tvoya*
2	S	*-ta*		*tvoyata*
2	O	*-ta*	*edna*	*tvoyata*
3	S	*-to*		
3	O	*-to*	*edno*	*tvoeto*

We can notice that for Classes 2 and 3 the definite suffix is the same for subject and object, and Class 2 uses the same possessive for subject and object as well. We can assume that for both Class 2 and Class 3 there is no difference between subject and object (thus deducing the Class 3 definite suffix for subject which we need for task 20). Moreover, we can notice some similarities between the

definite suffix and the form of the possessive: it seems that the 2SG possessive is formed by adding the form of the definite suffix to the morpheme *tvoya* (with the additional feature that if the definite suffix begins with a vowel, it gets dropped and that Class 3 is an exception, since *tvoya* becomes *tvoe*). Nevertheless, in order to solve the problem, we need not understand how the possessive is formed and the table above is enough. Therefore, we can write the official solution and solve the tasks.

Another important thing to notice is that we have not attempted to find any rules based on which the nouns are split into the three classes. This should be done *only* if new nouns are given in the tasks and we need to classify them into one of the three classes.

Rules:

- Word order: SOV (O = pronoun) or SVO (O = noun), Det-Noun.
- Noun is divided into three classes: Class 1 ('wizard', 'son'), Class 2 ('camel', 'cat', 'witch', 'monkey'), Class 3 ('girl', 'baby').
- Determiners:

Class		Def. suff.	Indef. art.	2SG poss.
1	S	*-ăt*		*tvoyat*
1	O	*-a*	*edin*	*tvoya*
2	S	*-ta*		*tvoyata*
2	O	*-ta*	*edna*	*tvoyata*
3	S	*-to*		
3	O	*-to*	*edno*	*tvoeto*

Solution 5.4a 1. I. 3. B. 5. H. 7. D. 9. A. 11. C.
2. G. 4. F. 6. E. 8. K. 10. L. 12. J.

Solution 5.4b 13. 'The monkey watched your witch.'
14. 'Your camel dressed a girl.'
15. 'The wizard cursed himself.'
16. 'You hugged the baby.'

17. ‘You walked.’
18. ‘You cursed a wizard.’

Solution 5.4c

19. *Veshtitsata te obleche.*
20. *Bebeto gleda momicheto.*
21. *Maymunata skochi.*
22. *Ti pregărna edin sin.*
23. *Tvoyat sin obleche edno bebe.*

5.4 Classifiers

A classifier is a word (or an affix) which accompanies nouns and serves to classify them (thus talking about noun class). In most cases, the classification is made on semantic considerations. In Chinese, the classifier is mandatory between numeral and noun. Let us consider the following examples from Chinese:

‘three dogs’ = 三只狗　　‘three cats’ = 三只猫　　‘five cats’ = 五只猫

We notice that the first character represents the number (三 = 3, 五 = 5), while the last one represents the noun (狗 = ‘dog’, 猫 = ‘cat’). The middle word represents a classifier and, in this case, it refers to small animals. Other semantic considerations for the choice of the classifier are, usually, related to the shape: long objects, flat objects, round objects, etc.

Problem 5.5

Japanese (Vlad A. Neacșu, RoLO 2017)

Here are some phrases in Japanese and their English translations:

isha kyūnin	‘9 doctors’
gakusei sannin	‘3 students’
hon yonsatsu	‘4 books’
inu kyūhiki	‘9 dogs’
kami hachimai	‘8 sheets of paper’
magajin nanasatsu	‘7 magazines’

neko nihiki	'2 cats'
purēto yonmai	'4 plates'
ratto gohiki	'5 rats'
uma rokutō	'6 horses'
zō rokutō	'6 elephants'

Problem 5.5a Translate into English:*purēto rokumai*, *isha gonin*, *uma yontō*.

Problem 5.5b Here are some more Japanese words:

mangabon	'comic books'	*piza*	'pizzas'
kaeru	'frogs'	*ushi*	'cows'

Translate into Japanese: '2 comic books', '5 pizzas', '7 frogs', '9 cows'.

Solution

Comparing examples '9 doctors' and '9 dogs', we notice that the only part that is repeated is the morpheme *kyū-* from the beginning of the second word. We can deduce that the number is represented by the first part of the second word. Moreover, comparing the last two examples, ('6 horses', '6 elephants'), we notice that only the first word is different, so we can assume that this one represents the noun. Therefore, the phrase structure is Noun Number-X.

Based on this structure, we can infer all the numbers and nouns in Japanese, as follows:

Japanese numbers are:

2	*ni-*	5	*go-*	8	*hachi-*
3	*san-*	6	*roku-*	9	*kyū-*
4	*yon-*	7	*nana-*		

Note that in order to figure out the morpheme for '6', we need to check the examples in task (a) in order to figure out which part corresponds to the number and which to the particle X.

Nouns are:

isha = ‘doctor’
hon = ‘book’
kami = ‘sheet of paper’
neko = ‘cat’
ratto = ‘rat’
zō = ‘elephant’
gakusei = ‘student’
inu = ‘dog’
magajin = ‘magazine’
purēto = ‘plate’
uma = ‘horse’

The only morpheme left to analyse is *X*. We notice that it can have five different forms: *-nin* (in the phrases ‘9 doctors’, ‘5 students’), *-satsu* (‘4 books’, ‘7 magazines’), *-mai* (‘8 sheets of paper’, ‘4 plates’), *-hiki* (‘9 dogs’, ‘2 cats’, ‘5 rats’), and *-tō* (‘6 horses’, ‘6 elephants’). Therefore, we deduce that they represent classifiers and correspond to: *-nin* for humans, *-satsu* for bound materials, *-mai* for flat objects, *-hiki* for small animals and *-tō* for large animals. Now we can write the rules and answer the tasks.

Rules:

- Structure: Noun Number–Class
- Class:
  - *-nin* ⟶ humans
  - *-satsu* ⟶ prints
  - *-mai* ⟶ flat objects
  - *-hiki* ⟶ small animals
  - *-tō* ⟶ large animals

Solution 5.5a *purēto rokumai* = ‘6 plates’
isha gonin = ‘5 doctors’
uma yontō = ‘4 horses’

Solution 5.5b ‘2 comic books’ = *mangabon nisatsu*
‘5 pizzas’ = *piza gomai*
‘7 frogs’ = *kaeru nanahiki*
‘9 cows’ = *ushi kyūtō*

5.5 Reduplication

Reduplication represents the formation of new words by partially or totally doubling a morpheme or a part of a word. An example of total reduplication is plural formation in Indonesian, as we can see in the following examples: *rumah* ('house') – *rumahrumah* ('houses'), *ibu* ('mother') – *ibuibu* ('mothers'), *lalat* ('fly') – *lalatlalat* ('flies').

We can notice that in order to form the plural we simply repeat the whole word, so we are talking about a process of total reduplication.

In linguistics problems, total reduplication is seldom encountered, since it is easy to spot and analyse; therefore, in most cases, partial reduplication is preferred. The following examples show two such processes:

Marshallese:	*kagir* ('belt')	*kagirgir* ('to wear a belt')
	takin ('socks')	*takinkin* ('to wear socks')
Samoan:	*savali* ('he walks')	*savavali* ('they walk')
	alofa ('he loves')	*alolofa* ('they love')

We can easily notice that in the first example the last syllable is reduplicated, while in the second example, the second (or penultimate) syllable is reduplicated. Moreover, we notice that the reduplication process can have different roles, such as forming the plural of the noun (in Indonesian), forming the plural of the verb (transforming a verb from 3SG to 3PL in Samoan) or even transforming a noun into a verb (the pairs *X* – 'to wear' *X*, in Marshallese).

Note

An alternative explanation for the aforementioned reduplication process can be "repeating the last three letters" (in Marshallese). Nevertheless, "the last three letters" do not represent a phonological or morphological unit of any sort, so this explanation is not very rigorous from a linguistic perspective.

5.6 Suppletion

Another important phenomenon that might be featured in linguistics problems is suppletion. This generally refers to the situation in which one stem (so one morpheme with a semantic component) has two or more completely different forms (in most cases due to different etymologies). Suppletion can be noticed to a lesser extent in most languages, including English (*good – better*, *bad – worse*, *go – went*), French (*être – sommes*, *aller – vais*), Spanish (*ir – voy*), German (*gut – besser*), etc. We can notice that, in each of the cases, it is not simply a mutation of the stem (as for example in English: *goose – geese* or *mouse – mice*), but rather the two forms are completely distinct.

This phenomenon seldom appears in linguistics problems, which makes it extremely hard to notice and, similar to the disfixes, it must be used with caution since there might only be some phonological transformations.

5.7 Expressing possession

A common way of marking possession in the world's languages is to use the genitive case. This is usually expressed by a morpheme attached to the noun (and, sometimes, its dependents) that functions as the possessor.

In certain languages, often ones belonging to the Austronesian family, possession can be marked by an affix or an independent word. In these cases, the word used is a classifier (similar to those defined above). Moreover, in a multitude of languages, there is a dichotomy between alienable and inalienable possession. *Inalienable* possession refers to those objects one owns and which cannot be borrowed/lent or taken (in most cases, they include body parts and family members), while *alienable* possession refers to all the other cases. In most languages where this distinction is relevant, inalienable possession is formed by attaching the possessor morpheme directly to the noun, while alienable possession is formed using possession classes.

Problem 5.6

Fijian (Viktoria Papp, UKLO 2018)

Here are some phrases in Fijian and their English translations:

1.	*na uluqu*	‘my head’
2.	*na nona wau*	‘her weapon (she owns)’
3.	*na memunī bia*	‘your$_{\text{PL}}$ beer’
4.	*na kemudrau itukutuku*	‘your$_{\text{DU}}$ story (about you two)’
5.	*na nona motokaa*	‘her car’
6.	*na meda tī*	‘our$_{\text{INCL}}$ tea’
7.	*na kelemu*	‘your$_{\text{SG}}$ belly’
8.	*na nona dio*	‘her oyster (she’ll sell)’
9.	*na kequ uvi*	‘my yam’
10.	*na noqu itukutuku*	‘my story (I tell)’
11.	*na watiqu*	‘my spouse’
12.	*na kemunī vuaka*	‘your$_{\text{PL}}$ pig (you’ll eat)’
13.	*na nomu kato*	‘your$_{\text{SG}}$ basket’
14.	*na tamana*	‘her father’
15.	*na memudrau dio*	‘your$_{\text{DU}}$ oyster (you’ll slurp)’
16.	*na nodra vuaka*	‘their pig (they raise)’
17.	*na keda wau*	‘our$_{\text{INCL}}$ weapon (we’ll be hit with)’
18.	*na kedra raisi*	‘their rice’

Problem 5.6a Now the Fijian words are given to you. Your task is to translate the phrase in the table into Fijian:

	Fijian	English	English phrase to translate
19.	*uto*	‘heart’	‘my heart’
20.	*yaqona*	‘kava’	‘her kava (she’s drinking)’
21.	*yaqona*	‘kava’	‘her kava (drunk in her honour)’
22.	*draunikau*	‘witchcraft’	‘my witchcraft (used on/against me)’
23.	*draunikau*	‘witchcraft’	‘your$_{\text{DU}}$ witchcraft (you’re making)’
24.	*dali*	‘rope’	‘your$_{\text{SG}}$ rope (you own)’
25.	*dali*	‘rope’	‘your$_{\text{PL}}$ rope (restraining you)’
26.	*ika*	‘fish’	‘your$_{\text{DU}}$ fish’
27.	*wai*	‘water’	‘your$_{\text{PL}}$ water’
28.	*luve*	‘child’	‘her child’
29.	*waqa*	‘canoe’	‘our$_{\text{INCL}}$ canoe’
30.	*yapolo*	‘apple’	‘their apple (they’ll sell)’
31.	*maqo*	‘mango’	‘their mango (for drinking)’

Problem 5.6b Explain your translation 21. (Why did you translate it this way?)

Problem 5.6c The word for 'coconut' is *niu*. List all the ways to say 'my coconut' and explain what they could mean.

'Weapon' refers to a club-like tool. A 'yam' is an edible starchy root. 'Kava' is a ceremonial drink.

The subscripts SG, DU, PL refer to singular (one person), dual (two persons), and plural (more than two persons), respectively. The subscript INCL means inclusive ('our$_{\text{INCL}}$' = belonging to me, you, and them, in contrast with 'our$_{\text{EXCL}}$' = belonging to me and them, but not you).

Solution

The first step we need to take is to determine the structure of the Fijian phrases. We can easily notice that all examples start with the word *na* followed by either one or two words. Separating the structures that contain a single word (i.e., the Fijian words for 'my head', 'your$_{\text{SG}}$ belly', 'my husband', 'her father'), we notice that all of them represent inalienable possessions. We therefore expect them to have the possessor marked directly on the noun, as an affix.

Indeed, comparing the structures 'my head' = *na uluqu* and 'my husband' = *na watiqu*, we discover that they both share the suffix *-qu*, which we then can deduce to mean 1SG possession. Therefore, for the inalienable possession, the phrase structure is *na* Noun–Poss.

In order to determine the structure of the rest of the phrases, we compare examples that contain the same noun (for example, 8 and 15, both containing the noun which means 'oyster'). We notice that both of them have the last word in common (*dio*), so the last word represents the noun, the possessed. Moreover, comparing examples 9 and 10 (both containing the possessive 'my'), we notice that they have in common the suffix *-qu* attached to the second word. Consequently, we can deduce that the structure of the alienable possession in Fijian is *na X*–Poss Noun. From here, we can easily determine the possession suffixes (applying similar processes as in Problem 5.5 where we had to separate the number from the classifier). We obtain:

1SG = *-qu*	3SG = *-na*	2DU = *-mudrau*	3PL = *-dra*
2SG = *-mu*	1PL incl. = *-da*	2PL = *-munī*	

Moreover, we understand that we have only three possession classes, marked by *no-*, *-me-*, and *ke-* (excluding the inalienable possession). In order to see how each classifier is used, we split the given phrases according to the classifier they use:

ke-	*me-*	*no-*
'my yam'	'your$_{PL}$ beer'	'her weapon (she owns)'
'their rice'	'our tea'	'her car'
'our weapon (we'll be hit with)'	'your$_{DU}$ oyster (you'll slurp)'	'her oyster (she'll sell)'
'your$_{PL}$ pig (you'll eat)'		'my story (I tell)'
'your$_{DU}$ story (about you two)'		'your$_{SG}$ basket'

We can easily notice that the morpheme *me-* is used for things that are to be drunk. Moreover, it seems that *no-* represents the class for owned objects, in general. The last class is *ke-* and, based on the fact that we already have a class for things that are drunk, we expect to also have a class for things that are eaten.

Indeed, the class *ke-* includes 'my yam', 'your$_{PL}$ pig (for eating)', and 'their rice', all of them being edible and meant to be eaten. On the other hand, we have two more structures ('our$_{INCL}$ weapon (we'll be hit with)' and 'your$_{DU}$ story (about you two)').

In order to figure out what these two have in common, we can also notice the pair of examples that contain the noun weapon: class *ke-* 'we'll be hit with', but class *no-* for 'she owns'. Based on these, we realise that class *ke-* includes not only food, but also things that do not belong to us directly, but affect us (the weapon is not ours, but will be used against us; the story is not yours, but it is about you two, etc.).

Thus, we can write the rules and solve the tasks.

Rules: Structure:

- Inalienable possession (kinship, body parts): *na* Noun–Poss

- Alienable possession: *na C*–Poss Noun

Poss = possessive (suffix):

1SG = *-qu*	3SG = *-na*	2DU = *-mudrau*	3PL = *-dra*
2SG = *-mu*	1PL incl. = *-da*	2PL = *-munī*	

C = class:

- *ke-* = food and objects we do not own, but affect us.
- *me-* = drinks
- *no-* = otherwise (owned objects)

Problem 5.6a

19. *na utoqu*
20. *na mena yaqona*
21. *na kena yaqona*
22. *na kequ draunikau*
23. *na nomudrau draunikau*
24. *na nomu dali*
25. *na kemunī dali*
26. *na kemudrau ika*
27. *na memunī wai*
28. *na luvena*
29. *na noda waqa*
30. *na nodra yapolo*
31. *na medra maqo*

Problem 5.6b In example 21 we used the classifier *ke-*, since the action is indirectly reflected towards the person. It is not she who drinks the kava, but someone else drinks it in her honour.

Problem 5.6c Since a coconut can clearly not be an inalienable possession, there are three different possible translations:

na kequ niu = 'my coconut' (for eating / I'll be hit with)
na mequ niu = 'my coconut' (for drinking)
na noqu niu = 'my coconut' (I own / I sell)

Problem 5.7

Ancient Greek (Todor Tchervenkov, NACLO 2007)

Here are some phrases in Ancient Greek (in a Latin-based transcription) and their English translations *in random order*:

1.	*ho tōn hyiōn dulos*	A.	‘the donkey of the master’
2.	*hoi tōn dulōn cyrioi*	B.	‘the brothers of the merchant’
3.	*hoi tu emporu adelphoi*	C.	‘the merchants of the donkeys’
4.	*hoi tōn onōn emporoi*	D.	‘the sons of the masters’
5.	*ho tu cyriu onos*	E.	‘the slave of the sons’
6.	*ho tu oicu cyrios*	F.	‘the masters of the slaves’
7.	*ho tōn adelphōn oicos*	G.	‘the house of the brothers’
8.	*hoi tōn cyriōn hyioi*	H.	‘the master of the house’

Problem 5.7a Determine the correct correspondences.

Problem 5.7b Translate into Ancient Greek: ‘the houses of the merchants’, ‘the donkeys of the slave’.

ō denotes a long *o*.

Solution

Step 1. We notice that all Ancient Greek phrases have the structure [*ho*/*hoi*] [*tu*/*tōn*] *X Y*, where *X* and *Y* represent the two nouns (the possessor and the possessed). Moreover, we notice that these nouns change their form. Furthermore, the first two words (which probably represent articles or possession markers) each have two forms. Checking the English translations, we notice that nouns appear both as singular and plural. Therefore, we can assume that the two markers from the beginning of the phrase change form, agreeing with the number of the noun.

Step 2. We make a frequency table based on the stem of the nouns (for now, we disregard the endings, which are variable).

Greek	Freq.	English	Freq.
hyi-	2	‘donkey’	2
dul-	2	‘master’	4
cyri-	4	‘brother’	2
empor-	2	‘merchant’	2
adelph-	2	‘son’	2
on-	2	‘slave’	2
oic-	2	‘house’	2

Since both in Greek and in English we have only one noun that appears four times (all the rest appearing only two times each), we infer that *cyri-* = 'master'.

Step 3. We separate all phrases which contain the noun master:

2.	*hoi tōn dulōn cyrioi*	A.	'the donkey of the master'
5.	*ho tu cyriu onos*	D.	'the sons of the masters'
6.	*ho tu oicu cyrios*	F.	'the masters of the slaves'
8.	*hoi tōn cyriōn hyioi*	H.	'the master of the house'

In Greek, the nouns that appear together with 'master' are *dul-*, *on-*, *oic-*, *hyi-*, while in English they are 'donkey', 'son', 'slave', 'house'. The only two nouns that do not appear here are 'merchant' and 'brother', so these must correspond to the two Greek nouns that do not occur: *empor-* and *adelph-*. Moreover, we notice that phrase 3 in Greek and phrase B in English contain both of these nouns. Therefore, we deduce the correspondence 3-B.

Step 4. We separate the sentences that contain the nouns 'brother' or 'merchant'.

3.	*hoi tu emporu adelphoi*		'the brothers of the merchant'
4.	*hoi tōn onōn emporoi*	C.	'the merchants of the donkeys'
7.	*ho tōn adelphōn oicos*	G.	'the house of the brothers'

Note

The fact that we didn't mention the letter in front of structure B. means that this phrase is already matched, its Greek correspondence being phrase 3.

Separating again the nouns which appear in these structures, we deduce that 'donkey' and 'house' are *on-* and *oic-* though we do not know which is which, and similarly with other word pairs.

Step 5. Recap.

Based on the information above, we know that:

- *cyri-* = ‘master’
- {*on-*/*oic-*} = {‘donkey’/‘house’}
- {*empor-*/*adelph-*} = {‘merchant’/‘brother’}
- {*dul-*/*hyi-*} = {‘son’/‘slave’}.

Moreover, we notice that there is one phrase in which the nouns ‘slave’ and ‘son’ co-occur, therefore 1-E.

Step 6. Based on this information, we split the phrases into subgroups:

1.	*ho tōn hyiōn dulos*	‘the slave of the sons’
3.	*hoi tu emporu adelphoi*	‘the brothers of the merchant’
4.	*hoi tōn onōn emporoi*	C. ‘the merchants of the donkeys’
7.	*ho tōn adelphōn oicos*	G. ‘the house of the brothers’
5.	*ho tu cyriu onos*	A. ‘the donkey of the master’
6.	*ho tu oicu cyrios*	H. ‘the master of the house’
2.	*hoi tōn dulōn cyrioi*	D. ‘the sons of the masters’
8.	*hoi tōn cyriōn hyioi*	F. ‘the masters of the slaves’

We know that phrases 1 and 3 are already matched and that phrases 4 and 7 correspond to C and G (although not necessarily in this order), phrases 5/6 with A/H, and 2/8 with D/F.

Moreover, we notice that phrases 5/6 use the same articles (the first two words are identical), while in the English translations, all nouns are singular. Therefore, we deduce that *ho* and *tu* represent the singular form, while the other two (*hoi* and *tōn*) represent the plural form (which is confirmed by phrases 2 and 8 which use these articles, and in their English translations all nouns are plural).

Step 7. We create a new frequency table for the four articles, knowing that they need to correspond to a combination of singular/plural and possessor/possessed.

Greek	Freq.	English	Freq
ho	4	SG, possessor	3
hoi	4	PL, possessor	5
tu	3	SG, possessed	4
tōn	5	PL, possessed	4

From the table, we can immediately deduce that *tu* is used for a singular possessor, and *tōn* when the possessor is plural. We are left with *ho*/*hoi* for the possessed. On the other hand, we already know that *ho* is used for the singular (from step 6). Therefore, we can make a table with all the forms of the article:

	SG	PL
possessor	*tu*	*tōn*
possessed	*ho*	*hoi*

Furthermore, knowing this, we can finish matching the phrases. Looking back at the table in step 6, in the pair 4-7 we have a singular possessed and a plural possessed (based on the articles), therefore 4 corresponds to C and 7-G. We get:

1. *ho tōn hyiōn dulos*	'the slave of the sons'
3. *hoi tu emporu adelphoi*	'the brothers of the merchant'
4. *hoi tōn onōn emporoi*	'the merchants of the donkeys'
7. *ho tōn adelphōn oicos*	'the house of the brothers'
5. *ho tu cyriu onos*	A. 'the donkey of the master'
6. *ho tu oicu cyrios*	H. 'the master of the house'
2. *hoi tōn dulōn cyrioi*	D. 'the sons of the masters'
8. *hoi tōn cyriōn hyioi*	F. 'the masters of the slaves'

Now we can deduce, from 3 and 4, that *empor-* = 'merchant' and then we can easily make the rest of the correspondences. Moreover, we deduce that the possessed is placed after the possessor. Therefore, the word order in the Ancient Greek phrase is:

[Art. possessed] [Art. possessor] Possessor Possessed

We notice an interesting phenomenon, that the word order between the noun and its corresponding article is "enclosed" (the possessor together with its article are placed in the middle and enclosed/surrounded by the possessor and its article).

Step 8. The last step is to figure out noun declension in Ancient Greek. To do so, we can make a table with the different forms of all nouns, based on number (singular/plural) and its role (possessor/possessed):

	Possessor		Possessed	
Noun	SG	PL	SG	PL
‘master’	*cyriu*	*cyriōn*	*cyrios*	*cyrioi*
‘donkey’		*onōn*	*onos*	
‘brother’		*adelphōn*		*adelphoi*
‘merchant’	*emporu*			*emporoi*
‘son’		*hyiōn*		*hyioi*
‘slave’		*dulōn*	*dulos*	
‘house’	*oicu*		*oicos*	

From this table, we can easily notice that each form has its own characteristic suffix:

	SG	PL
possessor	*-u*	*-ōn*
possessed	*-os*	*-oi*

Note

Note that this approach is a basic one: it is based strictly on logical observations and not necessarily on linguistic intuition. Another more solid starting point would have been to notice the fact that there are similarities between the article and the noun suffix (*tōn* – *-ōn*, *hoi* – *-oi*, *tu* – *-u*). This directly points towards the word order, which is the core phenomenon of the problem.

Based on these, we can write the rules and solve the task.

Rules:

- Structure: [Art. possessed] [Art. possessor] Possessor Possessed
- Articles:

	SG	PL
possessor	*tu*	*tōn*
possessed	*ho*	*hoi*

- Noun endings:

	SG	PL
possessor	*-u*	*-ōn*
possessed	*-os*	*-oi*

Solution 5.7a 1. E. 2. F. 3. B. 4. C. 5. A. 6. H. 7. G. 8. D.

Solution 5.7b 'the houses of the merchants' = *hoi tōn emporōn oicoi*
'the donkeys of the slave' = *hoi tu dulu onoi*

5.8 Colour terms

A special focus is given to the problems that include noun phrases in which the adjectives are colour terms. For example, let us look at the following problem:

Problem 5.8

Colours (Vlad A. Neacșu, RoLO 2018)

Brent Berlin and Paul Kay have studied colour terms of more than 100 languages by travelling the world and asking the locals to describe photos of different colours (Berlin & Kay 1969). After examining the results, they concluded that colour perception in all of these languages is governed by the same law.

Here are the colour terms in 12 of the languages the two scientists examined:[5]

[5] *Source:* Adapted from a problem by Ksenia Gilyarova (Elementy).

English	Fitzroy River	Nupe	Upper Pyramid	Ibo	Tzeltal	Hanunó'o
'white'	*bura*	*bókùn̩*	__(1)__	*nzu*	*sak*	*(ma)biru*
'blue'	*guru*	*dòfa*	__(2)__	__(3)__	*yaš*	__(4)__
'yellow'	*kalmur*	*wọn̩jin̩*	__(5)__	*odo*	*k'an*	*(ma)rara?*
'brown'	__(6)__	*dzúfú*	*mola*	*uhie*	__(7)__	*(ma)rara?*
'black'	__(8)__	*ẓìkò*	__(9)__	*oji*	*?ihk'*	*(ma)lagti?*
'red'	*kiran*	*dzúfú*	__(10)__	__(11)__	*cah*	__(12)__
'green'	__(13)__	*álígà*	*muli*	*oji*	*yaš*	*(ma)latuy*

English	Bari	Jalé	Hausa	Nasioi	Daza	Ibibio
'white'	*-kwe*	*hóló*	*fǎri*	*kakara*	*cuo*	*àfíá*
'blue'	*-murye*	*siŋ*	*shuḍi*	*mutaŋa*	*zẹdẹ*	__(14)__
'yellow'	*-forong*	__(15)__	*nawaya*	__(16)__	*mini*	*ńdàídàt*
'brown'	*-jere*	__(17)__	*ja*	__(18)__	*maaḍo*	__(19)__
'black'	*-rnö*	*siŋ*	*bāḳi*	__(20)__	*yasko*	*ébubit*
'red'	*-tor*	*hóló*	__(21)__	*erereŋ*	__(22)__	*ńdàídàt*
'green'	*-ngem*	__(23)__	*algashi*	__(24)__	__(25)__	*àwàwà*

Problem 5.8a Fill in the blanks.

Problem 5.8b Below are some colour terms in three other languages that follow the hypothesis of Berlin & Kay:

a. Urhobo: 'black' = *ɔbyibi*, 'brown' = *ɔBaBare*, 'green' = *ɔbyibi*, 'yellow' = *5do*;

b. N'gombe: 'white' = *bopu*, 'red' = *bopu*;

c. Tanna Island: 'yellow' = *laulau*, 'black' = *rapen*, 'brown' = *laulau*, 'blue' = *ramimera*.

For each language, specify which of the 12 languages above it most resembles. Explain your answer.

Problem 5.8c Here are some colour terms in two more languages:

d. Acehnese: 'white' = *iĵu*, 'red' = *pirã*, 'green' = *prãna*, 'yellow' = *iĵu*, 'blue' = *prãna*;

e. Alabama: ‘green’ = *okchakko*, ‘red’ = *homma*, ‘brown’ = *laana*, ‘blue’ = *okchakko*, ‘yellow’ = *laana*.

Explain why they do not follow the hypothesis of Berlin & Kay.

Problem 5.8d Formulate the hypothesis of Berlin & Kay.

Solution

Step 1. Looking closely at the table, we notice that almost all the languages in the table (save for English and Bari) have some colour terms which share the same name. Therefore, we deduce that, in order to fill in the blanks, we need to reuse some of the words in that language which are already given; in other words, we need to figure out which colour terms are translated identically in that language. This is also signalled by the fact that there does not seem to be any morphological process going on which allows us to derive some colour terms from others and there are few or no common morphemes.

In order to do so, we classify each language based on the number of distinct colour terms (i.e., colour terms which have different names in that language) and reorder the table in descending order based on the number of distinct colour terms.

	7 terms	6 terms		5 terms	
English	Bari	Nupe	Hausa	Tzeltal	Daza
‘white’	*-kwe*	*bókùn̩*	*fǎri*	*sak*	*cuo*
‘blue’	*-murye*	*dòfa*	*shuḍi*	*yaš*	*zẹ́dẹ́*
‘yellow’	*-forong*	*wọn̩jin̩*	*nawaya*	*k’an*	*mini*
‘brown’	*-jere*	*dzúfú*	*ja*		*maaḍo*
‘black’	*-rnö*	*ǹ̩kò*	*bāḳi*	*ʔihk’*	*yasko*
‘red’	*-tor*	*dzúfú*		*cah*	
‘green’	*-ngem*	*álígà*	*algashi*	*yaš*	

	4 terms			
English	Fitzroy River	Ibo	Hanunó'o	Ibibio
'white'	*bura*	*nzu*	*(ma)biru*	*àfíá*
'blue'	*guru*			
'yellow'	*kalmur*	*odo*	*(ma)rara?*	*ńdàídàt*
'brown'		*uhie*	*(ma)rara?*	
'black'		*oji*	*(ma)lagti?*	*ɛ́bubit*
'red'	*kiran*			*ńdàídàt*
'green'		*oji*	*(ma)latuy*	*àwàwà*

	3 terms	2 terms	
English	Nasioi	Jalé	Upper Pyramid
'white'	*kakara*	*hóló*	
'blue'	*mutaŋa*	*siŋ*	
'yellow'			
'brown'			*mola*
'black'		*siŋ*	
'red'	*erereŋ*	*hóló*	
'green'			*muli*

Step 2. In Nupe (which has six distinct colours), we notice that the same word is used for both 'brown' and 'red'. In Tzeltal (which has five terms), the same word is used for both 'blue' and 'green', etc.

We can assume that languages with the same number of colour terms will behave identically. We conclude that:

- if a language has six colour terms, 'red' = 'brown'
- if a language has five colour terms, 'red' = 'brown' and 'blue' = 'green'

Step 3. Checking the languages with four colour terms, we notice that we have two options:

- for Fitzroy River and Ibo: 'red' = 'brown', 'green' = 'blue' = 'black'
- for Ibobo and Hanunóo: 'red' = 'brown' = 'yellow', 'green' = 'blue'

So, in the case of languages with only four colour terms, we have two alternatives: 'black' is named identically with 'blue' and 'green', or 'yellow' is named identically with 'red' and 'brown'.

Step 4. The rest of the problem becomes trivial and we deduce that:

- for languages with three colour terms: 'black' = 'green' = 'blue', 'yellow' = 'brown' = 'red'
- for languages with two terms: 'black' = 'green' = 'blue', 'yellow' = 'brown' = 'red' = 'white'

Therefore, we can solve the tasks:

Solution 5.8a

(1)	*mola*	(8)	*guru*	(15)	*hóló*	(22)	*maaḍo*
(2)	*muli*	(9)	*muli*	(16)	*erereŋ*	(23)	*siŋ*
(3)	*oji*	(10)	*mola*	(17)	*hóló*	(24)	*mutaŋa*
(4)	*(ma)latuy*	(11)	*uhie*	(18)	*erereŋ*	(25)	*zẹdẹ*
(5)	*mola*	(12)	*(ma)rara?*	(19)	*ńdàídàt*		
(6)	*kiran*	(13)	*guru*	(20)	*mutaŋa*		
(7)	*cah*	(14)	*àwàwà*	(21)	*ja*		

Solution 5.8b

a. Urhobo is similar to Fitzroy River and Ibo.

b. N'Gombe is similar to Upper Pyramid and Jalé.

c. Tanna Island is similar to Hanunó'o and Ibibo.

Solution 5.8c

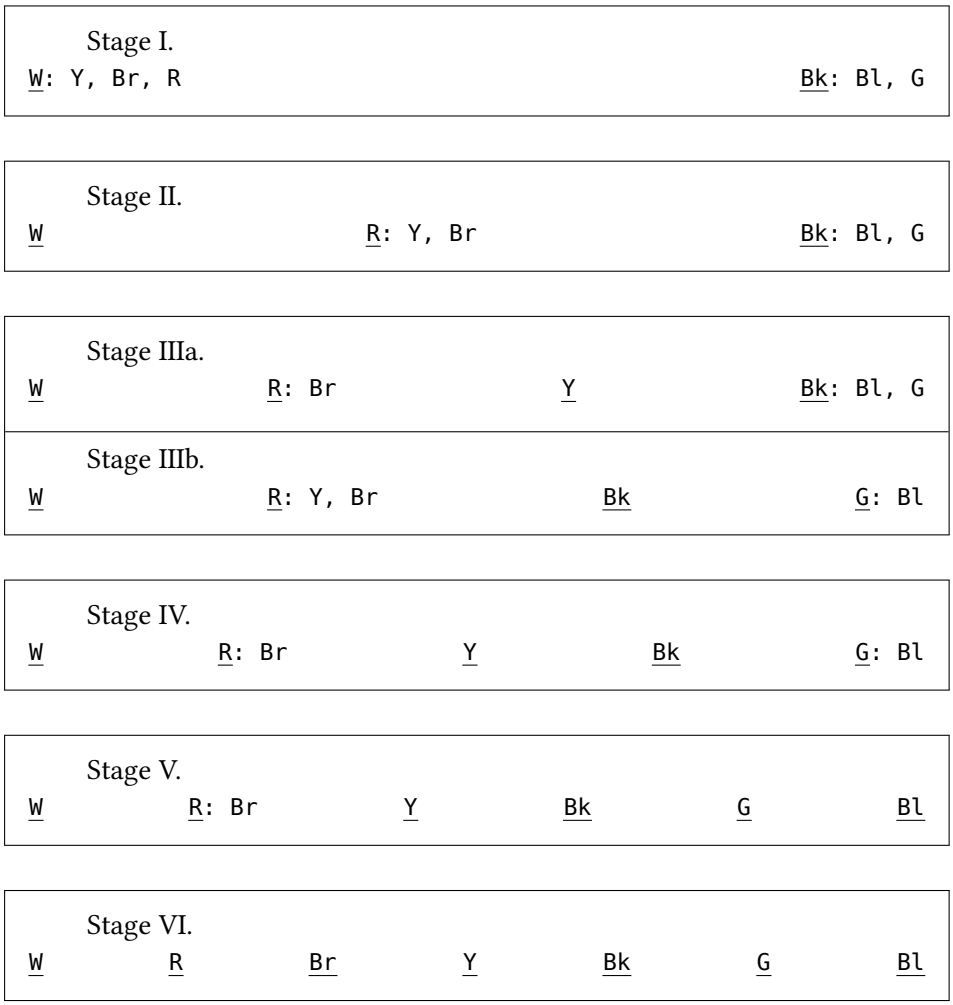
d. If the same word is used for both 'white' and 'yellow', there should only be two colour terms in that language, so 'red' should also be translated like 'white' and 'yellow'.

e. If 'green' and 'blue' use the same word, the language must have five terms, so 'brown' and 'red' should be translated identically.

Solution 5.8d Berlin & Kay hypothesised that the name of the colour terms in each language can be deduced by the number of colour terms each language has.

If we use the notation `W` = 'white', `Bk` = 'black', `Br` = 'brown', `R` = 'red', `G` = 'green', `Y` = 'yellow', `Bl` = 'blue', Berlin & Kay proposed six stages of evolution of colour terms in a language.

In the scheme below, the underlined colour terms exist in each language's vocabulary, while those following are perceived as being identical to the one underlined. For example, the notation G: Bl, Bk means that in the respective language, there is a word for 'green', while blue and black share the same term and are perceived as shades of green.



We notice that the third stage has two variants: either 'yellow' splits away (from 'red' and 'brown'), *or* 'blue' and 'green' (together) split from 'black'; while in the fourth stage, both splits take place: 'yellow' splits (from 'red') *and* 'blue' and 'green' from 'black'.

This fact can be useful in linguistics problems, considering the following aspect: historically speaking, it is likely that some languages had a limited number of colour terms and, in time, due to contact with other nations, they might have borrowed terms for other colours. For this reason, it is likely that, in some languages, some colour terms behave differently from others. In other words, it is likely that there are some basic colours (inherent to the language, which, according to the aforementioned hypothesis, should contain 'white', 'black', and 'red'), which follow the usual declension of the language and other colour terms (borrowings from other languages) with a different flexion (usually diminished, or even inflexible).

For example, in Romanian, the basic colour terms have a full set of inflected forms, with four different forms ('white': *alb – albă – albi – albe*, 'black': *negru – neagră – negri – negre*), while other, new terms, are usually invariable ('pink': *roz*, 'brown': *maro*, 'turquoise': *turcoaz*, etc.).

5.9 Practice problems

Problem 5.9

Ulwa (Peter Arkadiev, TurLom 2003)

Here are some words in Ulwa and their English translations *in random order*:

suulu, *suukilu*, *suumanalu*, *mismatu*, *miskatu*, *onkinayan*, *onkayan*, *onyan*

'bow', 'your$_{SG}$ cat', 'my dog', 'our bow', 'his cat', 'dog', 'his bow', 'your$_{PL}$ dog'

Problem 5.9a Determine the correct correspondences.

Problem 5.9b Translate into English: *suumalu* and *miskanatu*.

Problem 5.9c Translate into Ulwa: 'cat', 'my cat', 'your$_{SG}$ bow', 'their bow'.

Problem 5.10

Palauan (Michael Salter, NACLO 2018)

Here are some phrases in Palauan and their English translations:

eru ęl buil	'2 months'	*kltiu ęl hong*	'9 books'
ede ęl sils	'3 days'	*kllolem ęl lius*	'6 coconuts'
tede ęl chad	'3 people'	*teai ęl ngalęk*	'8 children'
kllolem ęl malk	'6 chickens'	*ongeru ęl buil*	'February'
teim ęl sensei	'5 teachers'	*ongede ęl ureor*	'Wednesday'
eim ęl rak	'5 years'	*etiu ęl klębęse*	'9 nights'
		tęruich me a tede ęl buik	'13 boys'
		tęruich me a euid ęl sikang	'17 hours'

Problem 5.10a Translate into English:

1. *telolem ęl sensei*
2. *tęruich me a etiu ęl buil*
3. *tęruich me a ongeru ęl buil*
4. *ongeim ęl ureor*

Problem 5.10b Translate into Palauan:

5. '8 days'
6. '19 people'
7. '7 teachers'
8. 'June'
9. 'August'

Problem 5.10c For each of the following, write the Palauan word that would be used to translate the word '3':

10. '**3** hours'
11. '**3** girls'
12. '**3** dolphins'

Problem 5.11

Norwegian (Babette Verhoeven-Newsome, NACLO 2017)

Here are some sentences in Norwegian and their English translations *in random order*:

1.	*Bussen stanser her.*	A.	'A woman has the apple.'
2.	*Jeg har en bil.*	B.	'I have an apple.'
3.	*Bilen stanser her.*	C.	'The bus stops here.'
4.	*Jeg har eplet.*	D.	'The woman has cars.'
5.	*Jeg har et eple.*	E.	'The car stops here.'
6.	*En kvinne har eplet.*	F.	'I have buses.'
7.	*Kvinna har biler.*	G.	'The woman has the cars.'
8.	*Kvinna har bilene.*	H.	'I have the apple.'
9.	*Jeg har busser.*	I.	'The women stop here.'
10.	*Kvinnene stanser her.*	J.	'I have a car.'

Problem 5.11a Determine the correct correspondences.

Problem 5.11b Nouns in Norwegian can belong to one of three classes: masculine, feminine, or neuter. The class determines how the noun can be used with determiners (words such as 'the', 'a', 'an') and be made plural. The nouns you encountered above are all regular and feature examples of all three classes:

kvinne – feminine, *bil* – masculine, *eple* – neuter

Here are three more regular Norwegian nouns and their translations:

jente (feminine) = 'girl', *hund* (masculine) = 'dog', *hotell* (neuter) = 'hotel'

Translate into Norwegian:

11. 'The girl stops here.'
12. 'A girl has a hotel.'
13. 'I have the dogs.'
14. 'The girl has dogs.'

Problem 5.11c Here are some more Norwegian words without any information about the classes the slightly irregular nouns belong to:

sko = 'shoe', *mann* = 'man', *ikke* = 'not'

Translate into English:

15. *Mennene har epler.*
16. *Kvinna har ikke skoene.*
17. *Jeg har ikke eplene.*

Problem 5.12

Afrihili (Michael Salter & Aleka Blackwell, UKLO 2019)

Here are some words in Afrihili and their English translations:

adu	'tooth'	*emeli*	'ship'	*olengi*	'horse'
afidi	'machine'	*enti*	'date tree'	*oluganda*	'dialect'
ajamuri	'republic'	*eshuli*	'principal'	*omola*	'child'
akalini	'pen'	*eture*	'flowers'	*omukazi*	'girl'
amadu	'dentist'	*ijamura*	'president'	*omuntundu*	'dwarf'
amkate	'bread'	*ikalini*	'pens'	*omuntu*	'man'
amola	'children'	*ilengi*	'horses'	*uruzindi*	'stream'
amukamo	'kingdom'	*imukazi*	'girls'	*uruzi*	'river'
aturesine	'bouquet'	*isabamatu*	'cobbler'		
emelisini	'fleet'	*ishule*	'school'		

Problem 5.12a Translate into English: *ajamura*, *amkamate*, *oluga*.

Problem 5.12b Translate into Afrihili: 'machinist', 'ships', 'flower', 'group of girls', 'date fruit', 'shoe', 'king'.

Problem 5.12c Below are three more Afrihili words and three options for a likely translation of the word:

imulenzi	a. 'fruit'	b. 'boys'	c. 'bridge'
aposino	a. 'baggage'	b. 'classroom'	c. 'parent'
iwelemase	a. 'book'	b. 'library'	c. 'librarian'

Pick the translation most likely to be correct and explain your choice.

Problem 5.13

Maltese (Simona Strizhevskaya, LLO 2020)

Here are some phrases in Maltese and their English translations:

serp aħdar	'green snake'	*mogħża sewda*	'__(1)__ __(2)__'
mogħżiet bojod	'white goats'	*ktieb aħmar*	'__(3)__ __(4)__'
kowt blu	'blue coat'	*mejda bajda*	'__(5)__ __(6)__'
mejda kannella	'brown chair'	*baqra* __(7)__	'blue cow'

żiemel iswed	‘black horse’	*fjuri* __(8)__	‘red flowers’
fjura vjola	‘purple flower’	*kelb* __(9)__	‘brown dog’
karozza ħamra	‘red car’	*kotba* __(10)__	‘yellow books’
rħula ħodor	‘green settlments’	*siġra* __(11)__	‘green tree’
kowtijiet roża	‘pink coats’	*mwejjed* __(12)__	‘purple chairs’
qomos blu	‘blue shirts’	*tuffieħa* __(13)__	‘yellow apple’
fenek isfar	‘yellow rabbit’	__(14)__ __(15)__	‘red snake’
qomos sowod	‘black shirts’		

Problem 5.13a Fill in the blanks. Each blank corresponds to *a single* word.

Problem 5.13b Based on the data given, one cannot translate ‘white book’. Why not?

Problem 5.14

Latvian (Maria Rubinstein, MSK 1999)

Here are some phrases in Latvian and their English translations:

1.	*augsts ozols*	‘tall oak’
2.	*vecs grāmatu veikals*	‘old bookstore’
3.	*veca meža*	‘of the old wood’
4.	*stikla galds*	‘glass table’
5.	*bruņinieka cimds*	‘knight’s glove’
6.	*sudraba ābols*	‘silver apple’
7.	*autora teksts*	‘author’s text’
8.	*operāciju galds*	‘surgery table’
9.	*pretīgu piena ēdienu*	‘of disgusting dairy products’
10.	*labs institūts*	‘good institute’
11.	*laba bērnu ārsta*	‘of the good paediatrician’
12.	*grāmatu veikala*	‘__(1)__’
13.	__(2)__*kolektīvs*	‘authors’ group’
14.	__(3)__*turnīrs*	‘knight tournament’
15.	*pretīgs* __(4)__	‘__(5)__child’
16.	*balts* __(6)__	‘white silver’
17.	__(7)__*zara*	‘of the oak branch’

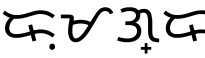
18. __(8)__ ‘of the oak wood’
19. __(9)__*ārstu* ‘of the institute’s doctors’

Problem 5.14a Fill in the blanks. Some blanks may correspond to multiple words. If you think some blanks can be filled in different ways, write all possibilities and explain the difference between them.

Problem 5.15

Ilocano (Patrick Littell, NACLO 2008)

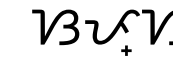
Below are 12 Ilocano words written in the Baybayin script, as well as their English translations given *in random order*:

1.
2. 
3.
4. 
5.
6. 
7.
8.
9.
10.
11.
12. 

‘to look’, ‘is skipping with joy’, ‘is becoming a skeleton’, ‘a skeleton’, ‘to buy’, ‘various skeletons’, ‘various appearances’, ‘to reach the top’, ‘is looking’, ‘appearance’, ‘summit’, ‘happiness’, ‘skeleton’

Problem 5.15a Determine the correct correspondences.

Problem 5.15b Fill in the blanks.

 ‘__(1)__’
 ‘__(2)__’
 ‘__(3)__’
__(4)__ ‘(a) purchase’
__(5)__ ‘is buying’

Problem 5.16

Irish (Tom Payne, UKLO 2011)

Below are some number phrases in Irish and their English equivalents:

1.	*garra amháin*	‘1 garden’
2.	*gasúr déag*	‘11 boys’
3.	*ocht mballa is dhá fichid*	‘48 walls’
4.	*dhá gharra déag is ceithre fichid*	‘92 gardens’
5.	*trí bhád*	‘3 boats’
6.	*seacht ndoras déag*	‘17 doors’
7.	*seacht mbád déag is dhá fichid*	‘57 boats’
8.	*naoi nduine déag is fiche*	‘39 people’
9.	*ceithre fichid doras*	‘80 doors’
10.	*cúig bhalla*	‘5 walls’
11.	*sé ghasúr is trí fichid*	‘66 boys’
12.	*deich mbád*	‘10 boats’
13.	*sé dhuine*	‘6 people’
14.	*trí dhoras is dhá fichid*	‘43 doors’
15.	*garra is ceithre fichid*	‘81 gardens’

Problem 5.16a Translate into English:

16. *naoi mbád déag is ceithre fichid*
17. *sé dhuine déag*
18. *naoi nduine*
19. *fiche gasúr*
20. *garra déag is fiche*

Problem 5.16b Translate into Irish:

21. ‘2 boys’
22. ‘38 walls’
23. ‘14 walls’
24. ‘71 doors’
25. ‘21 boats’
26. ‘90 people’

Problem 5.17

Iaai (Rujul Gandhi, APLO 2020)

Here are some phrases spoken by two speakers of Iaai and their English translations, *in random order*:

Speaker 1 (male, 50 years old):

1.	*hoom hu*	A.	‘your fire’
2.	*belem mââng*	B.	‘my water’
3.	*waau haalee Aiawa*	C.	‘my bus’
4.	*uutap taben than*	D.	‘your mango’
5.	*tabik kar*	E.	‘your boat’
6.	*anyik sawakiny*	F.	‘the chief’s chair’
7.	*anyim meic*	G.	‘my necklace’
8.	*belik köiö*	H.	‘Aiawa’s cat’

Speaker 2 (male, 12 years old):

9.	*belen koka*	I.	‘your goat’
10.	*anyik tang*	J.	‘his yam’
11.	*anyim karopëë*	K.	‘Kua’s mother’
12.	*haaleem nani*	L.	‘his watermelon’
13.	*an koko*	M.	‘his car’
14.	*hinyö anyi Kua*	N.	‘my basket’
15.	*belik nu*	O.	‘his Coke’
16.	*anyin loto*	P.	‘my coconut’
17.	*an waajem*	Q.	‘your dugout’

Problem 5.17a For each speaker, determine the correct correspondences.

Problem 5.17b Translate into English:

18. *belem waajem*
19. *karopëë hoon hinyö*

Mention any meaning that is not reflected in the literal translation.

Problem 5.17c Given below are some English words and their Iaai translations:

‘dog’ = *kuli* ‘tea’ = *trii* ‘canoe’ = *ok*

Translate into Iaai in all possible ways:

20. ‘the cat’s tea’
21. ‘Kua’s coconut’
22. ‘his canoe’
23. ‘my dog’

â, *ë*, *ö* are vowels. A ‘dugout’ is a long, narrow canoe made of a tree trunk. A ‘yam’ is a starchy vegetable, similar to a sweet potato. ‘Coke’ is a carbonated beverage sold by The Coca-Cola Company, an American multinational company. ‘Aiawa’ and ‘Kua’ are names of people.

5.10 Solutions of practice problems

Solution for practice problem 5.9. Ulwa

Solution 5.9a

suulu = ‘dog’
suukilu = ‘my dog’
suumanalu = ‘your$_{\text{PL}}$ dog’
onyan = ‘bow’
miskatu = ‘his cat’
onkinayan = ‘our bow’
mismatu = ‘your$_{\text{SG}}$ cat’
onkayan = ‘his bow’

Solution 5.9b

suumalu = ‘your$_{\text{SG}}$ dog’
miskanatu = ‘their cat’

Solution 5.9c

‘cat’ = *mistu*
‘my cat’ = *miskitu*
‘your$_{\text{SG}}$ bow’ = *onmayan*
‘their bow’ = *onkanayan*

Rules:

- The possessive is marked by an infix placed before the last syllable. The structure of the infix is Person–Number.
- Person: *-ki-* = 1 *-ma-* = 2 *-ka-* = 3
- Number: ∅ = SG *-na-* = PL

Solution for practice problem 5.10. Palauan

Solution 5.10a

1. ‘6 teachers’
2. ‘19 months’
3. ‘December’
4. ‘Friday’

Solution 5.10b

5. *eai ęl sils*
6. *tęruich me a tetiu ęl chad*
7. *teuid ęl sensei*
8. *ongelolem ęl buil*
9. *ongeai ęl buil*

Solution 5.10c

10. *ede*
11. *tede*
12. *klde*

Rules:

- Structure: Numeral + *ęl* + Noun
- Numerals have the following prefixes:
  - *e-* = time periods (days, months, years)
  - *te-* = people
  - *kl-* = non-human nouns (which do not refer to persons)
  - *onge-* = ordinal numerals
- For numbers higher than 10, the prefix is attached to the units.
- $10 + X$ = *tęruich me a X*

Solution for practice problem 5.11. Norwegian

Solution 5.11a

1. C.
2. J.
3. E.
4. H.
5. B.
6. A.
7. D.
8. G.
9. F.
10. I.

Solution 5.11b

11. *Jenta stanser her.*
12. *En jente har et hotell.*
13. *Jeg har hundene.*
14. *Jenta har hunder.*

Solution 5.11c

15. ‘The men have apples.’
16. ‘The woman does not have the shoes.’
17. ‘I do not have the apples.’

Rules: Sentence structure: SOV (Subject-Object-Verb)

	Last letter	SG indef. 'a dog'	SG def. 'the dog'	PL def. 'the dogs'	PL indef. 'dogs'
Neuter	*e*	*et X-e*	*X-et*	*X-ene*	*X-er*
	≠ *e*	*et X*	*X-et*	*X-ene*	*X-er*
Feminine	*e*	*en X-e*	*X-a*	*X-ene*	
	≠ *e*	*en X*	*X-a*	*X-ene*	
Masculine	*e*	*en X-e*	*X-en*	*X-ene*	*X-er*
	≠ *e*	*en X*	*X-en*	*X-ene*	*X-er*

Neuter	Feminine	Masculine
'apple'	'woman'	'bus'
'hotel'	'girl'	'car'
		'dog'

Note

The nouns for 'man' and 'shoe' are not included in the table, since their gender cannot be determined.

Solution for practice problem 5.12. Afrihili

Solution 5.12a *ajamura* = 'presidents' *oluga* = 'language'
amkamate = 'baker'

Solution 5.12b

'machinist' = *afimadi*		'date fruit' = *entindi*
'ships' = *imeli*		'shoe' = *isabatu*
'flower' = *ature*		
'group of girls' = *omukazisini*		'king' = *omukama*

Solution 5.12c

imulenzi	= b.	'boys' (first and last vowels are identical ⟹ plural)
aposino	= a.	'baggage' (affix *-sin-* ⟹ collective noun)
iwelemase	= c.	'librarian' (affix *-ma-* ⟹ profession)

Rules:

- All nouns begin and end with a vowel (V_1RV_2)
- Singular: $V_1 \neq V_2$, plural: $V_1 \rightarrow V_2$ (V_1 becomes V_2 $\Longrightarrow V_2RV_2$)
- Derived nouns (from a basic noun V_1RV_2):
  - head of an organisation: V_2RV_1 (first and last vowels switch places);
  - profession: infix *-ma-* is inserted before the last syllable of the stem ($V_1C...CV_2 \Longrightarrow V_1C...maCV_2$);
  - collective noun: $V_1RV_2sinV_2$ (add suffix *-sinV*, where V is the last vowel of the stem);
  - diminutive: $V_1RV_2ndV_2$

Solution for practice problem 5.13. Maltese

Solution 5.13a

(1) 'goat'	(6) 'white'	(11) *ħadra*
(2) 'black'	(7) *blu*	(12) *vjola*
(3) 'book'	(8) *ħomor*	(13) *safra*
(4) 'red'	(9) *kannella*	(14) *serp*
(5) 'chair'	(10) *sofor*	(15) *aħmar*

Solution 5.13b We do not know the first vowel of the stem for 'white' (V_1).

Rules: There are two types of colours: invariable ('pink', 'purple', 'brown', 'blue') – which have the same form in all contexts – and variable ('green', 'white', 'black', 'red', 'yellow').

Nouns are divided into three categories:

Cat. I. Singular, end in a consonant;
Cat. II. Singular, end in *a*;
Cat. III. Plural.

Based on these categories, the adjective is declined as follows:

Cat. I. $V_1C_1C_2V_2C_3$
Cat. II. $C_1V_2C_2C_3a$
Cat. III. $C_1oC_2oC_3$

where C and V refer to a consonant and a vowel, respectively.

We notice here that V_1 appears only for Cat. I. Therefore, knowing the forms for Cat. II and III is not sufficient to deduce the form for Cat. I; similarly, only knowing the form for Cat. III is not enough to deduce the forms for Cat. I and II.

Note

We refer here to problem 5.8 and the discussion thereafter. We notice that the five colours that can be inflected ('green', 'white', 'black', 'red', 'yellow') are the first to enter the language, according to the hypothesis of Berlin & Kay, while the invariable colours are those from a later stage ('blue', 'pink', 'purple', 'brown') – which, in this case, are loan words.

Solution for practice problem 5.14. Latvian

Solution 5.14a

(1) 'of the bookshop'
(2) *autoru*
(3) *bruņinieku*
(4) *bērns*
(5) 'disgusting'
(6) *sudrabs*
(7) *ozola*
(8) *ozolu*

(9) *institūta* or *institūtu* (the former is used if we refer to the doctors from a single institute, while the latter is used to refer to the doctors from different institutes)

- Suffixes of the noun head:

  -s = nominative *-a* = genitive SG *-u* = genitive PL

- Determiners can be split into three groups:
  - Those which agree with the noun head (receive the same suffix). In this category, we include all the qualifying adjectives ('tall', 'old', 'good', 'disgusting').
  - Those which receive the ending *-a* – they are represented by Latvian nouns in genitive singular ('dairy products' = products *of milk*).
  - Those receiving the ending *-u* – they are Latvian nouns in genitive plural ('paediatrician' = doctor *of children*, 'bookshop' = shop *of books*).

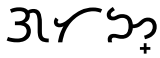
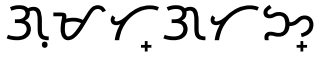
The difference between the last two groups is purely semantic, depending on whether they refer to a singular or a plural noun (a 'surgery table' is a table for *surgeries* since multiple surgeries are performed on the same table; a 'paediatrician' is a doctor of *children* because they treat more children, not a single one, etc.)

Solution for practice problem 5.15. Ilocano

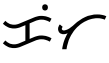
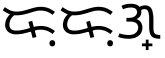
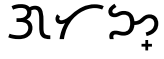
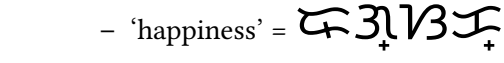
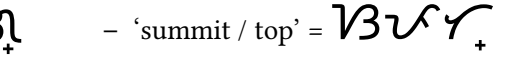
Solution 5.15a

1. 'appearance'
2. 'various appearances'
3. 'to look'
4. 'is looking'
5. 'happiness'
6. 'is skipping for joy'
7. 'skeleton'
8. 'various skeletons'
9. 'is becoming a skeleton'
10. 'to buy'
11. 'summit'
12. 'to reach the top'

Solution 5.15b

(1) 'to become a skeleton'
(2) 'various summits'
(3) 'is reaching the top'
(4) 
(5) 

Rules:

- Stems:
  - 'appearance' = 
  - 'skeleton' = 
  - 'purchase' = 
  - 'happiness' = 
  - 'summit / top' = 
- Derivation processes:
  a. *Partial reduplication*: copy the first two symbols and add them to the beginning of the word. The first symbol retains its diacritic, while the diacritic of the second symbol is replaced by a plus ().
  b. *Epenthesis*: insert  after the first symbol. The diacritic of the first symbol moves to this one and the first symbol receives an underdot ().

With these two processes, we can obtain three different transformations starting from the singular noun:

- Plural noun ('various...') (process a)
- Infinitive verb (process b)
- 3SG present cont. verb (process a, followed by b)

Solution for practice problem 5.16. Irish

Solution 5.16a

16. '99 boats'
17. '16 people'
18. '9 people'
19. '20 boys'
20. '31 gardens'

Solution 5.16b

21. *dhá ghasúr*
22. *ocht mballa déag is fiche*
23. *ceithre bhalla déag*
24. *doras déag is trí fichid*
25. *bád is fiche*
26. *deich nduine is ceithre fichid*

Rules: Irish uses base 20; numbers are written as:

$$(U) + (10) + (20X) \Longleftrightarrow (U)\ (déag)\ (is\ X\ fichid)$$

- U is between 2 and 9
- number 10 has two forms: *déag* – used if there is a unit number (from 2 to 9) – and *deich* – used only for 10 and multiples of 10 (30, 50, etc.)
- if $X = 1$, *is X fichid* becomes *is fiche*

Phrase structure: we can consider each structure to have four parts: I + II + III + IV.

- Part I is for the units (or *deich*).
- Part II is always the noun.
- Part III can only be filled by two words: *amháin* (meaning *1*, only for a singular noun) or *déag* (but not *deich*).
- Part IV is always a multiple of 20 (structures *is X fichid*/*is fiche*).

The only exception takes place when the number of objects is a multiple of 20 (20, 40, etc.). In this case, Part I is occupied by *X fichid*/*fiche* and the noun is placed at the end (in this case the particle *is* is not used). Therefore, the examples in the problem (and task (a)) can be analysed as follows:

	I	II	III	IV	Translation
1.		*garra*	*amháin*		'1 garden'
2.		*gasúr*	*déag*		'11 boys'
3.	*ocht*	*mballa*		*is dhá fichid*	'48 walls'
4.	*dhá*	*gharra*	*déag*	*is ceithre fichid*	'92 gardens'
5.	*trí*	*bhád*			'3 boats'
6.	*seacht*	*ndoras*	*déag*		'17 doors'
7.	*seacht*	*mbád*	*déag*	*is dhá fichid*	'57 boats'
8.	*naoi*	*nduine*	*déag*	*is fiche*	'39 people'
10.	*cúig*	*bhalla*			'5 walls'
11.	*sé*	*ghasúr*		*is trí fichid*	'66 boys'
12.	*deich*	*mbád*			'10 boats'

	I	II	III	IV	Translation
13.	*sé*	*dhuine*			‘6 people’
14.	*trí*	*dhoras*		*is dhá fichid*	‘43 doors’
15.		*garra*		*is ceithre fichid*	‘81 gardens’
16.	*naoi*	*mbád*	*déag*	*is ceithre fichid*	‘99 boats’
17.	*sé*	*dhuine*	*déag*		‘16 people’
18.	*naoi*	*nduine*			‘9 people’
20.		*garra*	*déag*	*is fiche*	‘31 gardens’
9.	*ceithre fichid*	*doras*			‘80 doors’
19.	*fiche*	*gasúr*			‘20 boys’

We separated examples 9 and 19 to highlight the case in which the number of objects is a multiple of 20.

Moreover, we can notice from the table that the difference between, for example, 20 and 21 (or 40/41, 60/61 etc.) is only based on the position of the noun.

Lastly, we notice that the noun has a variable form. In this case, it undergoes an initial consonant mutation. We notice three types of initial consonants:

- simple consonants (*b*, *d*, *g*) – used if the number of units is 1 or if the number of objects is a multiple of 20 (in other words, if position I is empty or occupied by a multiple of 20);
- consonants followed by *h* (*bh*, *dh*, *gh*) – used if the number of units is between 2 and 6 (or if position I is occupied by 2–6);
- consonants preceded by a nasal (*mb*, *nd*) – used if the unit number is 7–10 (10 refers only to the form *deich*, which appears in the first position). In the problem, there are no examples and we are not asked to write what happens with the consonant *g* in this context, but we can notice that the nasal added before assimilates to the place of articulation.

Solution for practice problem 5.17. Iaai

Solution 5.17a

1. E.	5. C.	9. O.	13. J.	17. L.
2. D.	6. G.	10. N.	14. K.	
3. H.	7. A.	11. Q.	15. P.	
4. F.	8. B.	12. I.	16. M.	

Solution 5.17b 18. 'your watermelon (for drinking)'

19. 'the mother's dugout'

Solution 5.17c

20. *trii belen waau*	22. *hoon ok* or *anyin ok*	
21. *nu a Kua* or *nu bele Kua*	23. *haaleik kuli*	

Rules: Structure:

X's *Y* =	*C*–Poss *Y*	(*X* = pronoun)
	Y C–Poss *X*	(*X* = common noun)
	Y C X	(*X* = proper noun)

Poss:	*-k*	(*X* = 1SG)	$e \longrightarrow i$ / _ *k*
	-m	(*X* = 2SG)	
	-n	(*X* = 3SG)	

C:	*a*	*Y* = food*
	bele	*Y* = drinks*
	haalee	*Y* = animals
	hoo	*Y* = boats†
	tabe	*Y* = things you can sit on/in†
	anyi	otherwise

* Fruits can be either 'food' or 'drink' depending on how the speaker intends them to be consumed.

† In the case of younger generations (Speaker 2), these types of noun also fall into the *anyi* category (new generations tend to simplify the classifier system and give up on very specific classifiers, preferring to use the general classifier).

Further reading

Aikhenvald, Alexandra Y. & Robert M. W. Dixon (eds.). 2013. *Possession and ownership*. Oxford: Oxford University Press.

Berlin, Brent & Paul Kay. 1969. *Basic color terms: Their universality and evolution*. 1st edn. Berkeley: University of California Press.
Booij, Geert. 2012. *The grammar of words*. Oxford: Oxford University Press.
Cabredo Hofherr, Patricia & Anne Zribi-Hertz (eds.). 2013. *Crosslinguistic studies on noun phrase structure and reference*. Leiden: Brill Academic Publishers.
Rijkhoff, Jan. 2002. *The noun phrase*. Oxford: Oxford University Press.

6 Verb and verb phrase

6.1 Introduction

A number of general concepts are important for a discussion of verbs and verb phrases. Here are some notional definitions (i.e., those based on meaning) for key concepts:

VERB: shows the action, existence, or state.[1]

SUBJECT: typically shows who or what performs the action.

(DIRECT) OBJECT: typically shows the person or object which is acted upon by the subject.

INDIRECT OBJECT: shows the entity upon which the action is reflected indirectly.

We can also think about connections between verbs and noun phrases. For example:

TRANSITIVE and INTRANSITIVE VERB: a transitive verb is one which has a direct object (*I broke the glass*) while an intransitive verb is one which has no direct object (*The baby yawned*). Note that in English we cannot say **I broke* or **The baby yawned his mouth*. Some verbs can vary in whether they take a direct object: thus, the verb *to eat* can be both transitive (*She is eating a pizza* – transitive use of *eat*, with *a pizza* as the direct object) and intransitive (*She is eating*).

6.2 Variables of the verb

By variables, we mean those parameters which change in the problem. For example, comparing the sentences:

He eats. *I ate.*

[1] In linguistics, there is a difference between *verb* and *predicate*, the former referring to the part of speech, while the latter to the part of a sentence. In linguistics problems, however, the term *predicate* is typically not used, and it is usually replaced by *verb*.

we notice that the variables are *subject* (*he* vs. *I*), tense (present vs. past), but not the verb itself (both sentences have the same verb - *eat*).

Generally, these variables can be classified into three categories:

1. TAM (Tense, Aspect, Mood)
2. Arguments (S, S+O, S+O+O)
3. Others

6.3 TAM (Tense, Aspect, Mood)

6.3.1 Tense

Tense is a grammatical category, related to the concept of time. Languages often have three subtypes of tense: past (often used when the action or state denoted by the verb occurs before the moment of speech), present (action happens while speaking) and future (action will take place after the moment of speech).

Although most of the Indo-European languages have all of these tenses, there are languages which only have two distinct tenses. Thus, there are languages which only distinguish *past* and *non-past*, such as Arabic and Japanese (non-past refers to everything that is not past, thus representing present and future) or languages which distinguish *future* and *non-future*, such as Greenlandic or Nivkh (where non-future represents past and present), or even languages that distinguish only *present* and *non-present* (one example is the constructed language Ithkuil). Moreover, there can be languages which make no tense distinctions at all, such as Chinese or Dyirbal.

Although there are only three broad categories of tense, some languages can have many more actual tenses, mainly depending on the specific moment at which the action occurred. For example, the Yagua language has five different past tense markers:[2]

a. *-jásiy* – for actions which took place a couple of hours ago (in the same day);

b. *-jay* – for actions occurring one day ago;

c. *siy* – one week to one month ago;

[2]This phenomenon was featured in a problem by Vlad A. Neacșu (RoLO 2018).

d. *tíy* – one or two months to one or two years ago;

e. *-jada* – more than two years ago (it is also called *distant past* or *legendary past*).

Additionally, some languages can have specific tenses, e.g., specific for actions occurring one day ago, the previous day (*hesternal* tense) or the following day (*crastinal* tense). There can even be *pre-hesternal* and *post-crastinal* tenses, referring to actions occurring two days before/after the moment of speech.

Some languages also have *hodiernal* tenses, which are specific to actions occurring on the same day as the moment of speech (today). These can be past tenses (actions happening earlier today) or future tenses (actions happening later today).

Therefore, returning to the Yagua example, we can define the marker a. (*-jásiy*) as a past hodiernal tense, while b. (*-jay*) can be considered a hesternal tense marker.

6.3.2 Aspect

Aspect shows the evolution in time of an action, state or event, independent of the moment of speech. The most common aspects are the *perfective* and *imperfective*.

Perfective aspect is used when the event denoted by the verb is bounded. The imperfective is used when the event is seen as unfolding, or when the event is repeated/habitual. Depending on the language, there can also be other aspects such as:

- Progressive = action is unfolding (progressing).
- Semelfactive = action is short-term.
- Accidental = action is done by mistake.

There are a lot of different aspects and their meaning can usually be inferred from their name (*punctual*, *resumptive*, *intensive*, *attenuative*, *moderative*, *experiential*, *durative*, *pausative*, *terminative*, *episodic*, *generic*, *habitual*, *discontinuous*, *prospective*, *intentional*, etc.). Memorising the names of all these aspects is not necessary, but it is important that you get used to the different meanings these aspects convey. Moreover, all of these aspects (not all of which exist in English so they may be expressed in more roundabout ways) need to be translated into English in a linguistics problem (most likely through a specific phrasing or by using an adverb). Therefore, instead of referring to the aspect itself, when solving

a linguistics problem, you can simply write "The structures which are translated as...". For example, the durative aspect can be translated into English by phrases such as 'for a while'. Therefore, even if you do not know the name of the aspect (*durative*), it is enough to realise that there is a distinction between, for example, *You ate* and *You ate for a while*, in which case you can explain the sense of the durative aspect marker through the phrase *for a while*. Other examples are: potential aspect *probably*, prospective aspect *I'm getting ready to...*, etc.

Remember

In English, at least at school, aspect is not often talked about and the idea of "tense" refers, in fact, to a combination of tense and aspect. English has two main aspects: perfect aspect (e.g., *I have/had eaten a sandwich*, as defined above) and progressive aspect (e.g., *I am/was eating a sandwich*, sometimes known as continuous aspect).

6.3.3 Mood

Mood is an inflectional category related to the concept of modality, which is often connected with the speaker's attitude towards the message being transmitted (whether it is a fact, a wish, a command, etc.). Moods are classified into two broad classes: *realis* moods which show that something is a statement or a fact; and *irrealis* moods which refer to actions which have not happened (or will certainly not happen).

6.3.3.1 Realis moods

1. Indicative mood: It is the most common mood and it is used for statements of fact. It is considered that all situations in a particular language that cannot be categorised as another mood will be classified as belonging to the indicative mood.

2. Certain languages can have another realis mood, a special mood which is used solely for general truths, as in the examples *Fish swim* or *Chickens have two legs.*

6.3.3.2 Irrealis mood

There are many different subtypes of irrealis. As with the Uralic cases listed in Section 5.3, there is no need for you to remember all these subtypes. However, some subtypes appear more frequently in linguistics problems and we briefly describe these below.

1. Subjunctive mood: marks imaginary/hypothetical events, as well as opinions and emotions. It is the main irrealis mood and it represents an "umbrella term" for all the instances in which one language does not have another mood to express that attitude.
2. Conditional mood: the action is conditioned (by another action).
3. Optative mood: shows desires or wishes.
4. Imperative mood: direct commands/request/interdictions.
5. Jussive mood: similar to the imperative, but it expresses commands towards a third person, not present. Since in English this mood is not expressed directly by the verb, we usually use the subjunctive mood to express a similar meaning, e.g., *I asked that he cook.*

6.3.4 Verbal expression of modality

In certain languages, modal verbs are used to express different kinds of modality. In many languages, modal verbs are considered auxiliary verbs. This means that they are accompanied by a "lexical verb", i.e., a verb with semantic content. In English, the central modal verbs (such as *must* and *should*) take the plain form of the verb as their complement (e.g., *I must leave*, *you should eat*). They can also combine with aspect markers (*he must be working*, *she should have stayed*): notice here that it is the following verb (the aspect marker) that is unchanged; the form of the lexical verb is determined by the aspect marker (*be +V ing*, *have +V ed*).

Other verbs can also be used to express modality, and such verbs often take the *to*-infinitive as their complement. For example, in the sentence *I want to go*, *want* is the verb which shows the modality (the desire), while *go* is the verb with semantic content, showing the action I *want/desire* to perform. Other such patterns in English include *wish to V*, *try to V*, etc.

The verbs in bold in sentences like *I* ***want*** *to swim* and *I* ***tried*** *to swim* are sometimes known as catenative verbs because they can form a sequence or chain of verbs (*catena* is Latin for 'chain'), as in *I* ***want*** *to* ***try*** *to swim.* Catenative verbs

take a non-finite verb (e.g., an infinitive or participle) as their complement in English. Such catenative verbs can express modality (*I **need** to leave*) and aspect (*He **kept** swimming*), and since catenative verbs can combine, a sentence can involve the marking of both modality and aspect (*I **need** to **keep** swimming*).

This concept is relevant because it involves an interdependency between the two verbs, which are grammatically and semantically interconnected. Thus, in linguistics problems, if sequences of verbs occur, we need to pay attention to the following potential parameters: the order of the modal/catenative verb and the semantic verb (whether it comes before or after it, or whether there are other words or morphemes in between), as well as which of the two verbs gets conjugated (in some languages, only the auxiliary verbs are conjugated; in others, only the lexical verb; and in yet others, both the auxiliary and the lexical verb are conjugated).

6.3.5 Evidentiality

Evidentiality, unlike tense, aspect, and mood, shows the way in which the uttered information was discovered or, in other words, what evidence there is for the transmitted information. For example, in Pomo, there are four types of evidentiality, each of them having its own marker:

1. Visual: the speaker witnessed the action;
2. Sensorial non-visual: the speaker felt (by hearing, smelling, etc.) something that pointed towards the action. For example, the sentence *The kids fought* can receive a sensorial non-visual evidentiality marker to point out that the speaker has not seen the children fighting, but heard them;
3. Inferential: the speaker did not witness the action but was able to see its consequence or result. For example, the sentence *The man cooked the fish* can carry an inferential evidentiality marker to show that the speaker has not seen the fish being cooked (the process of cooking), but, for example, saw someone holding a plate with the cooked fish (thus, being able to infer that, at some point, the fish went through the cooking process).
4. Reportative: the speaker found out the information from someone else.

For example, the English sentence *It rained* could be translated into Pomo in four different ways, depending on the source of the evidence. Visual – the speaker sees that it's raining; sensorial non-visual – they hear the rain; inferential – they notice it is wet outside; reportative – someone tells them it rained.

Other evidential contrasts can be:

5. Witness vs. non-witness: the speaker witnessed the action (the information was retrieved through direct observation) or not. This type of evidentiality occurs in Turkish, where there are two types of past, called *seen past* (*görülen geçmiş zaman*, witness evidentiality marker) and *heard past* (*duyulan geçmiş zaman*, non-witness evidential).

6. First-hand vs. second-hand vs. third-hand: first-hand information is equivalent to directly observed information (witness); second-hand information is used to show that the speaker found out the information from someone else (who witnessed the action), while third-hand information indicates that the speaker found out the information from another person (who, in turn, found it out from a third person who witnessed the action).

As in the case of TAM categories, there can be other evidential markers and, depending on the language, each of these categories can be further divided into subcategories. For example, the inferential evidential can have the following subtypes: information deduced based on direct evidence (seeing the result of the action), information deduced based on general knowledge, information deduced (or inferred) based on the speaker's past experiences in similar situations, etc.

Moreover, evidentiality can also be combined with tense, aspect, and mood, for which reason some linguists prefer using the abbreviation TAME (instead of TAM): tense, aspect, mood, evidentiality.

6.4 Arguments

In many linguistics problems, we need to identify the arguments of the verb. These are often the subject (S) and object(s) (O) of the verb. In some cases, we also distinguish between the subject of an intransitive verb and that of a transitive verb. This will be discussed more thoroughly in Section 7.4.

In verb (phrase) problems, S and O are often expressed as pronouns in the English translations, though they may be directly attached to the verb as affixes in the target language. Depending on the difficulty of the problem, we can have different situations:

- Easy problems: subject and object are marked through distinct, independent affixes.

- Medium problems: in which S and O are either combined into a single affix, or they undergo certain phonological changes.

- Hard problems: in which S and O are not expressed uniquely, but rather through a combination of affixes.

For more information about case alignment, see Section 7.4.

Problem 6.1

Swahili (Ronnie Sim, Princeton)

Here are some verbal forms in Swahili and their English translations:

1.	*Ninasema.*	'I speak.'
2.	*Wunasema.*	'You speak.'
3.	*Anasema.*	'She speaks.'
4.	*Wanasema.*	'They speak.'
5.	*Ninaona.*	'I see.'
6.	*Niliona.*	'I saw.'
7.	*Ninawaona.*	'I see them.'
8.	*Niliwuona.*	'I saw you.'
9.	*Ananiona.*	'She sees me.'
10.	*Wutakaniona.*	'You will see me.'
11.	__(1)__	'She saw them.'
12.	__(2)__	'I will see you.'
13.	__(3)__	'She saw me.'

Problem 6.1a Fill in the blanks.

Solution

We notice that all English examples containing the verb 'to see' end in *ona* in Swahili. Similarly, all English examples which contain the subject 'I' start with *ni* in Swahili. Separating these two morphemes, we obtain:

1.	*Ni-nasema.*	'I speak.'
2.	*Wunasema.*	'You speak.'
3.	*Anasema.*	'She speaks.'
4.	*Wanasema.*	'They speak.'
5.	*Ni-na-ona.*	'I see.'
6.	*Ni-li-ona.*	'I saw.'

7.	*Ni-nawa-ona.*	'I see them.'
8.	*Ni-liwu-ona.*	'I saw you.'
9.	*Anani-ona.*	'She sees me.'
10.	*Wutakani-ona.*	'You will see me.'

In examples 5 and 6 we are left with only one unidentified morpheme (*na* and *li*, respectively). The two sentences differ only in terms of their tense (past vs. present); therefore, we infer that *na* is the present-tense marker, while *li* is the past-tense marker. Separating these two morphemes, we obtain:

1.	*Ni-na-sema.*	'I speak.'
2.	*Wu-na-sema.*	'You speak.'
3.	*A-na-sema.*	'She speaks.'
4.	*Wa-na-sema.*	'They speak.'
5.	*Ni-na-ona.*	'I see.'
6.	*Ni-li-ona.*	'I saw.'
7.	*Ni-na-wa-ona.*	'I see them.'
8.	*Ni-li-wu-ona.*	'I saw you.'
9.	*A-na-ni-ona.*	'She sees me.'
10.	*Wutakani-ona.*	'You will see me.'

Now it is easy to identify that the morpheme order is Subject – Tense – Object – Stem (we can abbreviate it as S-Tense-O-V). Moreover, we notice that the subject and object markers are identical. Therefore, we can segment example 10 as well (knowing that 'you' is *-wu-*, and 'I/me' is *-ni-*) and we get *Wu-taka-ni-ona*. Therefore, the future marker is *taka*.

Once all the rules are discovered, it is time to structure them neatly. Usually, in this type of problem, we start by writing down the morpheme order, followed by explaining each of the morphemes.

Rules: Option 1

- Structure: S–Tense–O–V
- S: 1SG = *ni*, 2SG = *wu*, 3SG = *a*, 3PL = *wa*

- Tense: present = *na*, past = *li*, future = *taka*
- O: 1SG = *ni*, 2SG = *wu*, 3SG = *a*, 3PL = *wa*
- Verb: 'speak' = *sema*, 'see' = *ona*

Another option is combining the order of the morphemes with their structure, in a single table.

Rules: Option 2

S	Tense	O	V
ni = 1SG	*li* = past	*ni* = 1SG	*sema* = speak
wu = 2SG	*na* = present	*wu* = 2SG	*ona* = see
a = 3SG	*taka* = future	*a* = 3SG	
wa = 3PL		*wa* = 3PL	

The main disadvantage of these two options (in the case of this problem) is that we have to write the pronoun markers twice (once for S and once for O), although they are identical.

Usually, when we work with pronoun markers, it is favourable to structure them in a table in which we write the person in columns and the number in rows (or vice versa). Therefore, the subject markers become:

	1	2	3
SG	*ni*	*wu*	*a*
PL			*wa*

In this way, we can show the fact that the subject is identical to the object.

Rules: Option 3

- Structure: S–Tense–O–V
- S = O

	1	2	3
SG	*ni*	*wu*	*a*
PL			*wa*

- Tense: present = *na*, past = *li*, future = *taka*
- Verb: 'speak' = *sema*, 'see' = *ona*

All these options for writing out the rules are correct and complete, and they would all receive the maximum score. But, depending on the problem, one of them fits better in the sense that it is more succinct and helps save some time.

Once the rules are written, we can start solving the tasks.

11.	3SG–past–3PL–'see'	'She saw them.'
12.	1SG–future–2SG–'see'	'I will see you.'
13.	3SG–past–1SG–'see'	'She saw me.'

Thus, the answers are: 11. *aliwaona* 12. *nitakawuona* 13. *aliniona*

6.5 Segmenting

Segmenting is probably the most important part for this type of problem. It refers to dividing the verb into all its component morphemes, as we did in the previous problem (we divided the word *wutakaniona* into *wu-taka-ni-ona*, so we segmented it into its components).

For simple problems, once the verbs are fully (and correctly) segmented, it is just a matter of making the correspondences and figuring out the meaning of each morpheme. Sometimes segmentation can be complicated due to morphemes that undergo phonological changes.

In the case of chaos-and-order problems (those in which the data are given in random order), the general approach includes: 1) segmenting the verb, 2) deducing the verb structure (in the given language), 3) showing a frequency table.

Problem 6.2

Dabida (Ksenia Gilyarova, TurLom 2006)

Here are some verbal forms in Dabida and their English translations *in random order*:

> *dichakaδana*, *βichanirasha*, *kuchanikunda*, *dichakurasha*, *dicharashana*, *βichamukunda*, *muchadikaδa*, *βichakaδana*
>
> 'we will argue', 'you$_{PL}$ will beat us', 'we will curse you$_{SG}$', 'they will fight', 'they will curse me', 'they will fall in love with you$_{PL}$', 'we will fight', 'you$_{SG}$ will fall in love with me'

Problem 6.2a Determine the correct correspondences.

Problem 6.2b Translate into English:

1. *nichakukaδa*
2. *βichakundana*

Problem 6.2c Translate into Dabida:

3. 'you$_{PL}$ will argue'
4. 'you$_{SG}$ will curse them'

Solution

Step 1. Segmenting: when segmenting the verbs, we are not yet too concerned with the English translations. For this problem, we can start with the special characters since they are the easiest to follow. If we start with the letter *β*, we notice that it appears in three examples and in each of these examples we can separate the morpheme *βicha*. Nevertheless, we need to notice that the morpheme *-cha-* appears in every single given example. Therefore, it most likely represents a separate morpheme. Separating the morphemes *-cha-* and *-βi-* we get:

di-cha-kaδana, *βi-cha-nirasha*, *ku-cha-nikunda*, *di-cha-kurasha*, *di-cha-rashana*, *βi-cha-mukunda*, *mu-cha-dikaδa*, *βi-cha-kaδana*

Next, we notice the repeated strings *-rasha-*, *-kunda-*, and *-kaδa-*, obtaining:

di-cha-kaδa-na, *βi-cha-ni-rasha*, *ku-cha-ni-kunda*, *di-cha-ku-rasha*, *di-cha-rasha-na*, *βi-cha-mu-kunda*, *mu-cha-di-kaδa*, *βi-cha-kaδa-na*

Step 2. Now we can deduce the Dabida verb structure:

di *βi* *ku* *mu*	*cha*	*di* *ni* *ku* *mu* ∅	*kaδa* *rasha* *kunda*	*na* ∅

Remember

In case of morphemes 3 and 5, it is important to also mark ∅ (the null morpheme), meaning that the morpheme is optional and it does not appear in all examples.

Step 3. Checking the English examples, we notice only three variables: subject ('you$_{SG}$', 'we', 'you$_{PL}$', 'they'), object (∅, 'me', 'you$_{SG}$', 'us', 'you$_{PL}$') and verb ('to fight', 'to argue', 'to beat', 'to curse', 'to fall in love'). The tense is not a variable since all examples are in the future tense. We can probably assume that the morpheme *-cha-*, which appears in every single example, is the mark of the future.

Remember

For a correct and complete set of rules, we do not need to write the meaning of that morpheme, since it appears in all examples. Actually, we have no proof that it indicates future tense; we have no evidence as to what its function is as there are no contrasting examples without it.

Moreover, we can make some preliminary observations in order to help deduce what is the purpose of each morpheme. We notice that morphemes 1 and 3 are extremely similar (both of them can be *-di-*, *-ku-*, *-mu-*). Therefore, we can guess that these mark the subject and the object (both have the same markers, similar to the previous problem). Moreover, the long morpheme is, usually, the stem (morpheme 4). Nevertheless, we notice that in Dabida we have only three stems, while in English we have five.

We make a frequency table for morphemes 1 and 3 (which we presumed correspond to the pronouns):

	Dabida	
Morpheme	1st	3rd
di	3	1
βi	3	
ku	1	1
mu	1	1
ni		2
∅		3

	English	
Pronoun	S	O
1SG		2
2SG	1	1
1PL	3	1
2PL	1	1
3PL	3	
∅		3

In other words, this table shows, for example, that the morpheme *di* in Dabida appears in three examples in the first position and in a single example in the third position. In English, we have only one pronoun which matches this 3-and-1 pattern, namely the second person singular (2SG, 'you$_{SG}$').

We can easily notice that the first morpheme in Dabida corresponds to the subject in English while the third morpheme corresponds to the object. Moreover, we can infer that *-di-* = 1PL, *-ni-* = 1SG, and *-βi-* = 3PL. From the table, we cannot match the 2SG and 2PL since they both appear only once as a subject and once as an object. We can sum up what we have gathered so far as follows:

Structure: S-*cha*-O-?-?

S = O: *-di-* = 1PL, *-ni-* = 1SG, and *-βi-* = 3PL, {*-ku-*, *-mu-*} = {2SG, 2PL}.

In order to deduce the markers for 2SG and 2PL, we can make a bidimensional frequency table in which we mark the combinations of subject and object:

O / S	2SG	*-di-*	2PL	*-βi-*
∅		XX		X
-ni-	X			X
2sg		X		
-di-			X	X
2pl				

O / S	2sg	1pl	2pl	3pl
∅		XX		X
1sg	X			X
2sg		X		
1pl			X	X
2pl				

For instance, this table shows that we have only one example in which 1PL is subject and 2SG is object, and two examples in which 1PL is subject and there is no object. Since we already know the morphemes for 1SG, 1PL and 3PL, we notice

that 2SG is the only person which appears as an object in an example in which 1PL is subject. Therefore we deduce that *-ku-* = 2SG and *-mu-* = 2PL. Based on these results, we can make almost all the correspondences (save for the two examples which both have 1PL as subject and have no object).

βichanirasha	=	'They will curse me.'
kuchanikunda	=	'You$_{SG}$ will fall in love with me.'
dichakurasha	=	'We will curse you$_{SG}$.'
βichamukunda	=	'They will fall in love with you$_{PL}$.'
muchadikaδa	=	'You$_{PL}$ will beat us.'
βichakaδana	=	'They will fight.'

The two remaining examples are {*dicharashana*, *dichakaδana*} = {'We will fight.', 'We will argue.'}

It is time to analyse the fourth morpheme (which we previously assumed represents the stem):

rasha	*kunda*	*kaδa*
'They will curse me.'	'You$_{SG}$ will fall in love with me.'	'You$_{PL}$ will beat us.'
'We will curse you$_{SG}$.'	'They will fall in love with you$_{PL}$.'	'They will fight.'

We notice that *-kunda-* = 'to fall in love', and *-rasha-* = 'to curse'. On the other hand, *-kaδa-* seems to mean both 'to fight' and 'to beat', which, we notice, have similar meanings in English. Since one of the remaining examples contains the verb 'to fight', it will certainly correspond to the phrase containing *-kaδa-*.

Therefore, we can make the correspondences:

βichanirasha	=	'They will curse me.'
kuchanikunda	=	'You$_{SG}$ will fall in love with me.'
dichakurasha	=	'We will curse you$_{SG}$.'
βichamukunda	=	'They will fall in love with you$_{PL}$.'
muchadikaδa	=	'You$_{PL}$ will beat us.'
βichakaδana	=	'They will fight.'
dicharashana	=	'We will argue.'
dichakaδana	=	'We will fight.'

Based on these, we notice that the stem *-rasha-* can have two translations: 'to argue' and 'to curse'. In order to understand when to use one and when to use the other, we carefully analyse the examples containing *-rasha-*. We notice that sentences which do not have a direct object are translated with 'to argue', while those which have an object use 'to curse'. Similarly, *-kaδa-* means 'to fight' if there is no direct object, or 'to beat' if there is one. Therefore, the meaning of the stem depends on the transitivity of the verb. We can write:

rasha = 'to argue' (intransitive), 'to curse' (transitive)

kunda = 'to fall in love'

kaδa = 'to fight' (intransitive), 'to beat' (transitive)

We come back to the only morpheme left, the fifth one. In order to figure out its function, we make a table in which we separate the structures which contain it from those which do not:

-na	
βichakaδana	= 'They will fight.'
dicharashana	= 'We will argue.'
dichakaδana	= 'We will fight.'
∅	
βichanirasha	= 'They will curse me.'
kuchanikunda	= 'You$_{SG}$ will fall in love with me.'
dichakurasha	= 'We will curse you$_{SG}$.'
βichamukunda	= 'They will fall in love with you$_{PL}$.'
muchadikaδa	= 'You$_{PL}$ will beat us.'

Taking into account the previous observation (that the transitivity of the verb is relevant in this language), we easily notice that *-na* occurs only if the verb is intransitive. Therefore, we can call the marker *-na* an intransitivity marker.

Note

An alternative is to combine the intransitivity marker with the verb stem (hence saying that *rasha* = 'to beat' and *rashana* = 'to fight'). Although this is true and would not impede the correct solution of the tasks, the rules would probably not be awarded full marks, since we failed to identify the specific marker *-na*, which serves a precise and general purpose, independent of the stem.

Since we have discovered all the morphemes and other phenomena, we can sum up our findings.

Rules:

- Structure: S-*cha*-O-V-(*na*)
- S = O:

	1	2	3
SG	*ni*	*ku*	
PL	*di*	*mu*	*βi*

- Verb:
  - *rasha* = 'to argue' (intransitive), 'to curse' (transitive)
  - *kunda* = 'to fall in love'
  - *kaδa* = 'to fight' (intransitive), 'to beat' (transitive)
- *-na* = intransitivity marker

Problem 6.2a

βichanirasha	=	'They will curse me.'
kuchanikunda	=	'You$_{SG}$ will fall in love with me.'
dichakurasha	=	'We will curse you$_{SG}$.'
βichamukunda	=	'They will fall in love with you$_{PL}$.'
muchadikaδa	=	'You$_{PL}$ will beat us.'
βichakaδana	=	'They will fight.'
dicharashana	=	'We will argue.'
dichakaδana	=	'We will fight.'

Problem 6.2b
1. 'I will beat you$_{SG}$.'
2. 'They will fall in love.'

Problem 6.2c
3. *mucharashana*
4. *kuchaβirasha*

6.6 Patterns for arguments

If the arguments (subject and object) are not marked by a single morpheme, it is possible that they will be marked by a combination of morphemes. Thus, they can be split into individual morphemes representing the person (1, 2, or 3), number (singular, dual, plural), or gender (masculine, feminine). Therefore, if an argument cannot be encountered as a single morpheme, we need to check whether there are any correlations between individual variables. Moreover, it is rather common for the third person to be unmarked, i.e., not to have a specific morpheme.

Problem 6.3

Ge'ez (Peter Arkadiev, TurLom 2007)

Here are some verbal forms in Ge'ez and their English translations:

tawalada	'He was born.'
tawaladu	'They were born.'
tawaladna	'We were born.'
tawaladkəmu	'You$_{PL}$ were born.'
qatalkəwo	'I killed him.'
qatalkomu	'You$_{SG}$ killed them$_{M}$.'
qatalomu	'He killed them$_{M}$.'
qatalon	'He killed them$_{F}$.'
qatalnon	'We killed them$_{F}$.'
qatalkəməwon	'You$_{PL}$ killed them$_{F}$.'
qataləwo	'They$_{M}$ killed him.'
qataləwomu	'They$_{M}$ killed them$_{M}$.'

Problem 6.3a Translate into English:

1. *taʷaladku*
2. *qatalkəʷon*
3. *qatalo*

Problem 6.3b Translate into Ge'ez:

4. 'You$_{SG}$ were born.'
5. 'You$_{PL}$ killed him.'
6. 'We killed them$_{M}$.'
7. 'They$_{M}$ killed them$_{F}$.'

The subscripts M and F refer to masculine and feminine, respectively.

Solution

We easily notice the verb stem: *taʷalad-* = 'to be born' and *qatal-* = 'to kill'. Moreover, checking the English translations, we notice that the remaining morpheme needs to mark the subject and the object (all the other parameters are constant – tense, aspect, mood, etc.).

Since, at first glance, we cannot identify any patterns, we include all these morphemes in a table in which we mark the combinations:

O/S	1SG	2SG	3SG.M	1PL	2PL	3PL.M
∅			*-a*	*-na*	*-kəmu*	*-u*
3SG.M	*-kəʷo*					*-əʷo*
3PL.M		*-komu*	*-omu*			*-əʷomu*
3PL.F			*-on*	*-non*	*-kəməʷon*	

Looking at the row in which the object is 3PL.M ('them$_{M}$'), we notice that all the entries end in *-omu*. Therefore, we can assume that this morpheme marks

the 3PL.M object. Similarly, we can separate the object marker for 3PL.F = *-on*. If we separate these markers, the table becomes:

O/S	1SG	2SG	3SG.M	1PL	2PL	3PL.M
∅			*-a*	*-na*	*-kəmu*	*-u*
3SG.M	*-kəwo*					*-əwo*
3PL.M		*-k-**omu***	***-omu***			*-əw-**omu***
3PL.F			***-on***	*-n-**on***	*-kəməw-**on***	

If we look at the column 3SG.M, we notice that the subject marker is null if there is an object, and it is *-a* if there is no object (if the verb is intransitive). Therefore, we can assume that, when an object marker is added, $a \longrightarrow \emptyset$.

Similarly, looking at the columns 2PL and 3PL.M, we deduce that when an object marker is added, $u \longrightarrow əw$. Based on these two phonological rules, we can fill in the table with all the other combinations of subject and object:

O/S	1SG	2SG	3SG.M	1PL	2PL	3PL.M
∅	*-ku*	*-ka*	*-a*	*-na*	*-kəmu*	*-u*
3SG.M	*-kəwo*	*-ko*	*-o*	*-no*	*-kəməwo*	*-əwo*
3PL.M	*-kəwomu*	*-komu*	*-omu*	*-nomu*	*-kəməwomu*	*-əwomu*
3PL.F	*-kəwon*	*-kon*	*-on*	*-non*	*-kəməwon*	*-əwon*

Thus, we can write the rules and solve the tasks:

Rules:

- Structure: V-S-O
- Verb: *tawalad-* = 'to be born', *qatal-* = 'to kill'
- S and O:

	1	2	3M	3F
SG	S: *-ku*	S: *-ka*	S: *-a*, O: *-o*	
PL	S: *-na*	S: *-kəmu*	S: *-u*, O: *-omu*	O: *-on*

When an object morpheme is added, the final vowel of the subject changes: $a \longrightarrow \emptyset$ and $u \longrightarrow əw$.

Problem 6.3a

1. 'I was born.'
2. 'I killed them$_{F}$.'
3. 'He killed him.'

Problem 6.3b

4. *tawaladka*
5. *qatalkəməwo*
6. *qatalnomu*
7. *qatalәwon*

Problem 6.4

Itelmen (Yakov Testelets, MSK 1998)

Here are some verbal forms in Itelmen and their English translations:

aniaḳzovŏmnen	'He was asking me.'
nk'aniaḳzozvŏmnen	'They would ask me.'
naniaḳzone?n	'They were asking them.'
k'añchpnen	'He would have taught him.'
nañchpvŏmne?n	'They have taught us.'
añchpḳzozne?n	'He teaches them.'

Problem 6.4a Translate into English:

1. *añchpne?n*
2. *nk'aniaḳzovŏmnen*
3. *nanianen*

Problem 6.4b Translate into Itelmen:

1. 'He has asked them.'
2. 'They ask us.'
3. 'They would have taught me.'
4. 'He would ask him.'

Solution

The first step is to segment the verbs. In this process, two of the morphemes can be a bit problematic:

1. The final morpheme (*-nen* / *-ne?n*): perhaps, at first sight, we would be tempted to say that *-ne-* is a separate morpheme which appears in all examples, followed by the morpheme *-ʔ-*, which is optional, and finally followed by *-n* which appears in all examples. Although this explanation is correct and is applicable to all examples, it unnecessarily overcomplicates the verb structure, for which reason it is more convenient to treat the whole structure as if it was a single morpheme.

2. A similar problem appears in the case of the morpheme *-z-* which is sometimes placed after the morpheme *-ḳzo-*. Although we would perhaps be tempted to analyse it as a separate morpheme, we notice that it always appears after *-ḳzo-* and nowhere else. Therefore, we will analyse *-ḳzoz-* as a whole, not treating *-z-* as a separate morpheme.

Of course, these observations are preliminary. If it turns out that this hypothesis does not work, we can come back and re-segment the verbs.

Based on this, the verb segmentation is:

ania-ḳzo-vŏm-nen	'He was asking me.'
n-k'-ania-ḳzoz-vŏm-nen	'They would ask me.'
n-ania-ḳzo-ne?n	'They were asking them.'
k'-añchp-nen	'He would have taught him.'
n-añchp-vŏm-ne?n	'They have taught us.'
añchp-ḳzoz-ne?n	'He teaches them.'

And the Itelmen verb structure can be written as:

I	II	III	IV	V	VI
n-	*-k'-*	*-ania-*	*-ḳzo-*	*-vŏm-*	*-nen*
∅	∅	*-añchp-*	*-ḳzoz-*	∅	*ne?n*
			∅		

Since only morphemes III and VI are pervasive, it is very likely that one of these is the stem. Morpheme VI has only two forms, which are highly similar to

one another, so, most likely, morpheme III is the stem. Indeed, based on the given examples, we can suggest that *-ania-* = 'to ask' and *-añchp-* = 'to teach'.

Moreover, looking at the examples that contain morpheme I (*n-*), we notice that it marks the 3PL subject.

Note

We cannot know for sure whether this morpheme marks the subject 3PL or only that the subject is plural (not necessarily 3rd person), since, in all examples, the subject is 3rd person.

Morpheme II occurs only in the structures 'They would ask me' and 'He would have taught him', and these two have in common the (conditional) mood. Moreover, we notice that there are no other examples in this mood.

Therefore, we can summarise our findings:

I	II	III	IV	V	VI
n- = S3pl	*-k'-* = Cond.	*-ania-* = 'to ask'	*-ḳzo-*	*-vŏm-*	*-nen*
∅ = S3sg	∅ = Ind.	*-añchp-* = 'to teach'	*-ḳzoz-*	∅	*ne?n*
			∅		

For the morpheme V, we can make a table in which we compare the examples that contain that morpheme, and those which do not:

-vŏm-	∅
'He was asking me.'	'They were asking them.'
'They would ask me.'	'He would have taught him.'
'They have taught us.'	'He teaches them.'

The only difference we can notice is that the morpheme occurs every time the object is in the 1st person (singular or plural). Therefore, we deduce that morpheme V marks the person of the object (*-vŏm-* = O1 and ∅ = O3). Moreover,

since this morpheme only marks the person (not the number), we expect that one of the remaining morphemes marks the number. Looking at morpheme VI, we indeed notice that it marks the object's number: *-nen* = OSG, *-ne?n* = OPL.

I	II	III	IV	V	VI
n- = S3pl ∅ = S3sg	*-k'-* = Cond. ∅ = Ind.	*-ania-* = 'to ask' *-añchp-* = 'to teach'	*-ḳzo-* *-ḳzoz-* ∅	*-vŏm-* = O1 ∅ = O3	*-nen* = Osg *ne?n* = Opl

The only unidentified morpheme is morpheme IV and, again, we can make a table in order to notice where each form occurs:

-ḳzo-	*-ḳzoz-*	∅
'He was asking me.' 'They were asking them.'	'They would ask me.' 'He teaches them.'	'He would have taught him.' 'They have taught us.'

Based on what we have discovered so far (subject, object, mood), we expect that this morpheme will mark something related to tense and/or aspect. Therefore, we can easily notice that the morpheme *-ḳzoz-* marks the present tense. The other two morphemes both mark a past tense, but they discriminate two different aspects: *-ḳzo-* marks the imperfective (or the continuous aspect), while the null morpheme marks the perfective (or the perfect aspect).

Based on this, we can fill in the table above with all the meanings of the morphemes. When writing the rules, we propose a different version, which, although longer, is preferred in this situation since the table can become extremely wide, thus needing to be split into multiple rows.

Rules:

Verb structure

1. Subject: *n-* = 3PL, ∅ = 3SG
2. Mood: *-k'-* = Conditional, ∅ = Indicative
3. Stem: *-ania-* = 'to ask', *-añchp-* = 'to teach'
4. Tense/Aspect: *-ḳzo-* = Imperfective, *-ḳzoz-* = Present, ∅ = Perfective

5. Object – person: *-vŏm-* = 1, ∅ = 3
6. Object – number: *-nen* = SG, *-neʔn* = PL

Problem 6.4a

1. 'He has taught them.'
2. 'They would have been asking me.'
3. 'They have asked him.'

Problem 6.4b

4. *anianeʔn*
5. *naniaḳzozvŏmneʔn*
6. *nk'añchpvŏmnen*
7. *k'aniaḳzoznen*

In this way, we can start noticing some common patterns in this type of problem. We can have split patterns, in which a single concept or grammatical category in English is expressed using two or more morphemes (for example the person and number of an argument or the tense and mood) or combined patterns, in which two or more concepts in English are fused into a single morpheme for example, subject and object or tense and negation.

6.7 Pronoun hierarchy

In many languages that we are familiar with, the subject and object are distinguished solely based on the word or morpheme order. For example, the difference between *The dog sees the cat* and *The cat sees the dog* is strictly due to the word order. Since the subject is, generally, the first in the sentence, we know that in the first sentence *the dog* is the subject and *the cat* is the object. Swapping the positions of the two, we also change their roles: so that in the second sentence, *the cat* becomes the subject, while *the dog* is the object.

Even in the problems we have solved up to now, the same rules have applied. If we look back at Problem 6.1 (Swahili), we remember that the subject was always placed first, while the object was placed third. Thus, the structure *ninawuona* (*ni-na-wu-ona* = 1SG-present-2SG-'see') was translated as 'I see you$_{SG}$', but, if we reverse the order of morphemes I and III (*wunaniona*), we get 'you$_{SG}$ see me'.

Nevertheless, some languages work quite differently. The order of the two pronoun morphemes can be completely independent of their roles (subject or object), but rather depends on a predefined hierarchy of the persons together with

another morpheme which shows whether the subject and object follow that hierarchy or not. The most common pronominal hierarchies are 1 > 2 > 3 (meaning that the 1st person is prioritised over the 2nd, which is then prioritised over the 3rd) and 2 > 1 > 3. We can notice that, in both cases, the 3rd person is the lowest in the hierarchy. Moreover, the difference between the two hierarchies has socio-cultural implications (in languages with a 1 > 2 > 3 hierarchy, the 1st person, or the speaker, is hierarchically superior, meaning that we can talk about a speaker-focused language, while the languages with 2 > 1 > 3 hierarchy are listener-oriented).

Let us consider the above-mentioned Swahili examples and also imagine a language Y which has a 1 > 2 > 3 hierarchy. In the table below, T = tense, V = verb stem.

	'I see you$_{SG}$'	'You$_{SG}$ see me'
Swahili	*ni-na-wu-ona*	*wu-na-ni-ona*
	1SG-T-2SG-V	2SG-T-1SG-V
Y	1SG-2SG-V-X	1SG-2SG-V-X′

If we analyse the (artificial) examples from language Y, we notice, in this case, that the order of the two pronominal morphemes is identical (1st person appears before 2nd person, since it is hierarchically superior). The two examples do not differ in the order of two morphemes, but rather in the morpheme X (or X′), which shows that the hierarchy is or is not observed respectively. Thus, in the example 'I see you$_{SG}$', morpheme X shows that the hierarchy is respected (i.e., that the person of the subject is hierarchically superior to that of the object), while in the sentence 'you$_{SG}$ see me', the morpheme X′ shows the opposite: the hierarchy is not followed, since the person of the object is hierarchically superior to the person of the subject.

Problem 6.5

Proto-Algonquian (Heather Newell, UKLO 2017)

Here are some verbal forms in Proto-Algonquian (in a simplified transcription) and their English translations:

1.	*kewa:pameθehm*	'I see you$_{SG}$.'
2.	*kewa:pameθehmwa:*	'I see you$_{PL}$.'
3.	*newa:pama:ehma*	'I see him.'
4.	*newa:pama:ehmaki*	'I see them.'
5.	*kewa:pameθehmwa:ena:n*	'We see you$_{PL}$.'
6.	*newa:pama:ehmena:na*	'We see him.'
7.	*kewa:pamiehm*	'You$_{SG}$ see me.'
8.	*kewa:pama:ehma*	'You$_{SG}$ see him.'
9.	*kewa:pamiehmwa:*	'You$_{PL}$ see me.'
10.	*kewa:pamiehmwa:ena:n*	'You$_{PL}$ see us.'
11.	*newa:pamekwehmena:naki*	'They see us.'

Problem 6.5a Translate into English: *kewa:pamiehmena:n*.

Problem 6.5b Translate into Proto-Algonquian: 'We see them' and 'They see me'.

The mark *:* after a vowel denotes length. *θ* = 'th' in 'thin'. All 'we' pronouns in this problem refer to 'we$_{EXCL}$' ('we exclusive', meaning 'me' and 'them', not including the listener).

Solution

After segmentation, we obtain the following verb structure in Proto-Algonquian:

I	II	III	IV	V
k-	*ewa:pam*	*-eθ-*	*-ehm-*	*-a*
n-		*-a:-*		*-aki*
		-i-		*-wa:ena:n*
		-ekw-		*-ena:na*
				-ena:naki
				∅

Morphemes II and IV are constant, so we do not need to offer a translation for them. Nevertheless, we can easily assume that one of them represents the stem ('to see') and the other one the tense (present). As mentioned at the beginning of this chapter, the stem tends to be the longest morpheme, so we can assume that morpheme II is the stem (of the verb), while morpheme IV represents the tense (present indicative).

Morpheme I: we make a table in order to highlight the contrast between the two possible forms:

k-	*n-*
'I see you$_{SG}$.'	'I see him.'
'I see you$_{PL}$.'	'I see them.'
'We see you$_{PL}$.'	'We see him.'
'You$_{SG}$ see him.'	'They see us.'
'You$_{SG}$ see me.'	
'You$_{PL}$ see me.'	
'You$_{PL}$ see us.'	

We infer that *k-* marks the presence of the 2nd person (either as a subject or as an object), while *n-* marks the absence thereof.

Analysing morpheme V, we infer that *-ena:n-* = 1PL, *-wa:-* = 2PL, *-a-* = 3SG, *-aki-* = 3PL. Moreover, we notice that persons 1SG and 2SG are unmarked. One of the interesting elements is that these morphemes have a specific order, independent of their role as a subject or object, namely *-wa:-* is always the first, followed by *-ena:n-* and, finally, by *-a-* or *-aki-*. In other words, it seems as if these morphemes are placed following a pronominal hierarchy: 2 > 1 > 3. The fact that this hierarchy is focused on the listener (2nd person) is also supported by the fact that the first morpheme refers strictly to the 2nd person.

The only morpheme left to analyse is morpheme III. In order to figure out its function (although we expect it to be the additional morpheme pointing towards whether the hierarchy is respected or not), we make the following table:

-eθ-	*-a:-*	*-i-*	*-ekw-*
'I see you$_{SG}$.'	'I see him.'	'You$_{SG}$ see me.'	'They see us.'
'I see you$_{PL}$.'	'I see them.'	'You$_{PL}$ see me.'	
'We see you$_{PL}$.'	'We see him.'	'You$_{PL}$ see us.'	
	'You$_{SG}$ see him.'		

Since in the case of pronominal hierarchies the number is usually not relevant, but only the person, we can reduce the table above to only the essential data, i.e., the person of the subject and the object:

-eθ-	*-a:-*	*-i-*	*-ekw-*
S1 O2	S1 O3 S2 O3	S2 O1	S3 O1

Even if the concept of pronoun hierarchy is unknown, the table above can help us solve most of the problem. Nevertheless, if, for example, we were asked to translate an example like S3 O2, we would not know which form of morpheme III to use.

Another possible explanation for these data is that morpheme *-a:-* is used if the object is in the 3rd person (independent of the subject) and *-ekw-* is used for S3 (independent of the object). This rule would force the example S3 O2 to use the morpheme *-ekw-*.

A more thorough approach would be to use the notion of pronominal hierarchy. We already know that this language has a 2 > 1 > 3 hierarchy, so the morphemes *-eθ-* and *-a:-* show that the object is superior to the subject (S < O), while morphemes *-i-* and *-ekw-* show that S > O. In order to discriminate between the two, we can add another parameter, namely the presence of the 3rd person (as either subject or object), in a similar way to how morpheme I works.

Thus, morpheme III can be explained in the following table:

	∃3	∄3
S > O	*-a:-*	*-i-*
S < O	*-ekw-*	*-eθ-*

Note

The symbols ∃ and ∄ are equivalent to “there is, there exists” and “there isn’t, there does not exist” respectively. These symbols do not need an explanation/legend when used.

Now we can structure all the rules:

Rules (Option 1).

- Verb structure: A–*ewa:pam*–B–*ehm*–C
- A: *n-*, if there is no 2nd person; *k-*, if there is 2nd person.
- B: info about the hierarchy of subject and object (considering 2 > 1 > 3), as well as the presence of 3rd person.

	∃3	∄3
S > O	*-a:-*	*-i-*
S < O	*-ekw-*	*-eθ-*

- C: pronoun markers: *-ena:n-* = 1PL, *-wa:-* = 2PL, *-a-* = 3SG, *-aki-* = 3PL. These are added hierarchically (2 > 1 > 3). Persons 1SG and 2SG are unmarked.

Rules (Option 2). A more succinct way of writing the rules is as a table:

k- (∃2) *n-* (∄2)	*ewa:pam*	*-eθ-* (∄3, S<O) *-i-* (∄3, S>O) *-ekw-* (∃3, S<O) *-a:-* (∃3, S>O)	*-ehm-*	*-wa:-* (2PL)	*-ena:n-* (1PL)	*-a* (3SG) *-aki* (3PL)
I	II	III	IV	V		

Note

We divided morpheme V into three positions in order to show that 2nd person needs always to be before 1st person, which needs always to be before 3rd person.

Based on these rules, we can solve the tasks:

Solution 6.5a 'You$_{SG}$ see us.'

Solution 6.5b 'We see them.' = *newa:pama:ehmena:naki*
'They see me.' = *newa:pamekwehmaki*

6.8 Verb semantics

Different English verbs can have the same stem in some languages, using some additional morphemes to alter the meaning. Let us remember from Problem 6.2 that the pairs 'to argue' – 'to curse' and 'to beat' – 'to fight' use the same verb stem in Dabida, although they are completely different in English. In this case, the meaning depends on the transitivity of the verb. The most common three types of morphemes which affect the verb semantics are:

- Intensifiers: show that an action is performed "excessively" or "a lot" (compare, for example, *to eat – to devour*, or *to hurt – to kill*).
- Mitigators: are the opposite of intensifiers and show that the intensity of an action is diminished (compare, for example, *to eat – to taste* or *to see – to glance*).
- Causative marker: would be translated into English as *to make someone do…*. For example, *to eat* can become *to feed* ('to make someone eat') or *to learn* ⟶ 'to make someone learn' = *to teach*).

Moreover, certain languages can display a dichotomy (a contrast) between stative and active verbs. Active (or dynamic) verbs involve movement, an action which is performed (e.g., *to hit*, *to kill*, *to look*, *to say*, *to eat*, etc.), while stative verbs show a state or a feeling (*to love*, *to hate*, *to like*, *to be good*, *to be ill*, etc.). Certain languages treat these two verbal categories quite differently, using different TAM markers or even morpheme order.

Other semantic considerations involve the notion of *agent* (*A*) and *patient* (*P*). These two terms are complementary to the pair subject-object. The agent is defined as the argument which performs the action, while the patient is that upon whom it is acted. Although these definitions may seem similar to those of the subject and object, these two arguments (agent and patient) are purely semantic. Thus, in the sentence *The cat eats the fish*, the subject is *the cat*, while the object is *the fish*. Moreover, *the cat* is also the agent (because it performs the action), while *the fish* is the patient (since it undergoes the action). In this case, therefore, $S = A$ and $O = P$.

On the other hand, in the sentence *The fish is eaten by the cat*, the (grammatical) subject is *the fish* (since it displays concord with the verb), while the (prepositional) object is *the cat*. Nevertheless, from a purely semantic point of view, the agent is still *the cat* (it is still the cat that performs the action), while *the fish* is the patient.

Thus, we can consider the subject and object as *syntactic* arguments (which depend strictly on the sentence structure), while agent and patient are *semantic* arguments (they do not depend on the sentence structure, but strictly on the meaning of it).

Some languages can mark the argument of an intransitive verb either as a patient or as an agent. The choice is made by the speaker based on semantic considerations. The agentive marking (i.e., the argument of the intransitive verb is marked as agent) is used to show that the action is undertaken by the subject, while the patientive marking is used to show that the action is rather undergone by the subject.

For example, in English this patientive marking is usually implied by using the structures *to get X* (e.g., *to get lost*, *to get drowned*, etc.). Another example is verbs that indicate accidental actions, for which a patientive marking would probably be favoured (*to fall*, *to trip*, etc.).

This type of marking is usually referred to as *fluid-S* alignment (which is a type of morphosyntactic alignment, as shown in the next chapter). There are two subtypes of fluid-S alignment:

- *Agentive-default* in which the "default" (neutral) way of marking the intransitive argument is as agent. In this case, the patientive marking is used to highlight a lack of volition / control.
- *Patientive-default* on the other hand, uses patient marking as default. By using the agentive marking, the speaker highlights a certain degree of volition / control.

Problem 6.6

Tabasaran (Yakov Testelets, MSK 1998)

Here are some verbal forms in Tabasaran and their English translations. The verb endings have been separated from the verb root with a hyphen.

bisun-čačʷu	‘we caught you$_{\text{PL}}$’	*ilbicun-za*	‘I turned’
ʁapun-čʷa	‘you$_{\text{PL}}$ said’	*kčʷuχun-vu*	‘you$_{\text{SG}}$ slipped’
ʁarʁun-zu	‘I froze’	*kčʷuχun-za*	‘I sleighed’
ʁiliχun-čʷa	‘you$_{\text{PL}}$ worked’	*uldugun-zu*	‘I got lost’
ʁipʼun-za	‘I ate’	*ursun-čʷa*	‘you$_{\text{PL}}$ jumped’
ʁurčːʷun-vazu	‘you$_{\text{SG}}$ beat me’	*qergun-zu*	‘I woke up’
daqun-za	‘I stretched’	*šadʁaxun-čʷu*	‘you$_{\text{PL}}$ got happy’
duʁmišʁaxun-zu	‘I was born’	*ergun-vu*	‘you$_{\text{SG}}$ got tired’

Problem 6.6a You are given some additional verb roots:

aqun = ‘to fall’, *ʁilirqʼun* = ‘to get scared’, *učʼʷun* = ‘to enter’

Translate into English:

1. *aqun-za*
2. *aqun-zu*
3. *bisun-čʷazu*
4. *ʁilirqʼun-ču*
5. *učʼʷun-va*

Problem 6.6b You are given some additional verb roots:

ʁalrʁun = ‘to inflate’, *dusun* = ‘to stay’, *ʁergun* = ‘to escape’

Translate into Tabasaran:

6. ‘you$_{\text{SG}}$ escaped’
7. ‘I got scared’
8. ‘I beat you$_{\text{SG}}$’
9. ‘we inflated’
10. ‘we stayed’

Solution

It is easy to notice that the arguments are marked at the end by the morphemes *č* (1PL), *čʷ* (2PL), *z* (1SG), *v* (2SG). The only challenge in this problem is figuring out the choice between the vowels *a* and *u* which follow them.

We notice that in transitive phrases (which have both subject and object), the subject always has the vowel *a*, while the object has the vowel *u*. Therefore, for transitive verbs, the structure of the phrase is $\boxed{\textit{V-SaOu}}$.

For intransitive verbs, we notice the examples *kčʷuχun-vu* and *kčʷuχun-za*, which have the same verbal stem in Tabasaran, but are translated as ‘to slip’ and ‘to sleigh’. Moreover, they differ only in the final vowel *a* vs. *u*. We notice that the difference between the two verbs is the degree of volition: ‘to slip’ is an involuntary or accidental action, while ‘to sleigh’ is used to describe a voluntary

action, done on purpose, over which we have control. We therefore notice that the vowel *a* is used if the action is done on purpose (it represents an agentive marker), while *u* is used if the action is done by mistake or unintentionally (patientive marker). Moreover, this reasoning also explains the choice of vowels in the case of transitive constructions, since, in these, the subject is the same as the agent and the object is the same as the patient.

Therefore, the rules are:

- Structure: $\boxed{V\text{-}S\mathbf{x}(O\mathbf{x})}$
- **x**: *a* = agent marker, *u* = patient marker
- S = O: *z* = 1sg, *v* = 2sg, *č* = 1PL, *č*^{w} = 2pl, or as a table:

	1	2
SG	*z*	*v*
PL	*č*	*č*^{w}

Answers: 1. and 2.: in both cases, the stem is 'to fall', but one is marked as agentive, while the other is marked as patientive. The action of falling is accidental, so example 2 is translated as 'I fell'. In order to find the right verb for the agentive marking, we need to think of the signification of falling: you are standing and then you end up sitting or lying. Therefore, a suitable verb for it is 'to sit', 'to lie'. Thus:

Solution 6.6a

1. 'I sat'
2. 'I fell'
3. 'you$_{\text{PL}}$ caught me'
4. 'we got scared'
5. 'you$_{\text{SG}}$ entered'

Solution 6.6b

6. *ʁergun-va*
7. *ʁilirq'un-zu*
8. *ʁurč:*^{w}*un-zavu*
9. *ʁalrʁun-ča*
10. *dusun-ča*

6.9 Practice problems

Problem 6.7

Swahili (Catherine Sheard, UKLO 2014)

Here are two sets of Swahili verbs and their English translations, *in random order*:

Set I.

1.	*Alikula.*	A.	'He ate.'
2.	*Atacheza.*	B.	'He will play.'
3.	*Mlifahamu.*	C.	'I eat.'
4.	*Mnapika.*	D.	'I played.'
5.	*Nilicheza.*	E.	'I cook.'
6.	*Ninakula.*	F.	'I will cook.'
7.	*Ninapika.*	G.	'They understand.'
8.	*Nitapika.*	H.	'They will cook.'
9.	*Tulifahamu.*	I.	'They played.'
10.	*Unacheza.*	J.	'We understood.'
11.	*Utapika.*	K.	'You$_{\text{PL}}$ understood.'
12.	*Wanafahamu.*	L.	'You$_{\text{PL}}$ cook.'
13.	*Watapika.*	M.	'You$_{\text{SG}}$ play.'
14.	*Walicheza.*	N.	'You$_{\text{SG}}$ will cook.'

Set II.

15.	*Hakucheza.*	O.	'He did not play.'
16.	*Hamkupika.*	P.	'He will not cook.'
17.	*Hamli.*	Q.	'He will not play.'
18.	*Hatacheza.*	R.	'I did not play.'
19.	*Hatapika.*	S.	'I do not eat.'
20.	*Hatukufahamu.*	T.	'I will not fear.'
21.	*Hatupiki.*	U.	'They do not fear.'
22.	*Hawachi.*	V.	'They do not understand.'
23.	*Hawafahamu.*	W.	'We did not understand.'
24.	*Huchezi.*	X.	'We do not cook.'

25.	*Sikucheza.*	Y.	'You$_{PL}$ do not eat.'
26.	*Sili.*	Z.	'You$_{PL}$ did not cook.'
27.	*Sitakucha.*	AA.	'You$_{SG}$ do not play.'

Problem 6.7a Determine the correct correspondences for each of the two sets.

Problem 6.7b Knowing that *Ninatembelea* means 'I visit' and *Ninakufa* means 'I die', translate into Swahili:

28. 'You$_{SG}$ visit.'	32. 'He dies.'
29. 'You$_{SG}$ do not visit.'	33. 'He does not die.'
30. 'You$_{SG}$ did not visit.'	34. 'He died.'
31. 'You$_{SG}$ will visit.'	35. 'He will not die.'

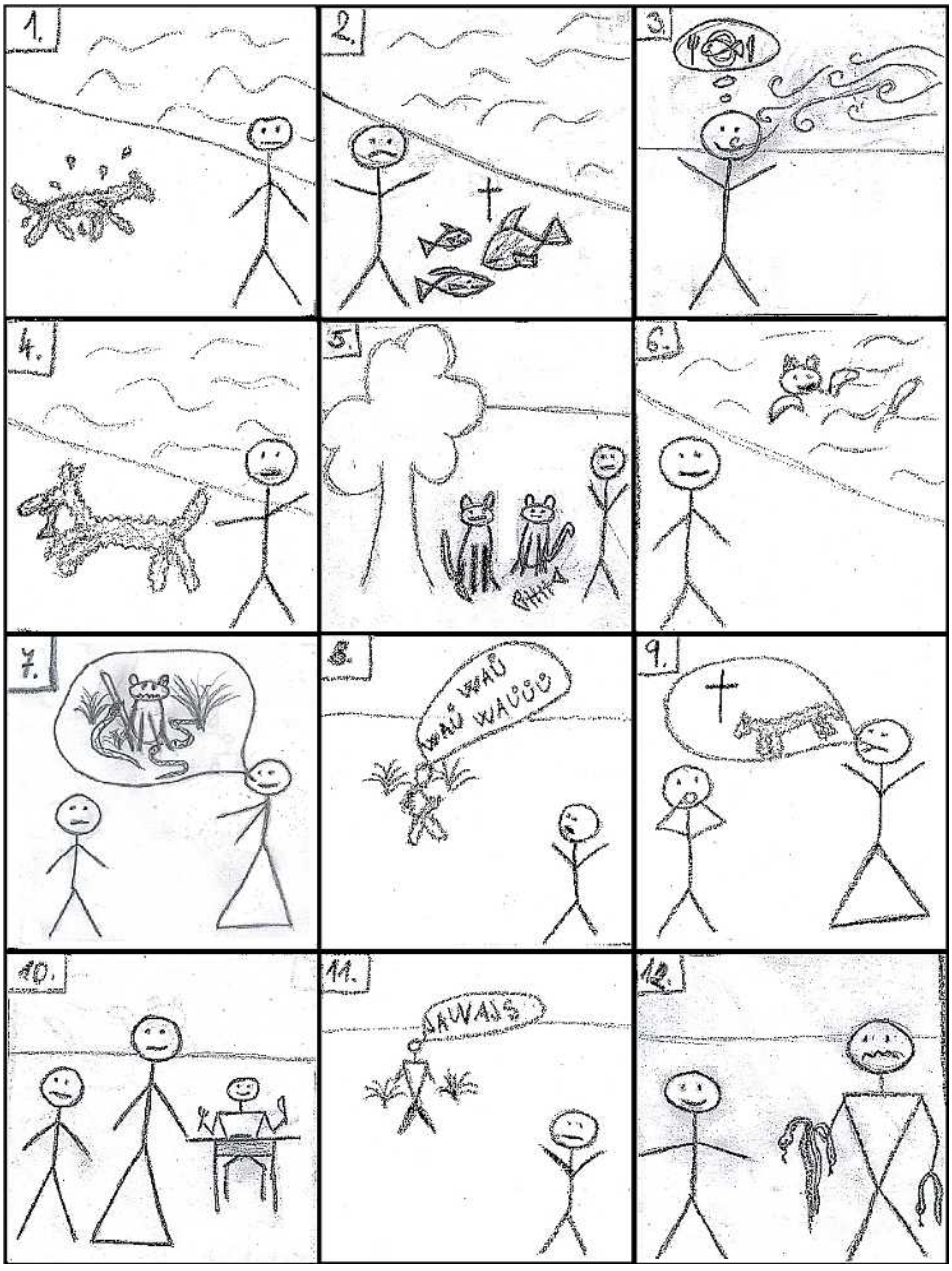
Problem 6.8

Tariana (Michaela Svatošová, ČLO 2019)

On the following page we can see some drawings depicting Jovino's experiences on a summer morning.

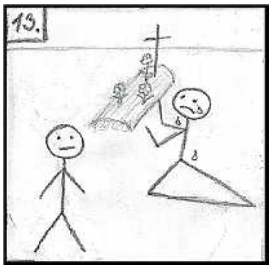
Later that day, Jovino met his brother Gracilian and told him what happened (the English sentences correspond to the numbers in the figures above; the Tariana sentences are given *in random order*):

1. 'My dog bathed.'	A. *nahã pisanane kuphe nañhasika*
2. 'The fish$_{PL}$ died.'	B. *ʧãri kuphene diyanamahka*
3. 'The man cooked the fish$_{PL}$.'	C. *pisana dipitaka*
4. 'The dog stole his fish$_{SG}$.'	D. *mawari nuhã ʧinu diwhãmahka*
5. 'Their cats ate the fish$_{SG}$.'	E. *duhã ʧãri mawarine diinusika*
6. 'The cat bathed.'	F. *mawarine nahã pisana nawhãpidaka*
7. 'The snakes bit their cat.'	G. *nuhã ʧinu dipitasika*
8. 'The snake bit my dog.'	H. *mawari ʧãri diwhãmahka*
9. 'My dog died.'	I. *kuphene nayãmika*
10. 'Her husband ate.'	J. *nuhã ʧinu diyãmipidaka*
11. 'The snake bit the man.'	K. *duhã ʧãri diñhaka*
12. 'Her husband killed the snakes.'	L. *ʧinu dihã kuphe diituka*

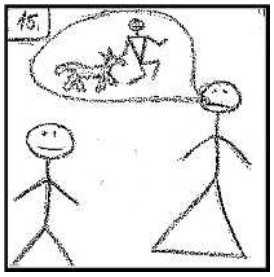


Problem 6.8a Determine the correct correspondences.

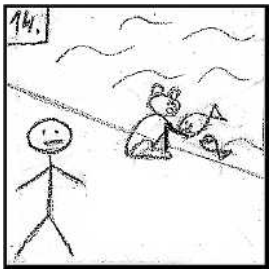
Problem 6.8b How would Jovino describe the following situations in Tariana?



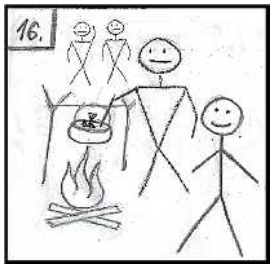
13. 'Her cat died.'



15. 'Her husband stole their dog.'



14. 'The fish$_{PL}$ bit my cat.'



16. 'The men cooked my fish$_{SG}$.'

Problem 6.8c Translate into English the following sentences and explain in which situation might Jovino utter them:

17. *dihã tʃinune nañhamahka*
18. *pisana nahã kuphene diinuka*
19. *mawaɾi tʃãɾi diwhãsika*
20. *duhã kuphene nañhapidaka*

Problem 6.9

Gee (Paul Helmer, RoLO 2018)

Here are some verbal forms in Gee and their English translations *in random order*:

1.	*baipaʔmedo*	A.	'I came first.'
2.	*baitenedoleʔ*	B.	'Will you$_{PL}$ not come first?'
3.	*biʔʃunerisa*	C.	'You$_{SG}$ do not go.'
4.	*biʔteme*	D.	'You$_{PL}$ will only talk.'

5.	*biʔtemirisadoleʔ*	E.	‘I only ran.’
6.	*dospaʔmi*	F.	‘Will I not go?’
7.	*dospaʔmiduʔaleʔ*	G.	‘Do you$_{\text{PL}}$ only run?’
8.	*dosʃuneduʔa*	H.	‘You$_{\text{SG}}$ only talked.’
9.	*dosʃuneduʔadoleʔ*	I.	‘You$_{\text{SG}}$ will not come.’
10.	*meʔpaʔmerisaleʔ*	J.	‘Do you$_{\text{SG}}$ talk first?’
11.	*meʔʃumeduʔa*	K.	‘Did I not just run?’
12.	*meʔtemiduʔa*	L.	‘You$_{\text{PL}}$ run.’

Problem 6.9a Determine the correct correspondences.

Problem 6.9b Translate into English:

13. *meʔpaʔmi*	15. *biʔʃunidoleʔ*
14. *baiʃune*	16. *meʔtemeleʔ*

Problem 6.9c Translate into Gee:

17. ‘Do I talk?’	20. ‘Do we just not come?’
18. ‘You$_{\text{SG}}$ will only run.’	21. ‘You$_{\text{SG}}$ talk first.’
19. ‘You$_{\text{PL}}$ did not go first.’	22. ‘Will I not run first?’

Problem 6.10

Gyarung (Svetlana Britova, MSK 1998)

Here are some sentences in Gyarung and their English translations:

1.	*ŋəñe no tast’on*	‘We will take care of you$_{\text{SG}}$.’
2.	*no ŋəñe kəust’oi*	‘You$_{\text{SG}}$ will take care of us.’
3.	*wəjonžəskə ŋəñe wəst’oi*	‘They two will take care of us.’
4.	*ŋənže ño tast’oñ*	‘We two will take care of you$_{\text{PL}}$.’
5.	*wəjoñek nžo təust’onč*	‘They will take care of you two.’
6.	*wəjok ŋəñe wəst’oi*	‘He will take care of us.’
7.	*wəjok no təust’on*	‘He will take care of you$_{\text{SG}}$.’
8.	*wəjonžəskə ño təust’oñ*	‘They two will take care of you$_{\text{PL}}$.’

9.	*ŋəñe ño tast'oñ*	'We will take care of you$_{PL}$.'
10.	*wəjoñek ŋənʒe wəst'očh*	'They will take care of us two.'
11.	*ño ŋa kəust'oŋ*	'You$_{PL}$ will take care of me.'

Problem 6.10a Translate into English:

12. *no ŋa kəust'oŋ*
13. *wəjonʒəskə no təust'on*
14. *ño ŋənʒe kəust'očh*

For tasks (b) and (c), you are given the following additional sentences:

15. *wəjonʒəskə ŋənʒe nɐrə ño t'has wəst'oi*
'They two will take care of us two and you$_{PL}$ together.'
16. *wəjok no nɐrə wəjoñe t'has təust'oñ*
'He will take care of you$_{SG}$ and them together.'
17. *no ŋənʒe nɐrə wəjo t'has kəust'oi*
'You$_{SG}$ will take care of us two and him together.'

Problem 6.10b Here is a Gyarung sentence, in which *a single word* is missing:

18. *ŋəñe* _____ *nɐrə wəjo t'has tast'onč*

Fill in the missing word and translate the sentence into English.

Problem 6.10c Translate into Gyarung:

19. 'I will take care of you two.'
20. 'They two will take care of me.'
21. 'They will take care of me and you two together.'
22. 'You$_{SG}$ will take care of me and him together.'

č, *ñ*, *ŋ*, *t'*, *t'h* and *ʒ* are consonants; *ɐ* and *ə* are vowels.

Problem 6.11

Hakhun (Peter Arkadiev, IOL 2018)

Here are some sentences in Hakhun and their English translations:

1.	*ŋa ka kɤ ne*	'Do I go?'
2.	*nɤ ʒip tuʔ ne*	'Did you$_{SG}$ sleep?'
3.	*ŋabə ati lapkʰi tɤʔ ne*	'Did I see him?'
4.	*nirum kəmə nuʔrum cʰam ki ne*	'Do we know you$_{PL}$?'
5.	*nɤbə ŋa lapkʰi rɤ ne*	'Do you$_{SG}$ see me?'
6.	*tarum kəmə nɤ lan tʰu ne*	'Did they beat you$_{SG}$?'
7.	*nuʔrum kəmə ati lapkʰi kan ne*	'Do you$_{PL}$ see him?'
8.	*nɤbə ati cʰam tuʔ ne*	'Did you$_{SG}$ know him?'
9.	*tarum kəmə nirum lapkʰi ri ne*	'Do they see us?'
10.	*ati kəmə ŋa lapkʰi tʰɤ ne*	'Did he see me?'

Problem 6.11a Translate into English:

11. *nɤ ʒip ku ne*
12. *ati kəmə nirum lapkʰi tʰi ne*
13. *tarum kəmə nuʔrum cʰam ran ne*
14. *nirum kəmə tarum lan ki ne*
15. *nirum kəmə nɤ cʰam tiʔ ne*
16. *nirum ka tiʔ ne*

Problem 6.11b Translate into Hakhun:

17. 'Did I beat you$_{SG}$?'
18. 'Did they seem me?'
19. 'Does he know you$_{SG}$?'
20. 'Do you$_{PL}$ sleep?'

cʰ, *kʰ*, *ŋ*, *tʰ*, *ʒ* and *ʔ* are consonants; *ə* and *ɤ* are vowels.

Problem 6.12

Cree (Ivan Derzhanski, MSK 2008)

Here are some verbal forms in Cree (in the Plains Cree dialect) and their English translations:

1.	*kiwīminahitin*	'I want to make you drink.'
2.	*ninanāskomik*	'He thanks me.'
3.	*kiwīminahāw*	'You want to make him drink.'
4.	*kikīwāpamin*	'You saw me.'
5.	*nikīminahikwak*	'They made me drink.'
6.	*nikananāskomāw*	'I will thank him.'
7.	*kikīnanāskomik*	'He thanked you.'
8.	*kikaminahāwak*	'You will make them drink.'

Problem 6.12a Translate into English:

9. *niwīwāpamāwak*
10. *kiminahin*
11. *ninanāskomikwak*
12. *kikawāpamik*

Problem 6.12b Translate into Cree:

13. 'I saw you.'
14. 'I want to thank him.'
15. 'You will thank me.'
16. 'I make them drink.'
17. 'He wants to see me.'
18. 'They see you.'

A bar above a vowel denotes length.

Problem 6.13

Ainu (Vlad A. Neacșu, RoLO 2021)

Here are some sentences in Ainu (in the Shizunai dialect) and their English translations:

1.	*ikupa as wa isam*	'We drank.'
2.	*inkartek an wa an*	'I was glancing.'
3.	*e inkar wa an*	'You$_{SG}$ were seeing.'
4.	*inu wa isam*	'He listened.'
5.	*iperepa wa oka*	'They were feeding.'
6.	*e ipe wa an*	'You$_{SG}$ were eating.'
7.	*eci inuruypa wa oka*	'You$_{PL}$ were listening a lot.'
8.	*cie koretek wa isam*	'We lent you$_{SG}$.'
9.	*cieci nukarruypa wa isam*	'We stared at you$_{PL}$.'
10.	*eun nurepa wa oka*	'You$_{SG}$ were telling us.'
11.	*un etekpa wa oka*	'He was tasting us.'
12.	*ecien nutek wa an*	'You$_{PL}$ were listening to me a little.'
13.	*an yaynu wa isam*	'I thought.'
14.	*an eruypa wa oka*	'I was devouring them.'
15.	*inuruypa as wa isam*	'We listened a lot.'
16.	*en e wa an*	'They were eating me.'
17.	*e yaykore wa isam*	'You$_{SG}$ gave yourself.'
18.	*cieci nurepa wa oka*	'We were telling you$_{PL}$.'

Problem 6.13a Translate into English in all possible ways:

19. *e nukarepa wa isam*
20. *e koreruy wa an*
21. *ci yaynukarpa wa oka*
22. *nuruypa wa isam*
23. *iperuy an wa isam*

Problem 6.13b Translate into Ainu:

24. 'He was listening to you$_{PL}$.'
25. 'We ate.'
26. 'You$_{SG}$ were thinking a lot.'
27. 'They were staring.'
28. 'We were borrowing him.'
29. 'I glanced at them.'
30. 'You$_{PL}$ fed yourselves.'
31. 'You$_{SG}$ were chattering to us.'

Problem 6.14

Rotokas (Theodor Cucu, RoLO 2019)

Here are some verbal forms in Rotokas and their English translations *in random order*:

1.	*aloravirovo*	A.	'I went'
2.	*iparaepa*	B.	'I threw it'
3.	*ourovo*	C.	'I just talked'
4.	*oraoupaveiepa*	D.	'I just devoured it'
5.	*orareoveiepo*	E.	'I just confessed'
6.	*reoraepo*	F.	'he moved it'
7.	*reoraviroepo*	G.	'he just took it'
8.	*rupupaveiepo*	H.	'we two just discussed'
9.	*rururova*	I.	'we two were just swimming'
10.	*vikirava*	J.	'we two were getting married'

Problem 6.14a Determine the correct correspondences, knowing that:

oraruruveiepa = 'we two moved ourselves'
aloparovo = 'he was just eating it'

Problem 6.14b Translate into English:

11. *rupuraepo*
12. *ouparava*
13. *reoparoepa*
14. *oraruruveviroepa*

Problem 6.14c Translate into Rotokas:

15. 'I was devouring him'
16. 'he was just getting married'
17. 'we two just jumped'
18. 'we two arrived'
19. 'we two ate it'

Problem 6.15

Dinka (Michal Láznička, ČLO 2019)

Here are some sentences in Dinka (in the Agar dialect) and their English translations:

1.	*dàam báng*	'Do I catch the chief?'
2.	*báng àdɔ́ɔm jò*	'It's the chief that the dog catches.'
3.	*rów àpíik wèng*	'It's the hippo that the cow pushes.'
4.	*rów àdɔ̀m jó*	'The hippo catches the dog.'
5.	*tèɛɛt wéng*	'Do I curse the cow?'
6.	*ghêɛn àgèl rów*	'I protect the hippo.'
7.	*tèeet kwàc ghêɛn*	'Does the leopard curse me?'
8.	*gèeet bàng jó*	'Does the chief cook the dog?'
9.	*jó àgéeet bàng*	'It's the dog that the chief cooks.'
10.	*gèel jò kwác*	'Does the dog protect the leopard?'
11.	*jó àlɔ̀ɔk ê*	'The dog washes him.'
12.	*gɔ̀ɔɔr kwác*	'Do I seek the leopard?'
13.	*báng àgòor kwác*	'The chief seeks the leopard.'
14.	*rów àwèc wéng*	'The hippo hits the cow.'
15.	*wéng àwèc ê*	'The cow hits him.'

Problem 6.15a Translate into Dinka:

16. 'I cook the leopard.'
17. 'Do I cook the dog?'
18. 'The leopard washes the hippo.'
19. 'It's the leopard that the chief washes.'
20. 'Do I wash the chief?'
21. 'I hit him.'
22. 'Does the hippo push the dog?'
23. 'Does the cow hit the dog?'
24. 'The chief curses him.'
25. 'It's me that the hippo protects.'

Vowel doubling and tripling denotes length (short *a*, medium *aa*, long *aaa*). The marks ◌̀, ◌́, and ◌̂ above the vowel denote low, high,

and falling tones respectively.

ɛ and *ɔ* are vowels similar to *e* and *o* respectively, but pronounced with a more open mouth (but less open than *a*).

The language features two types of vowel phonation, but they were not included in the problem for simplicity.

6.10 Solutions of practice problems

Solution for practice problem 6.7. Swahili

Solution 6.7a Set I:

1. A.	4. L.	7. E.	10. M.	13. H.
2. B.	5. D.	8. F.	11. N.	14. I.
3. K.	6. C.	9. J.	12. G.	

Set II:

15. O.	18. Q.	21. X.	24. AA.	27. T.
16. Z.	19. P.	22. U.	25. R.	
17. Y.	20. W.	23. V.	26. S.	

Solution 6.7b

28. *unatembelea*	31. *utatembelea*	34. *alikufa*
29. *hutembelei*	32. *anakufa*	35. *hatakufa*
30. *hukutembelea*	33. *hafi*	

Rules:

1. Negation: *ha-*
2. Subject:

	1	2	3
SG	*ni*	*u*	*a*
PL	*tu*	*m*	*wa*

3. Tense/Negation:

	Past	Present	Future
Affirmative	*li*	*na*	*ta*
Negative	*ku*	∅	*ta*

4. Stem

   Other changes:

   - Negative marker can merge with the subject marker:
     - *ha-* + *-ni-* ⟶ *si-*
     - *ha-* + *-V-* ⟶ *hV-* (*V* = *a* or *u*)
   - For present negative:
     - last vowel of the stem (*a*) becomes *i*.
     - if the stem starts with the morpheme *ku-*, it gets dropped.

Solution for practice problem 6.8. Tariana

Solution 6.8a 1. G. 3. B. 5. A. 7. F. 9. J. 11. H.
2. I. 4. L. 6. C. 8. D. 10. K. 12. E.

Solution 6.8b
13. *duhã pisana diyãmisika*
14. *kuphene nuhã pisana nawhãka*
15. *duhã tʃãri nahã tʃinu diitupidaka*
16. *tʃãrine nuhã kuphe nayanaka*

Solution 6.8c
17. 'His dogs ate.' (Jovino heard the dogs eating, tearing the meat.)
18. 'The cat killed their fish$_{PL}$.' (Jovino witnessed the action.)
19. 'The snake bit the man.' (Jovino saw the bleeding wound.)
20. 'Her fish$_{PL}$ ate.' (Someone told that to Jovino.)

Rules:

- Word order: SOV, Possessor-Possessed
- Possessives: *nuhã* = 1SG, *dihã* = 3SG masc, *duhã* = 3SG fem, *nahã* = 3PL
- *-ne* = plural (for nouns)
- Verb:
  1. Subject number (*di-* = SG, *na-* = PL)
  2. Stem
  3. Evidential markers:
     - ∅ = visual (Jovino was a witness.)
     - *-mah-* = sensory non-visual (hearing/smell)
     - *-si-* = inferential (Jovino sees the result of the action.)
     - *-pida-* = reportative (Jovino finds out about it from someone else.)
  4. *-ka* – Alternatively, this mark can be combined with the evidential markers, resulting in: *-ka*, *-mahka*, *-sika*, *-pidaka*).

Solution for practice problem 6.9. Gee

Solution 6.9a

1. C.	4. I.	7. G.	10. J.
2. F.	5. B.	8. E.	11. H.
3. A.	6. L.	9. K.	12. D.

Solution 6.9b

13. 'You$_{PL}$ talk.'
14. 'Did we not come?'
15. 'I went.'
16. 'Will you$_{SG}$ talk?'

Solution 6.9c

17. *me?pa?nele?*
18. *dostemedu?a*
19. *baiʃumirisado*
20. *bi?pa?nidu?adole?*
21. *me?pa?merisa*
22. *dostenerisadole?*

		Subject				
Stem	Tense	Pers.	Number	Adverb	Neg	Q
biiʔ (‘come’) *bai* (‘go’) *dos* (*run*) *meʔ* (‘talk’)	*ʃu* (past) *paʔ* (present) *te* (future)	*n* (1) *m* (2)	*e* (SG) *i* (PL)	*duʔa* (‘only’) *risa* (‘first’)	*do*	*leʔ*

Solution for practice problem 6.10. Gyarung

Solution 6.10a

12. ‘You$_{SG}$ will take care of me.’
13. ‘They two will take care of you$_{SG}$.’
14. ‘You$_{PL}$ will take care of us two.’

Solution 6.10b 18. Missing word: *no*
Translation: ‘We will take care of you$_{SG}$ and him together.’

Solution 6.10c

19. *ŋa nʒo tast’ončh*
20. *wəjonʒəskə ŋa wəst’oŋ*
21. *wəjoñek ŋa nɐrə nʒo t’has wəst’oi*
22. *no ŋa nɐrə wəjo t’has kəust’očh*

Rules:

- Structure: S O (*nɐrə* O’ *t’has*) V
  For complex sentences, V refers to the sum of the objects (you$_{SG}$ + me = us two; you$_{SG}$ + us two = us; him + you$_{SG}$ = you two, etc.)

- Pronouns (S = O):

	SG	DU	PL
1	*ŋa*	*ŋənžje*	*ŋañe*
2	*no*	*nžo*	*ño*
3	*wəjok*†	*wəjonžəskə*	*wəjoñek*†

† final *k* is dropped if it's an object

- Verb:
  1. Information about 2nd person: *t* = O2, *k* = S2, *w* = ∄2
  2. Subject and object person: *a* = S1–O2, *ə* = S3–O1, *ə* = S2–O1 or S3–O2

Note

An alternative (and simpler) explanation can combine the first two morphemes into one. We can write: *ta* = S1–O2, *təu* = S3–O2, *wə* = S3–O1, *kəu* = S2–O1.

  3. *st'o* – it most likely represents the stem, possibly including the TAM marker
  4. Information about the subject:

	SG	DU	PL
1	*ŋ*	*čh*	*i*
2	*n*	*nčh*	*ñ*

Solution for practice problem 6.11. Hakhun

Solution 6.11a

11. 'Do you$_{SG}$ sleep?'
12. 'Did he see us?'
13. 'Do they know you$_{PL}$?'
14. 'Do we beat them?'
15. 'Did we know you$_{SG}$?'
16. 'Did we go?'

Solution 6.11b

17. *ŋabə nɤ lan tɤʔ ne*
18. *tarum kəmə ŋa lapkʰi tʰɤ ne*
19. *ati kəmə nɤ cʰam ru ne*
20. *nuʔrum ʒip kan ne*

Rules:

- Word order: S O V X *ne*
- S: in transitive sentences, it receives the suffix *-bə* (if S is 1SG or 2SG) or the word *kəmə* placed after the subject (otherwise).
- X: hierarchy 1 > 2 > 3:

t- -ʔ	past, S>O	*ɤ*	∃1SG
tʰ-	past, S<O	*i*	∃1PL
k-	present, S>O	*u*	∄1 & ∃2SG
r-	present, S>O	*an*	∄1 & ∃2PL

Solution for practice problem 6.12. Cree

Solution 6.12a

9. 'I want to see them.'
10. 'You make me drink.'
11. 'They thank me.'
12. 'He will see you.'

Solution 6.12b

13. *kikīwāpamitin*
14. *niwīnanāskomāw*
15. *kikananāskomin*
16. *niminahāwak*
17. *niwīwāpamik*
18. *kiwāpamikwak*

Rules:

Option 1 – Detailed

1. Existence of 2SG:
   - *ki* = exists (S or O)
   - *ni* = does not exist
2. TAM markers:
   - *wī* = volitive ('to want to')
   - *kī* = past
   - ∅ = present
   - *ka* = future
3. Stem
   - *minah* = 'to make drink'
   - *nanāskom* = 'to thank'
   - *wāpam* = 'to see'
4. Information about 3rd person
   - *ik* = S3SG
   - *ikwak* = S3PL
   - *āw* = O3SG
   - *āwak* = O3PL
   - *in* = ∄3 & S2SG
   - *itin* = ∄3 & O2SG

Option 2 – Condensed

2nd pers.	TAM	Stem	S/O
∃ ⟶ *ki* ∄ ⟶ *ni*	∅ = present *wī* = volitive *kī* = past *ka* = future	*minah* = 'to make drink' *nanāskom* = 'to thank' *wāpam* = 'to see'	*-ik*(*wak*) = S3SG(PL) *-āw*(*ak*) = O3SG(PL) *-in* = ∄3 & S2SG *-itin* = ∄3 & O2SG

Solution for practice problem 6.13. Ainu

Solution 6.13a

19. 'You$_{SG}$ showed them.'
20. 'He was giving a lot to you$_{SG}$.' / 'They were giving a lot to you$_{SG}$.' / 'You$_{SG}$ were giving a lot to him.'
    You can also use other ways to express 'give a lot', such as 'be generous' etc.
21. 'We were seeing ourselves.'
22. 'He listened a lot to them.' / 'They listened a lot to them.'
23. 'I ate a lot.'

Solution 6.13b

24. *eci nupa wa oka*
25. *ipepa as wa isam*
26. *e yaynuruy wa an*
27. *inkarruypa wa oka*
28. *ci koretekre wa an*
29. *an nukartekpa wa isam*
30. *eci yayerepa wa isam*
31. *eun nureruypa wa oka*

Rules:

- *Tense/Aspect*: placed at the end of the structure.
  - past simple (perfective): *wa isam*
  - past continuous (imperfective):
    - *wa an* – if the verb is considered singular[3]
    - *wa oka* – if the verb is considered plural[4]

[3]The plurality of the verb is determined by the subject (for intransitive verbs) or by the object (for transitive verbs). In other words, the verb plurality follows an ergative marking – see Section 7.4.

[4]See previous footnote.

- *Pronoun markers*

Person	S (Intransitive)	S (Transitive)	Object
1 sg	*-an*	*an-*	*en-*
2 sg	*e-*	*e-*	*e-*
3 sg	∅	∅	∅
1 pl	*-as*	*ci-*	*un-*
2 pl	*eci-*	*eci-*	*eci-*
3 pl	∅	∅	∅

  The hyphen that precedes or follows the morpheme shows its position with respect to the verb (before or after). In the case of transitive verbs, if both arguments are placed before the verb, they are placed in the order subject – object and they fuse together into a single word.

- *Marker for verb plurality* (as defined above): suffix *-pa* is placed after the verbal affixes (if any).

- *Verbal affixes*:
  - *yay-* = reflexive ('to think' = 'to hear yourself')
  - *-ruy* = intensifier (can be translated by 'a lot', but it can also be lexicalised in the choice of verb: 'to eat' – 'to devour', 'to see' – 'to stare')
  - *-tek* = mitigator (the opposite of *-ruy*: 'to eat' – 'to taste', 'to see' – 'to glance')
  - *–(r)e*: causative ('to eat' ⟶ to make someone eat = 'to feed', 'to see' ⟶ to make someone see = 'to show'). Additionally, *-re* ⟶ *e / r _*

- *Verbal stems*: Each verb has two different stems, for transitive and intransitive:

English	Intransitive	Transitive
'to eat'	*ipe*	*e*
'to see'	*inkar*	*nukar*
'to listen'	*inu*	*nu*
'to drink'	*iku*	
'to give'		*kore*[a]

[a] In reality, the stem *kore* ('to give') comes from the stem *kor* ('to have') + causative marker ('to have' ⟶ to make someone have = 'to give').

Solution for practice problem 6.14. Rotokas

Solution 6.14a

1. D.	3. G.	5. H.	7. E.	9. F.
2. A.	4. J.	6. C.	8. I.	10. B.

Solution 6.14b

11. 'I just swam'
12. 'I was taking him'
13. 'he was talking'
14. 'we two emigrated'

Solution 6.14c

15. *aloparavirova*
16. *oraouparoepo*
17. *oravikiveiepo*
18. *ipaveviroepa*
19. *aloveva*

Rules:

1. *ora-* = reciprocal / reflexive ('to move' → 'to move oneself'; 'to take' → to take oneself = 'to marry'; 'to speak' → 'to discuss'; 'to throw' → 'to jump');
2. Stem: *-alo-* = 'to eat', *-ipa-* = 'to go', *-ou-* = 'to take', *-reo-* = 'to speak', *-rupu-* = 'to swim', *-ruru-* = 'to move', *-viki-* = 'to throw';
3. *-pa-* = imperfect (progressive aspect);
4. Subject: *-ra-* = 1SG, *-ro-* = 3SG, *-ve-* = 1d;
5. *-viro-* = intensifier / to do till the end ('to eat' → 'to devour'; 'to talk' → 'to confess'; 'to walk' → 'to arrive' = to walk till the end; 'to move oneself' → 'to emigrate');
6. Transitivity: *-v-* = transitive, *-(i)ep-* = intransitive (*ep* → *iep* / *e* _);
7. *-a* = far past, *-o* = recent past ('just').

Note

In this language, the reflexive is considered intransitive (it takes the marker *-ep-*), while in Ainu (previous problem), the reflexive is considered transitive (it uses the transitive form of the stem).

Solution for practice problem 6.15. Dinka

Solution 6.15a

16. *ghêɛn àgèet kwác*
17. *gɛ̀ɛɛt jó*
18. *kwác àlɔ̀ɔk róẃ*
19. *kwác àlɔ́ɔɔk bàng*
20. *làaak báng*
21. *ghêɛn àwèc ê*
22. *pìik ròẃ jó*
23. *wèec wèng jó*
24. *báng àtèet ê*
25. *ghêɛn àgéel ròẃ*

Rules:

- Sentence structure:
  - Active (e.g., 'The dog catches the hippo.') – word order S V O
  - Active with focused object (e.g., 'It's the hippo that the dog catches.') – word order O V S;
    the tone of the subject becomes low.
  - Interrogative (e.g., 'Does the dog catch the hippo?') – word order V S O;
    the tone of the subject becomes low.

- Verb

  Stems:

  dɔ̀m = ‘to catch’ *gèet* = ‘to cook’ *lɔ̀ɔk* = ‘to wash’ *tèet* = ‘to curse’
  gèl = ‘to protect’ *gòor* = ‘to seek’ *pìk* = ‘to push’ *wèc* = ‘to hit’

  Active:
  Add prefix *à-*.

  Interrogative:
  Vowels lengthen by one degree (short ⟶ medium ⟶ long).
  If the subject is 1SG, vowel opens by one degree (*i* ⟶ *e* ⟶ *ɛ* ⟶ *a* and *u* ⟶ *o* ⟶ *ɔ* ⟶ *a*).

  Active (focused object):
  Add prefix *à-*.
  The vowel lengthens by one degree.
  The tone becomes high.

Further reading

Donohue, Mark & Søren Wichmann (eds.). 2005. *The typology of semantic alignment*. Oxford: Oxford University Press.
Fortescue, Michael, Marianne Mithun & Nicholas Evans (eds.). 2017. *The Oxford handbook of polysynthesis*. Oxford: Oxford University Press.
Hengeveld, Kees, Heiko Narrog & Hella Olbertz (eds.). 2017. *The grammaticalization of tense, aspect, modality and evidentiality*. Berlin: De Gruyter Mouton.

7 Syntax

7.1 Introduction

Syntax is concerned with the study of sentences and phrases. We remember that sentences can typically be considered to be formed from a noun phrase and a verb phrase (sentence = NP + VP)[1] (though not necessarily in that order). Therefore, syntax problems are nothing more than a noun morphology problem combined with a verb morphology problem.

Nevertheless, since NP and VP are related in the same structure, there can be some additional interactions between the two. For example, the noun phrase can contain morphemes indicating the roles of nouns in the sentence (subject, object, agent, patient, etc.). Since the arguments can be expressed though nouns (not only pronouns, as was the case in the previous chapter), there can also be a wider variety of distinctions in the verb phrase. For verb phrase problems, the principal distinctions were person (1, 2, 3), number (singular, dual, plural) and, sometimes, gender (masculine, feminine), but for syntax problems we can also add distinctions such as ±human (difference between nouns which refer to humans versus the rest).

7.2 Word order

Word order is one of the most important phenomena in syntax problems. While in the NP and VP problems the number of words is relatively small, in sentences and phrases the number of words can increase considerably; therefore the order in which these words are placed plays a very important role.

Generally, when talking about the word order of a language, we refer to the order of the subject, verb, and object; therefore, we can have six possible patterns: SOV, SVO, VSO, VOS, OSV, OVS. Cross-linguistically, the SOV and SVO patterns are the most common; they account for patterns in around 90% of the world's languages. The next one is VSO, which occurs in approx. 8% of languages (such as Classical Arabic, Tagalog, or Celtic languages like Irish, Welsh, Breton,

[1]The abbreviations used are: NP = noun phrase, VP = verb phrase.

and Manx). The least common patterns are VOS, OVS and OSV, the latter being encountered in fewer than 0.5% of languages.[2] If we look closely, we notice that the most common patterns (SOV, SVO, VSO) are those in which the subject is placed before the object, while the languages in which the subject is placed after the verb account for under 5% of the languages. Thus, it is statistically probable that in a problem the subject is placed before the object.

Moreover, it is important to mention the dichotomy between languages with fixed or rigid word order and those with a flexible word order.

Fixed word order can be directly assigned to one of the six patterns above, i.e., when a language has fixed word order, (almost) every sentence follows the same word order. All the other word orders are either ungrammatical or rarely used. Languages with fixed word order often have a simpler (inflectional) morphology in subjects and objects since the role of the arguments is well determined by their position in the sentence. Thus, if we consider the English examples *The cat saw the boy* and *The boy saw the cat*, we notice that both sentences have the same structure and the only difference is the word order. Nevertheless, there is no ambiguity between subject and object (even though they both appear morphologically identical – neither the subject nor the object is marked), since the fixed word order dictates that the subject is first (before the verb) and the object is placed after the verb.

Flexible (variable) word order, on the other hand, cannot be directly assigned to one of the six patterns above, i.e., when a language has flexible word order, not all sentences have the same word order. All six patterns (or most of them) are grammatical and commonly used. Nevertheless, there can be a dominant word order, i.e., one word order that is favoured. Unlike the languages with fixed word order, those with variable word order tend to have a much richer inflectional system. This is due to the fact that, since word order is flexible, each argument needs to be morphologically marked in order to avoid any ambiguities. Let us consider the following three examples from Romanian:

*Uriașul îl dă pe copil tatăl**ui**.*
*Tatăl**ui** îl dă uriașul pe copil.*
*Pe copil îl dă uriașul tatăl**ui**.*

All of these sentences have the same meaning ('The giant gives the child to the father.'), but the order of the arguments (subject, direct object, indirect object) differs in each case. Nevertheless, the meaning is clear and the role of each noun is

[2] Based on the data from https://wals.info/chapter/81.

well defined due to the morphological marking: the subject (*uriașul*) is unmarked, the direct object (*copil*) is preceded by *pe*, while the indirect object (*tatălui*) ends in the suffix *-ui*. In this instance, the suffix *-ui* is an example of the dative case, marking the indirect object.

Due to this morphological marking, the flexible word order does not create ambiguities in sentences such as:

Tatăl îl dă pe copil uriașului. = 'The father gives the child to the giant.'
Tatălui îl dă copilul pe uriaș. = 'The child gives the giant to the father.'
Pe tată îl dă copilului uriașul. = 'The giant gives the father to the child.'

Typologically speaking, the word order in a sentence is classified only based on subject, object, and verb. However, in writing the rules for linguistics problems, it is important to include all the components of the sentence. Thus, for the sentence *The child writes a letter with the pen*, it is not enough to write S V O, but rather we should write S V O Instr. (subject, verb, object, instrumental). Moreover, when describing the word order, we usually talk about the main constituents and not the internal structure of those constituents. This means we describe the relative positions of the subject, object, verb, instrumental, location, time, etc. As for the internal structure of the constituents, this covers the order of the noun and its modifiers (adjectives, possessives, etc.), which we can write as Adj. – Noun, or, more generally, Modifier – Noun. The reason why it is preferable to separate the two is that modifiers can occur together with different constituents (subject, object, instrumental, etc.), as in the sentence *The happy child writes a new letter with the black pen*

Thus, if we wanted to combine the two structures above (S V O Instr. and Adj. – Noun), we would have to write:

[Adj – S] V [Adj – O] [Adj – Instr]

So we would have to show that the adjective can be placed before the noun in each constituent (subject, object, instrumental). Moreover, the adjective is part of the noun phrase, so its position is strictly relative to the head of the phrase (the noun), independent of its role in the sentence (subject, object, etc.). Interestingly, in many languages the word order for the sentence is mirrored in the word order for the NP, assuming the verb/noun is the head of the VP/NP respectively. So if the sentence word order is, say verb-final, the NP order will also be head-final.

Problem 7.1

Nung (Alex Wade, UKLO 2016)

Here are some sentences in Nung and their English translations:

1. *Cáu ca vửhn nhahng kíhn.*
   'I was about to continue to eat it.'
2. *Cáu cháhn slờng páy mi?*
   'Do I truly want to go?'
3. *Cáu mi slày kíhn.*
   'I don't have to eat it.'
4. *Cáu ngám hẻht pehn tế.*
   'I did it like that just now.'
5. *Cáu tan đohc hảhn mửhng.*
   'I only see you.'
6. *Cáu vửhn nhahng bô sạhm tảhng hẻht hơn.*
   'I also continue to build the house alone.'
7. *Da kíhn!*
   'Don't eat it!'
8. *Da khải hơn!*
   'Don't sell the house!'
9. *Mửhn chống ca cháhn fải khải.*
   'Then she truly was about to have to sell it.'
10. *Mửhn mi cháhn đày non.*
    'She truly can't sleep.'
11. *Mửhn náhc-thày chống bô sạhm kíhn.*
    'Then she also just previously ate it.'
12. *Mửhng náhc-thày slờng tảhng páy.*
    'You wanted to go alone just previously.'

Problem 7.1a Translate into English:

13. *Cáu cháhn đày non.*
14. *Da páy non!*
15. *Mửhn bô sạhm mi slờng hẻht hơn mi?*
16. *Mửhn ngám bô sạhm páy hơn.*

Problem 7.1b Translate into Nung:

17. 'I wasn't about to eat it just previously.'
18. 'She didn't have to eat it alone like that just now.'

19. ‘The house truly can’t eat you.’
20. ‘Then were you also about to go just previously?’

Solution

Step 1. We notice the repetition of the first word, which we can easily correlate with the subject pronoun (*cáu* = ‘I’, *mưhng* = ‘you’, *mưhn* = ‘she’). This is also confirmed by sentence 5 where we notice that the object has the same form (therefore, S = O).

Moreover, we notice that sentences 7 and 8 both start with *da* and both are in the (negative) imperative mood. We therefore deduce that *da* is the negative imperative marker.

Step 2. In sentence 7, we have only one word remaining to be translated (*kíhn*). This must be the verb (with semantic content), so *kíhn* = ‘to eat’. This is also confirmed by sentences 1, 3, and 11.

In sentence 8, we have only two remaining words, so one must be the verb, while the other must represent the noun ‘house’. Comparing with sentence 9, we deduce that *khải* = ‘to sell’ and *hơn* = ‘house’. Moreover, we infer that for negative imperative sentences the word order is *Da* V O. Furthermore, since in sentence 7 ‘it’ is not translated, we infer that the 3SG pronominal object is unmarked.

Since we have already identified two verbs, we continue to focus on the verbs. From sentences 2 and 12, we deduce that *pảy* = ‘to go’.

Step 3. Based on sentence comparison, we can also identify most of the vocabulary, especially the adverbs: *chớng* = ‘then’, *náhc-thày* = ‘just previously’, *bô sạhm* = ‘also’, *cháhn* = ‘truly’, *tảhng* = ‘alone’.

Based on the same principle, we can identify some modal verbs: *ca* = ‘was about to’, *vữhn nhahng* = ‘continue to’.

Lastly, since we already know that *cháhn* = ‘truly’, we are left with *slờng* = ‘want to’.

Step 4. Based on the above words, sentence 6 is left with only one untranslated word, mainly *hẻht* = ‘to build’. On the other hand, we notice the same word in sentence 4, this time meaning ‘to do’ (we can assume it also represents a verb). Therefore, we deduce that in Nung, similar to many languages derived from Latin, ‘to build a house’ = ‘to do (make) a house’.

Comparing sentences 3 and 10, we infer that *mi* is a negative marker. Nevertheless, the same marker occurs in sentence 2. Since the only thing left undiscovered in sentence 2 is the interrogation, we deduce that *mi* is also an interrogative marker. Moreover, we notice that if *mi* marks a question, it will be placed at the very end of the sentence (it is rather common that the interrogative marker is placed at the end of the sentence). Therefore, *mi* represents two distinct morphemes: an interrogative marker (in which case it is placed last in the sentence) or a negative marker (in which case it is placed directly after the subject).

Sentences 3 and 10 have only one untranslated phrase, 'to have to'. Nevertheless, we notice that in Nung there are two distinct words: *slày* and *fải*. The simplest explanation for the variation between the two forms is that one is used in positive (affirmative) sentences, while the other is used in negative sentences. There can be many other explanations, such as one appears if the subject is 1SG, while the other if the subject is 3SG, but this explanation does not make sense linguistically.[3] Moreover, there are two possible interpretations regarding these two forms: either there are two completely different verbs (e.g., in English the negation of *must* is usually *not have to*), or there are two suppletive forms of the same verb, one being used in affirmative sentences and the other in negative ones.

Step 5. In sentence 10, we are left with two words: *đày* and *non*, which mean (not necessarily in this order) 'to sleep' and 'can'. Comparing the examples we have got so far, we notice that the modal verb is always placed before the semantic verb. Thus, we deduce that *đày* = 'can' and *non* = 'to sleep'.

Using a similar thought process in sentence 4, in which we need to identify the words meaning 'like that' and 'just now', we can assume that 'just now' will behave similarly to 'just previously' since they both have a similar form and meaning. Since 'just previously' appears before the verb (and even immediately after the subject), it is likely that *ngám* = 'just now' and *pehn tế́* = 'like that'.

The only words we have not identified yet are 'only' and 'to see'. Since we do not have enough details to determine which is which, we check task (a) to see whether either of them occurs in those examples, offering us

[3]The linguistic explanation for which the polarity distinction (affirmative vs. negative) is more likely to be the one that is responsible for the different verb form (rather than the subject of the sentence) is that, usually, suppletive forms are driven by intrinsic characteristics of the verb (polarity, TAM, etc.) and not by their interaction with the arguments.

additional information. Moreover, we also check task (b) to see whether we need these two words. Since these two words do not occur in any of the tasks, we can just ignore them and not try to assign them to their meaning. If, nevertheless, we would like to take a guess, we would base it on the fact that most of the adverbs are placed before the verb and that the verb is usually a single word. Therefore, we could assume that *håhn* = 'to see' and *tan đohc* = 'only'.

Step 6. Since we have identified all the vocabulary, we can solve task (a). For each of the sentences, we will first write the translation of each word (in the order in which they appear in Nung) and afterwards we write the English translation. Thus:

13. I – truly – can – sleep ⟹ 'I truly can sleep.'
14. negative imperative – go – sleep ⟹ 'Don't go to sleep!'
15. She – also – not – want – do – house – question ⟹ 'Does she also not want to build the house?'

    We take into account that 'to do a house' is translated as 'to build a house'. Moreover, we notice two things: the negation *mi* does not appear immediately after the subject as we assumed (but it still does appear before the verb). Moreover, another possible translation (and more adequate, grammatically) is 'Doesn't she want to build a house either?', since the meaning of the adverb 'also' is changed in English when in the negative (compare: *She also came – She didn't come either*).
16. she – just now – also – go – house ⟹ 'She also goes home just now'.

    In this case as well, the sentence could have been translated more literally as 'She also goes to the house just now' (and still be awarded full marks), but, generally, it is preferable to use the more natural translation.

Step 7. The only thing left to do is figure out the word order. We already know that the subject comes first. Considering the verb has a fixed position, we notice that after the verb there are only three possible morphemes: the object, the question marker, and the adverb 'like that'. We can hypothesise that the question marker is always last. Since we do not have any example in which the object and 'like that' coexist, we cannot determine

the order between them; therefore, we can assume that they occupy the same position. Consequently, we can consider the Nung word order:

Negative imperative: *Da* V O

Indicative: S [...] V O/'like that' Question

By "[...]" we mean all the adverbs and modal verbs that occur between subject and verb, whose order we are still to determine. Since we know the meaning of each of them and we are only interested in the relative order between them, we can just rewrite that part of the sentence (excluding the subject, the verb and everything after the verb). The negative imperative sentences can be excluded since their word order is clear and they do not contain any adverbs or modal verbs.

ca vửhn nhahng	'was about to continue'
cháhn	'truly'
mi slày	'not have to'
ngám	'just now'
tan đohc hảhn	'see only'
vửhn nhahng bô sạhm tảhng	'also continue alone'
chống ca cháhn fải	'then truly was about to have to'
mi cháhn đày	'truly can't'
náhc-thày chống bô sạhm	'then also just previously'
náhc-thày slờng tảhng	'want alone just previously'
cháhn đày	'truly can'
bô sạhm mi slờng	'also not want'
ngám bô sạhm	'also just now'

The second, fourth, and fifth examples can be excluded: the second and the fourth because they each have only one word, and are thus not helpful in figuring out the relative positions of the adverbs; the fifth example since we have already discussed it in Step 5 and we cannot segment it.

Moreover, since we already know the meaning of all the structures, we can just write their English translations in the order in which they appear in Nung. Thus, we get:

'was about to – continue to'
'not – have to'
'continue to – also – alone'
'then – was about to – truly – have to'
'not – truly – can'

‘just previously – then – also’
‘just previously – want to – alone’
‘truly – can’
‘also – not – want to’
‘just now – also’

What is left now is no more than a logic puzzle in which we need to arrange all of these words into an overall order which is consistent with each example above. For convenience, we refer to the words by their translations. A simple method is to start with the first word (‘was about to’). It cannot be the first one in the sentence since in the fourth row the word ‘then’ appears before it. Next, ‘then’ cannot be first since ‘just previously’ appears before it in row 6. We notice that ‘just previously’ always appears first, so we can consider it as occupying the first position. Moreover, we keep in mind that we assumed that ‘just previously’ and ‘just now’ behave similarly. Therefore, we can check whether ‘just now’ also appears in first position. It occurs in only one example and it is indeed in the first position, so we can deduce that the first position is occupied by the temporal adverb {‘just now’/‘just previously’}. Now we can delete these two adverbs from the list, as well as all examples which now contain a single word. We get:

‘was about to – continue to’
‘not – have to’
‘continue to – also – alone’
‘then – was about to – truly – have to’
‘not – truly – can’
‘then – also’
‘want to – alone’
‘truly – can’
‘also – not – want to’

Using a similar thought process, we notice that ‘then’ appears only in one example (among those left), so we can consider it to be the next in line.

So far, we have: {‘just previously’/‘just now’} ⟶ {‘then’}. We are left with:

‘was about to – continue to’
‘not – have to’
‘continue to – also – alone’
‘was about to – truly – have to’
‘not – truly – can’
‘want to – alone’

‘truly – can’
‘also – not – want to’

Now ‘was about to’ appears first, so it can be the next in line.

Note that we get the same result if we start from another word. For example, if we start with the negation ‘not’ (which appears in the second row), it cannot be the first one since, in the third row, ‘continue to’ appears before it, while ‘continue to’ cannot be first since ‘was about to’ appears before it. Thus, we again end up with ‘was about to’ as being next in line. Using the same process, we establish the overall order:

{‘just previously’ / ‘just now’} → {‘then’} → {‘was about to’} → {‘continue to’} → {‘also’} → {‘not’}

We are left with:

‘truly – have to’
‘truly – can’
‘want to – alone’
‘truly – can’

We notice that ‘truly’ appears in three out of the four remaining examples and after it we have ‘have to’/‘can’/‘want to’. We therefore deduce that the modal verbs follow it and the last word placed is the adverb ‘alone’. In order to get to this result, it is important to assume that ‘want to’ and ‘can’ behave similarly, both being modal verbs. Thus, the Nung order of adverbs/modal verbs is:

{‘just previously’ / ‘just now’}
→ {‘then’}
→ {‘was about to’}
→ {‘continue to’}
→ {‘also’}
→ {‘not’}
→ {‘truly’}
→ {‘want to’, ‘have to’, ‘can’}
→ {‘alone’}

One question that might arise is: why are the modal verbs at the end (‘want to’, ‘can’, ‘have to’) separated from the verbs ‘was about to’ and ‘continue to’, which we also referred to as modal verbs? In reality, the constructions ‘was about

to' and 'continue to' are aspectual verbs, rather than modal verbs. Therefore, in broad terms, the Nung order is: TIME → ASPECT → 'also' → 'not' → 'truly' → MODAL → 'alone'.

We need to observe that the adverb *tan đohc* = 'only' does not appear in the hierarchy above. This is explained by the fact that it appears in a single sentence and it is not accompanied by any other adverb, so it cannot be compared with any other word.

Now we can solve the task (b) and write the rules.

Rules: Word order:

Negative imperative: *Da* V O
Indicative: S [...] V O/'like that' (*mi* = Question)

[...] represents all the other adverbs/modals, which are written in the following order:

{*náhc-thày* = 'just previously'/*ngám* = 'just now'}
→ {*chống* = 'then'}
→ {*ca* = 'was about to'}
→ {*vửhn nhahng* = 'continue to'}
→ {*bô sạhm* = 'also'}
→ {*mi* = 'not'}
→ {*cháhn* = 'truly'}
→ {*slờng* = 'want to'/*fải* = 'have to'/*slày* = 'not have to'/*đày* = 'can'}
→ {*tảhng* = 'alone'}

Moreover, {*tan đohc* = 'only'} belongs to this category as well, but its place in the order cannot be determined.

Problem 7.1a

13. 'I truly can sleep.'
14. 'Don't go to sleep!'
15. 'Doesn't she want to build the house either?'
16. 'She also goes to the house just now.'

Problem 7.1b

17. *Cáu náhc-thày ca mi kíhn.*
18. *Mửhn ngám mi slày tảhng kíhn pehn tế.*
19. *Hơn mi cháhn đày kíhn mửhng.*
20. *Mửhng náhc-thày chống ca bô sạhm páy mi?*

7.3 Focusing

Focusing describes the syntactic process in which one part of the sentence is emphasized. In English, focusing can be done by intonation in speaking (compare: *He **hit** the dog* and *He hit **the dog*** In the first sentence, the emphasis is on the action of hitting, while in the latter the emphasis is on the object). Moreover, a common way to focus the object in English is by using a cleft sentence (e.g., *I saw a cat* vs. *It is a cat that I saw*).

In certain languages, focusing can be done by changing the word order (in most cases, this means that the focused part is moved nearer to the beginning of the sentence) or by using certain specific markers. For example, there might be definite or indefinite articles specific for the focused form or there can be specific morphemes which signal the fact that some words are focused. In Wolof,[4] for example, there are four sets of pronouns: subject (1SG = *man*), object (1SG = *ma*), verb-focus (1SG = *damay*), and object-focus (1SG = *laa*). The first two types (subject and object) are used by default; the verb-focus pronoun is used for the subject if the verb is emphasized, while the object-focus form is used instead of the usual object pronoun, if it is emphasized. Moreover, if the verb or the object are focused, they are moved to the beginning of the sentence. Below the four forms for 1SG and 2SG are given:

	S	O	V-focus	O-focus
1SG	*man*	*ma*	*damay*	*laa*
2SG	*yow*	*la*	*dangay*	*nga*

Let us consider the sentence 'I saw you$_{SG}$' (in Wolof, the corresponding verb is *gisoon*). We can have the following three cases:

1. Neutral sentence = no part of the sentence is focused. The word order is the default one, SOV.
   Man la gisoon. ('I saw you$_{SG}$.')
2. Sentence with focused verb = the V-focus pronoun is used to replace the subject and it is placed at the beginning of the sentence (it can be considered a focus marker), being immediately followed by the verb.
   Damay gisoon la. ('I *saw* you$_{SG}$.')

[4]This phenomenon was featured in a problem by Vlad A. Neacșu (RoLO 2019).

3. Sentence with focused object = object becomes the first in the sentence, the rest of the order is preserved.
   Nga man gisoon. ('I saw *you*$_{SG}$.')

7.4 Morphosyntactic alignment

This refers to the way in which three verbal arguments (subject of intransitive verb, subject of transitive verb, object) behave. For simplicity, in this chapter we will use the following notation: Subject of intransitive verb = Subject = S, Subject of transitive verb = Agent = A, Direct object = Object = O.

In order to better illustrate this concept, let us consider the following problem:

Problem 7.2

Morphosyntactic alignments (Vlad A. Neacșu, RoLO 2016)

The following 16 sentences represent the translations of four different English sentences into four different languages, *in random order*:

1. *Bayi jugumbil baŋgul gúdaŋgu buṛan.*
2. *'ua hi'o te tamari'i.*
3. *Bayi yaṛa buṛan.*
4. *Pol'eʔi hina nawta.*
5. *Cíq'ămqalnim peexne 'áyatne.*
6. *'ua hi'o te 'ūrī 'i te vahine.*
7. *Háma peexne.*
8. *Bayi gagara baŋgul yaṛaŋgu buṛan.*
9. *Met'aii pol'eʔ nawta.*
10. *'áyatnim peexne hámane.*
11. *'ua hi'o te tamari'i 'i te 'āva'e.*
12. *Pol'eʔi nawta.*
13. *'ua hi'o te vahine 'i te tamari'i.*
14. *Hámanim peexne hísemtuksne.*
15. *Bayi yaṛa baŋgul jugumbilŋgu buṛan.*
16. *Tsu'itsui met'ai nawta.*

Problem 7.2a Group the 16 sentences into four groups, based on the language they are in.

Problem 7.2b Group the 16 sentences into four groups, based on their meaning.

Problem 7.2c Here are eight more sentences:

17. *'ua hi'o 'i te vahine.*
18. *Met'ai pol'eʔi nawta.*

19. *Bayi gúda buṛan.*
20. *Cíq'ămqalnim peexne.*
21. *'ua te hi'o 'i te tamari vahine.*
22. *Bayi jugumbilŋgu baŋgul yaṛa buṛan.*
23. *Pol'e?i pol'e? nawta.*
24. *Peexne 'áyatnim háma.*

Out of these sentences, six are wrong. Which are these and why are they wrong?

Problem 7.2d Translate the two correct sentences from task (c) into the other three languages.

Solution

Step 1. The sentence grouping based on language can be easily done considering that all sentences in Language 1 (Dyirbal) contain the word *buṛan*, all sentences in Language 2 (Tahitian) start with *'ua hi'o*, all sentences in Language 3 (Nez-Perce) contain the word *peexne*, and all sentences in Language 4 (Wappo) end in *nawta*.

Step 2. Based on the previous grouping, we have the categories:

Lang. 1 (Dyirbal)

1. *Bayi jugumbil baŋgul gúdaŋgu buṛan.*
3. *Bayi yaṛa buṛan.*
8. *Bayi gagara baŋgul yaṛaŋgu buṛan.*
15. *Bayi yaṛa baŋgul jugumbilŋgu buṛan.*

Lang. 2 (Tahitian)

2. *'ua hi'o te tamari'i.*
6. *'ua hi'o te 'ūrī 'i te vahine.*
11. *'ua hi'o te tamari'i 'i te 'āva'e.*
13. *'ua hi'o te vahine 'i te tamari'i.*

Lang. 3 (Nez-Perce)

5. *Cíq'ămqalnim peexne 'áyatne.*

7. *Háma peexne.*

10. *'áyatnim peexne hámane.*

14. *Hámanim peexne hísemtuksne.*

Lang. 4 (Wappo)

4. *Pol'e?i hina nawta.*

9. *Met'aii pol'e? nawta.*

12. *Pol'e?i nawta.*

16. *Tsu'itsui met'ai nawta.*

We notice that in each language there is one sentence which is shorter than the others. We assume that these sentences are translations of each other, so 3 = 2 = 7 = 12.

We notice that one part of these sentences occurs in all other sentences of that language (*buṛan*, *'ua hi'o*, *peexne*, *nawta*). This represents the verb.

Step 3. Looking at the sentences in each language, we notice that in Dyirbal *bayi* and *buṛan* appear in every sentence. Expanding on this, the structure of Dyirbal sentences is:

Bayi X buṛan. ⟶ for short sentences

Bayi X baŋgul Y-ŋgu buṛan. ⟶ for long sentences

Doing the same for the other languages, we get:

Language	Short sentence	Long sentence
Dyirbal	*Bayi X buṛan.*	*Bayi Y baŋgul Z-ŋgu buṛan.*
Tahitian	*'ua hi'o te A.*	*'ua hi'o te B 'i te C.*
Nez-Perce	*M peexne.*	*N-nim peexne P-ne.*
Wappo	*R-i nawta.*	*S-i T nawta.*

Step 4. By analysing the nouns in sentences 2, 3, 7, 12, we know that *yaṛa* = *tamari'i* = *pol'e?i* = *háma*.

We notice that, in each language, these nouns appear in two more sentences. Therefore, each language has a sentence which does not contain these words, so 1 = 6 = 5 = 16.

Each of these sentences contains two nouns: one which occurs in one more sentence, and one which does not occur in any other place. Checking the sentences in which that noun also occurs, we deduce 15 = 13 = 10 = 9 and *jugumbil = vahine = met'ai = 'áyat.*

We are left with the other noun from sentences 1 = 6 = 5 = 16, so *gúda = 'ūrī = cíq'ămqal = tsu'itsu.*

Now we are left with only the sentences 8 = 11 = 14 = 4 and the nouns *gagara = 'āva'e = hísemtuks = hina.*

Thus, the sentences are:

Sentence A:

3. *Bayi yaṛa buṛan.* (Dyirbal)
2. *'ua hi'o te tamari'i.* (Tahitian)
12. *Pol'e?i nawta.* (Wappo)
7. *Háma peexne.* (Nez-Perce)

Sentence B:

1. *Bayi jugumbil baŋgul gúdaŋgu buṛan.* (Dyirbal)
6. *'ua hi'o te 'ūrī 'i te vahine.* (Tahitian)
16. *Tsu'itsui met'ai nawta.* (Wappo)
5. *Cíq'ămqalnim peexne 'áyatne.* (Nez-Perce)

Sentence C:

8. *Bayi gagara baŋgul yaṛaŋgu buṛan.* (Dyirbal)
11. *'ua hi'o te tamari'i 'i te 'āva'e.* (Tahitian)
4. *Pol'e?i hina nawta.* (Wappo)
14. *Hámanim peexne hísemtuksne.* (Nez-Perce)

Sentence D:

15. *Bayi yaṛa baŋgul jugumbilŋgu buṛan.* (Dyirbal)
13. *'ua hi'o te vahine 'i te tamari'i.* (Tahitian)
9. *Met'aii pol'e? nawta.* (Wappo)
10. *'áyatnim peexne hámane.* (Nez-Perce)

Step 5. In the table at Step 3, we marked each noun with different letters, not knowing in which order they occur in the long sentences. Now, based on the correspondences, we deduce the final sentence structure:

Language	Short sentence	Long sentence
Dyirbal	*Bayi S buṛan.*	*Bayi O baŋgul A-ŋgu buṛan.*
Tahitian	*'ua hi'o te S.*	*'ua hi'o te A 'i te O.*
Nez-Perce	*S peexne.*	*A-nim peexne O-ne.*
Wappo	*S-i nawta.*	*A-i O nawta.*

We can easily solve tasks (c) and (d):

Solution 7.2c

17. *'i* occurs
18. Word order
20 Ending of the first word
21 Word order
22 Word order
24 Word order and ending of last word

Solution 7.2d

	19.	23.
Dyirbal	*Bayi gúda buṛan.*	*Bayi yaṛa baŋgul yaṛaŋgu buṛan.*
Nez-Perce	*Cíq'ămqal peexne.*	*Hámanim peexne hámane.*
Tahitian	*'ua hi'o te 'ūrī.*	*'ua hi'o te tamari'i 'i te tamari'i.*
Wappo	*Tsu'itsui nawta.*	*Pol'eʔi pol'eʔ nawta.*

This problem allows us to better understand the fundamental difference between morphosyntactic alignments across languages. Returning to the general structure of the above sentences, we notice we have four situations:

1. In Wappo, *S* and *A* are marked identically (using the suffix *-i*), but differently from *O*. In this case, we talk about a NOMINATIVE-ACCUSATIVE ALIGNMENT. This is the most common type of alignment. In this case, *S* and *A* are the nominative arguments, while *O* is the accusative argument.

2. In Nez-Perce, *S*, *A* and *O* are all marked differently. This is, by definition, the TRIPARTITE ALIGNMENT.
3. In Tahitian, there is no difference between *S*, *A* and *O* (all three are unmarked), so we say that this language features a DIRECT ALIGNMENT.
4. In Dyirbal, *S* is marked like *O* (in this case, unmarked or using a null morpheme), but differently from *A* (which receives the suffix *-ŋgu*). This language exhibits an ERGATIVE-ABSOLUTIVE ALIGNMENT. The agent is the only ergative argument, while the absolutive arguments are the subject and the object.
5. There is another type of alignment, extremely rarely used: the TRANSITIVE alignment in which *A* and *O* are marked the same, while *S* is marked differently.

A schematic representation of the five types of alignments is shown below, where identically marked arguments are highlighted in the same colour:

Alignment	Nom-Acc	Erg-Abs	Tripartite	Direct	Transitive
S					
A					
O					

It is important to understand that the morphosyntactic alignment is intrinsic to the language. Thus, in English we cannot talk about an ergative argument simply because English does not follow an ergative-absolutive alignment. Thus, the first step is determining the type of alignment that the language follows.

In the solution to Problem 6.13, we mentioned that the plurality of the verb is determined by the subject of the intransitive verb or the object of the transitive verb. In this problem, the two arguments have the same role (determining the verb plurality), so we can combine them under the specific of the absolutive case, stating that the marker *-pa* appears if the absolutive argument of the verb is plural.

7.5 Split alignment

Certain languages can have two (or more) types of alignments, each of them appearing in a specific linguistic context. For example, Pashto has a split align-

ment: it follows a nominative-accusative alignment in the present tense, but an ergative-absolutive alignment in the past. We can compare the following examples:

Ze wlarrem.	Dai *me* woleed.	*Ze* yay woleedelem.
I went	him I saw	me he saw
'I went.'	'I saw him.'	'He saw me.'
Ze dzem.	*Ze* yay weenem.	Dai *me* weenee.
I go	I him see	he me sees
'I go.'	'I see him.'	'He sees me.'

We notice that 1SG can be expressed in two ways: *ze* and *me*. In the past tense sentences (first row), *ze* is used for subject and object, while *me* is used for agent. Thus, the past follows an ergative-absolutive alignment, with *me* being used as the ergative form of 1SG and *ze* the absolutive form. At the same time, in the present tense sentences, we notice that *ze* is used for S and A, while *me* is used for O. Thus, the present tense follows a nominative-accusative alignment, with *ze* being the nominative form of 1SG and *me* the accusative form.

Another situation in which we can talk about the coexistence of two different alignments in the same language (but which is not considered split alignment) is that in which, historically speaking, a language had a certain alignment but, in time, due to changes in morphophonology, it came to use the direct alignment (completely unmarked). This can be easily observed in English, where we talk about a nominative-accusative alignment of the pronouns (*he – him*, *I – me*), but about a direct alignment of the nouns.

In linguistics problems the ergative-absolutive alignment is commonly found (together with the nominative-accusative one), while the split alignment is usually highlighted between these two types of alignment. Although in Pashto the context of the two alignments depends solely on the tense, the distinction is motivated for other reasons too: person, discourse prominence of arguments (noun vs pronoun), etc. Every time we see both transitive and intransitive sentences in a linguistics problem, we need to take into account the possibility that the language has a different alignment. Although nominative-accusative languages are more common, especially in the West, and therefore more familiar, ergative languages represent about a quarter of all world languages, and are particularly found in less known language families, and so are proportionally more likely to occur in linguistics problems! Another feature of ergative languages is that they are almost all verb-initial or verb-final, almost never SVO.

7.6 Practice problems

Problem 7.3

Luiseño (Richard Hudson, UKLO 2012)

Here are some sentences in Luiseño written in the International Phonetic Alphabet and their English translations:

1.	*nawitmalqajwukalaqpoki:k*	'The girl does not walk home.'
2.	*ja?aʃpolo:v*	'The man is good.'
3.	*hu:?unikatqajtʃipomkat*	'The teacher is not a liar.'
4.	*haxʂuxetʃiqʂuŋa:li*	'Who hits the woman?'
5.	*ja?aʃwukalaq*	'The man walks.'
6.	*to:wqʂuʂuŋa:lihu:?unikat*	'Does the teacher see the woman?'
7.	*?iviʂuŋa:lnona:jixetʃiq*	'This woman hits my father.'
8.	*nona:jiʂuxetʃiq?iviʂuŋa:l*	'Does this woman hit my father?'
9.	*?iviʂuŋa:lxetʃiqnona:ji*	'This woman hits my father.'
10.	*hu:?unikattʃipomkat*	'The teacher is a liar.'
11.	*?ivihu:?unikatnona:jito:wq*	'This teacher sees my father.'
12.	*hu:?unikatʂuto:wqʂuŋa:li*	'Does the teacher see the woman?'

Problem 7.3a Translate into English:

13. *ja?aʃwukalaqpoki:k*
14. *xetʃiqʂuʂuŋa:linona:j*
15. *haxʂuqajtʃipomkat*
16. *ʂuŋa:liʂuto:wqhu:?unikat*

Problem 7.3b Translate into Luiseño. Use vertical lines to represent word spaces (*a*|*b*):

17. 'Is the teacher a liar?'
18. 'The teacher sees the woman.'
19. 'This girl does not see my father.'
20. 'Who is good?'

Problem 7.4

Beja (Harold Somers & Richard Hudson, NACLO 2018)

Here are some Beja sentences and their English translations *in random order*. Two Beja sentences have the same English translation.

1.	*Tak rihan.*	A.	'I saw a man that is strong.'
2.	*Yaas rihan.*	B.	'I know a man that I saw.'
3.	*Akra tak rihan.*	C.	'I saw a man that is small.'
4.	*Dabalo yaas rihan.*	D.	'I saw a small dog.'
5.	*Tak akraab rihan.*	E.	'I saw a strong man.'
6.	*Tak dabaloob rihan.*	F.	'I saw a dog.'
7.	*Tak akteen.*	G.	'I saw a man.'
8.	*Rihane tak akteen.*	H.	'I know a man.'
9.	*Tak rihaneeb akteen.*		

Problem 7.4a Determine the correct correspondences.

Problem 7.4b Here are some more words from the Beja language with their translations:

araw = 'friend', *mek* = 'donkey', *kwati* = 'happy'

Translate the following sentences into Beja. If there are different ways to translate the sentence, show all the alternatives.

10. 'I saw a donkey.'
11. 'I saw a happy man.'
12. 'I know a strong donkey.'
13. 'I saw a friend that is happy.'
14. 'I know a dog that is small.'
15. 'I saw a donkey that I know.'

Problem 7.4c Translate the following sentences into English. One of them has a mistake. Write the correct version of this sentence.

16. *Kwati mek rihan.*
17. *Akraab araw akteen.*
18. *Akteene yaas rihan.*
19. *Mek dabaloob akteen.*

Problem 7.5

Mundari (Peter Arkadiev, MSK 2014)

Here are some sentences in Mundari and their English translations:

1. *senkena-ñ*
   'I left.'
2. *koɽa-eʔ senkena*
   'The man left.'
3. *otere-m dubkena*
   'You$_{SG}$ sat on the ground.'
4. *coke-ñ lelkiʔia*
   'I saw the frog.'
5. *pulis honko-eʔ lelkedkoa*
   'The policeman saw the children.'
6. *biŋ coke-ʔ huakiʔia*
   'The snake bit the frog.'
7. *seta pulisko-eʔ huakedkoa*
   'The dog bit the policemen.'
8. *biŋ setaʔre-m sabkiʔia*
   'You$_{SG}$ caught the snake in the morning.'
9. *pulisko kumbuɽu hola-ko sabkiʔia*
   'The policemen caught the thief yesterday.'
10. *kuɽiko honko hature-ko ʈokoeʔkedkoa*
    'The women scolded the children in the village.'

Problem 7.5a Translate into English:

11. *kumbuɽuko-ko dubkena*
12. *hola-ñ senkena*
13. *biŋko-m lelkedkoa*
14. *hon seta setaʔre-ʔ ʈokoeʔkiʔia*
15. *koɽa coke-ʔ sabkiʔia*

Problem 7.5b Translate into Mundari:

16. 'They left.'
17. 'The woman sat on the ground.'
18. 'The thieves saw the men.'
19. 'The dogs bit the thief.'
20. 'He caught the frogs yesterday.'

Problem 7.6

Swahili (Harold Somers, NACLO 2011)

Here are some sentences in Swahili and their English translations:

1. *Mtu ana watoto wazuri.*
   'The man has good children.'
2. *Mto mrefu una visiwa vikubwa.*
   'The long river has large islands.'
3. *Wafalme wana vijiko vidogo.*
   'The kings have small spoons.'
4. *Watoto wabaya wana miwavuli midogo.*
   'The bad children have small umbrellas.'
5. *Kijiko kikubwa kinatosha.*
   'The large spoon is enough.'
6. *Mwavuli una mfuko mdogo.*
   'The umbrella has a small bag.'
7. *Kisiwa kikubwa kina mfalme mbaya.*
   'The large island has a bad king.'
8. *Watu wana mifuko mikubwa.*
   'The men have large bags.'
9. *Viazi vibaya vinatosha.*
   'The bad potatoes are enough.'
10. *Mtoto ana mwavuli mkubwa.*
    'The child has a large umbrella.'
11. *Mito mizuri mirefu inatosha.*
    'The good long rivers are enough.'
12. *Mtoto mdogo ana kiazi kizuri.*
    'The small child has a good potato.'

Problem 7.6a Translate into Swahili:

13. 'The small children have good spoons.'
14. 'The long umbrella is enough.'
15. 'The bad potato has a good bag.'
16. 'The good kings are enough.'
17. 'The long island has bad rivers.'
18. 'The spoons have long bags.'

Problem 7.6b If the Swahili word for 'the prince' is *mkuu*, what do you think the word for 'the princes' is? Explain.

Problem 7.7

Arabic (Grigory Durnovo, MSK 1997)

Here are some sentences in Arabic and their English translations:

1. *ʾaḥraqa lmudarrisu lḥayawāna*
   'The teacher burnt the monster.'
2. *ʾabdaʾa lqāmūsa llaðī ʾaḥraqtuhu*
   'He created the dictionary that I burnt.'
3. *ʾaðlaltu lmudarrisa llaðī ʾaṣammaka*
   'I beat [=did beat] the teacher who surprised you$_{SG}$.'
4. *ʾaxrağtu lxādima llaðī ʾaṣmamtahu*
   'I brought the servant whom you$_{SG}$ surprised.'
5. *ʾaðalla lḥayawānu lkalba llaðī ʾafazzahu*
   'The monster beat [=did beat] the dog which scared him.'

Problem 7.7a One of the sentences above is ambiguous and can be translated into English in a different way. Which sentence is it and what is the alternative translation?

Problem 7.7b Translate into English:

6. *ʾabdaʾtuhu*
7. *ʾaxrağta lmudarrisa llaðī ʾafazzaka*
8. *ʾaṣamma lxādimu lkalba llaðī ʾaðallahu lmudarrisu*
9. *ʾaḥraqtu lḥayawāna llaðī ʾaðalla lxādima*

Problem 7.7c Translate into Arabic:

10. 'You$_{SG}$ scared the servant who surprised the monster.'
11. 'The dog brought the teacher who beat [=did beat] you$_{SG}$.'
12. 'I burnt the dictionary that you$_{SG}$ created.'

ʾ, *ð*, *ğ*, *h*, *ḥ*, *q*, *ṣ*, *x* are consonants. A bar above a vowel denotes length.

Problem 7.8

Welsh (Timur Maisak, MSK 1998)

Here are some sentences in Welsh and their English translations:

1. *Mae tad canllaith gan ei fanon e.*
   'His queen has a good father.'
2. *Mae banon ganllaith gan ei blentyn e.*
   'His child has a good queen.'
3. *Mae brawd teg gan ei gyfaill e.*
   'His friend has a beautiful brother.'
4. *Mae tywysoges deg gan 'y nhad i.*
   'My father has a beautiful princess.'
5. *Mae cyfaill penffol gan 'y newynes i.*
   'My witch has a stupid friend.'
6. *Mae plentyn talentog gan 'y manon i.*
   'My queen has a talented child.'
7. *Mae dewynes gall gan 'y nghyfaill i.*
   'My friend has a wise witch.'

Problem 7.8a Translate into English:

8. *Mae banon deg gan ei frawd e.*
9. *Mae tywysoges gall gan ei ddewynes e.*
10. *Mae cyfaill canllaith gan 'y nhywysoges i.*

Problem 7.8b Translate into Welsh:

11. 'His father has a stupid princess.'
12. 'His princess has a wise father.'
13. 'My child has a talented witch.'

c = 'c' in 'car'.

Problem 7.9

Tadaksahak (Bozhidar Bozhanov, UKLO 2011)

Here are some sentences in Tadaksahak and their English translations:

1. *ayagon cidi*
   'I swallowed the salt.'
2. *atezelmez hamu*
   'He will have the meat swallowed (by someone).'
3. *atedini a*
   'He will take it.'
4. *hamu anetubuz*
   'The meat was not taken.'
5. *jifa atetukuš*
   'The corpse will be taken out.'
6. *amanokal anešukuš cidi*
   'The chief didn't have the salt taken out.'
7. *ayakaw hamu*
   'I took out the meat.'
8. *itegzem*
   'They were slaughtered.'
9. *ayasezegzem a*
   'I'm not having him slaughtered.'
10. *anešišu aryen*
    'He didn't have the water drunk (by anybody).'
11. *feji abnin aryen*
    'The sheep is drinking the water.'
12. *idumbu feji*
    'They slaughtered the sheep.'
13. *cidi atetegmi*
    'The salt will be looked for.'
14. *amanokal abtuswud*
    'The chief is being watched.'
15. *cidi asetefred*
    'The salt is not being gathered.'
16. *amanokal asegmi i*
    'The chief had them looked for.'

Problem 7.9a Translate into English:

17. *aryen anetišu*
18. *ayasuswud feji*
19. *cidi atetelmez*
20. *asedini jifa*

Problem 7.9b If the stem of the verb 'to walk' is *izuwenket*, translate into Tadaksahak:

21. 'He is having the water taken.'
22. 'I'm having them walked.'
23. 'The chief did not drink the water.'
24. 'The salt was not looked for.'
25. 'He will have the salt gathered.'

ʒ = 's' in 'vision', *š* = 'sh' in 'shop', *y* is a consonant.

Problem 7.10

Sandawe (Shen-Chang Huang, APLO 2021)

Here are some sentences in Sandawe and their English translations:

1. *!'ìnéỳsù kòŋkórìsà xéʔe̩wáá*
   'A hunter$_{F}$ brought roosters.'
2. *thíméỳsù kókósà ǀ'èésú*
   'A cook$_{F}$ skinned a hen.'
3. *múk'ùmè kókó xéʔe̩wáátshú*
   'A cow didn't bring hens.'
4. *kòŋkórì múk'ùmèʔà khàású*
   'Roosters hit [=did hit] a cow.'
5. *!'ìnéỳsò k'ámbà khàáyétshógé*
   'Apparently, hunters didn't hit a bull.'
6. *thíméỳ !'ìnéỳ xééyétshèégé*
   'Apparently, a cook$_{M}$ didn't bring a hunter$_{M}$.'

7. *!'ìnéỳsò kókógéʔà ǁ'èésú*
   'Apparently, hunters skinned a hen.'
8. *!'ìnéỳsò kókóʔà khǎʔa̰wáá*
   'Hunters hit [=did hit] hens.'
9. *kòŋkórì !'ìnéỳà xééyé*
   'A rooster brought a hunter$_{M}$.'
10. *thíméỳsù kókó khǎʔa̰wáátshúgé*
    'Apparently, a cook$_{F}$ didn't hit hens.'
11. *!'ìnéỳ thíméỳsògéà khàáʔíŋ*
    'Apparently, a hunter$_{M}$ hit [=did hit] cooks.'
12. *!'ìnéỳsò thíméỳsò ǁ'èéʔíntshó*
    'Hunters didn't skin cooks.'
13. *kòŋkórì !'ìnéỳsò xééʔíntshó*
    'Roosters didn't bring hunters.'
14. *thíméỳ kòŋkórì khǎʔa̰wáátshèé*
    'A cook$_{M}$ didn't hit roosters.'

Given below are some more words in Sandawe and their English translations:

ŋ!àméỳ = 'blacksmith$_{M}$'

bálóó = 'to herd'

théká = 'leopard (any gender)'

Problem 7.10a Translate into English:

15. *thíméỳ kòŋkórìgéà ǁ'èéyé*
16. *ŋ!àméỳsù thíméỳsùsà xéésú*
17. *k'ámbà théká khàásútshógé*
18. *múk'ùmè !'ìnéỳsòsà bálóóʔíŋ*

Problem 7.10b Translate into Sandawe:

19. 'Cooks herded hens.'
20. 'Apparently, a blacksmith$_{F}$ didn't skin leopards.'
21. 'A leopard$_{F}$ didn't herd a rooster.'
22. 'Apparently, a bull didn't bring cooks.'
23. 'Apparently, a hunter$_{M}$ brought blacksmiths.'

x, *th*, *tsh*, *kh*, *k'*, *ŋ*, *ŋ!*, *ʔ*, *!'*, and *ǁ'* are consonants. The marks ◌́, ◌̀, and ◌̌ above a vowel denote high, low and rising (low ↗ high) tones, respectively. A circle under a vowel (e.g., *ḁ*) indicates a devoiced vowel. The subscripts M and F refer to masculine and feminine, respectively.

Problem 7.11

Burushaski (Danylo Mysak, UkrLO 2019)

Here are some sentences in Burushaski and their English translations:

1. *khue gušiŋanc uwaran.*
   'These women will get tired.'
2. *ise ṣiqar iɣurci.*
   'That wasp will drown.'
3. *biṭayue amin dasin musarkan?*
   'Which girl will the shamans let in?'
4. *ɣeniṣ muwalo.*
   'The queen will fall.'
5. *ue dasiwance šugulimuc usarkan.*
   'Those girls will let the friends$_{F}$ in.'
6. *guse ɣurqune ṣiqarišo uɣarki.*
   'This frog will catch the wasps.'
7. *qhudaae ice jakuyo uyeeci.*
   'The god will see those donkeys.'
8. *khine hilese belišo uɣarki.*
   'This boy will catch the rams.'
9. *hoolalase amic talabuudomuc uyeeci?*
   'Which spiders will the butterfly see?'
10. *ue thamišue ɣeniṣanc uyaranan.*
    'Those kings will deceive the queens.'
11. *hilešue šugulo isarkan.*
    'The boys will let the friend$_{M}$ in.'
12. *ɣaṣepe khine biṭan iyarani.*
    'The magpie will deceive this shaman.'

Problem 7.11a Translate into English:

13. *ice belišo uwalan.*
14. *qhudaamuce tham iyaranan.*
15. *talabuudue khine gus muyeeci.*
16. *amin guse ɣurquyo uɣarko?*

Problem 7.11b Translate into Burushaski:

17. 'Those shamans will drown.'
18. 'Which magpies will the women catch?'
19. 'The kings will see these butterflies.'
20. 'Which friend$_{M}$ will let the boys in?'
21. 'That boy will deceive the friend$_{F}$.'
22. 'The queen will let that girl in.'
23. 'This girl will see the friends$_{M}$.'
24. 'The wasp will deceive that frog.'
25. 'Which donkey will get tired?'

ɣ, *ɟ*, *ŋ*, *ʂ*, *š*, and *ʈ* are consonants. The subscripts M and F refer to masculine and feminine, respectively.

7.7 Solutions of practice problems

Solution for practice problem 7.3. Luiseño

Solution 7.3a

13. 'The man walks home.'
14. 'Does my father hit the woman?'
15. 'Who is not a liar?'
16. 'Does the teacher see the woman?'

Solution 7.3b

17. *hu:ʔunikat | ʂu | tʃipomkat*
18. *hu:ʔunikat | to:wq | ʂuŋa:li*
19. *ʔivi | nawitmal | qaj | to:wq | nona:ji*
20. *hax | ʂu | polo:v*

Note

Any other word order is accepted, as long as it follows the rules below.

Rules:

- Flexible word order; Det. – Noun; Neg. – Verb;
- The interrogative particle is placed before the verb. If the verb is first in the sentence, the particle is placed after it; i.e., the interrogative particle is always second;
- *-i* = object marker.

Solution for practice problem 7.4. Beja

Solution 7.4a

1. G
2. F
3. E
4. D
5. A
6. C
7. H
8. B
9. B

Solution 7.4b

10. *Mek rihan.*
11. *Kwati tak rihan.*
12. *Akra mek akteen.*
13. *Araw kwatiib rihan.*
14. *Yaas dabaloob akteen.*
15. *Akteene mek rihan.* or *Mek akteeneeb rihan.*

Solution 7.4c

16. ‘I saw a happy donkey.’
17. Two options:
    - ‘I know a friend that is strong.’ (Correct: *Araw akraab akteen.*)
    - ‘I know a strong friend.’ (Correct: *Akra araw akteen.*)
18. ‘I saw a dog that I know.’
19. ‘I know a donkey that is small.’

Rules:

- Three sentence patterns (`V` = Verb, `A` = Adjective, `N` = Noun):
  1. ‘I’ `V` ‘a’ (`A`) `N`. ⟹ (`A`) `N` `V`.
  2. ‘I’ `V` ‘a’ `N` ‘that is’ `A`. ⟹ `N` `A`-V^*b `V`, where V^* refers to the last vowel of the word.
  3. ‘I’ `V` ‘a’ `N` ‘that I’ `V′`. ⟹ `V′`-*e* `N` `V` or `N` `V′`-*eeb* `V`.
- The parts separated by hyphen (-) are suffixes.

Discussion (not part of the solution)

Word order: Object—Verb, Adjective—Noun. The adjective can also act like a verb (e.g., ‘small’ – ‘to be small’).

The relative clause (introduced by ‘that’) can be placed:

- immediately after the noun (in our case, between object and verb) – in which case it receives the suffix $-V^*b$, where V^* is the final vowel of the verb);
- before the noun – in which case it receives no suffix (but when the relative clause consists of a verb, the verb receives a suffix *-e*; compare, for example, sentences 6 and 8).

If the relative clause contains a predicative expression, it can only be placed after the noun, otherwise it would simply be translated as an adjective (‘I saw a *small* donkey.’ vs. ‘I saw a donkey *that is small*.’).

Solution for practice problem 7.5. Mundari

Solution 7.5a

11. 'The thieves sat.'
12. 'I left yesterday.'
13. 'You$_{SG}$ saw the snakes.'
14. 'The child scolded the dog in the morning.'
15. 'The man caught the frog.'

Solution 7.5b

16. *senkena-ko*
17. *kuɽi otere-ʔ dubkena*
18. *kumbuɽuko koɽako-ko lelkedkoa*
19. *setako kumbuɽu-ko huakiʔia*
20. *cokeko hola-eʔ sabkedkoa*

Rules:

- Word order: S O (Location/Time) V
- Plural: *-ko* added to the end of the noun (before the hyphen).
- Verbal suffixes:
  - *-kena* = intransitive verb;
  - *-kedkoa* = transitive verb, plural object;
  - *-kiʔia* = transitive verb, singular object.
- The agreement between verb and subject is marked through a suffix separated by a hyphen. It is attached to the word before the verb. If the sentence only contains a verb (one word), it is attached to the verb instead. The forms of this suffix are:
  - *-ñ* = 1SG;
  - *-m* = 2SG;
  - *-eʔ* = 3SG (*eʔ* ⟶ *ʔ* / *e* _);
  - *-ko* = 3PL.

Solution for practice problem 7.6. Swahili

Solution 7.6a

13. *Watoto wadogo wana vijiko vizuri.*
14. *Mwavuli mrefu unatosha.*
15. *Kiazi kibaya kina mfuko mzuri.*
16. *Wafalme wazuri wanatosha.*
17. *Kisiwa kirefu kina mito mibaya.*
18. *Vijiko vina mifuko mirefu.*

Solution 7.6b *wakuu.* 'Prince' belongs to Class 1 (human), so the plural is formed by replacing the singular prefix *m-* with the plural *wa-*.

Rules:

- Word order: SVO, Noun – Adj.
- Nouns are grouped into three classes:

  Class 1. Human nouns: 'child', 'king', 'man';

  Class 2. Non-human nouns: 'umbrella', 'river', 'bag';

  Class 3. Non-human nouns: 'spoon', 'potato', 'island';

- The class and number are marked by a prefix on the noun. The adjective agrees with the noun in class and number, while the verb is conjugated according to the class and number of the subject, as follows:

	Class 1		Class 2		Class 3	
	SG	PL	SG	PL	SG	PL
Noun/Adj.	*m-*	*wa-*	*m-*	*mi-*	*ki-*	*vi-*
Verb	*a-*	*wa-*	*u-*	*i-*	*ki-*	*vi-*

Note

The identification of noun classes which are marked by a prefix is specific to the Bantoid languages. Depending on the language, the nouns can be classified based on certain semantic considerations (similar to the way in which classifiers work – see Chapter 5), but not necessarily. For example, in this problem, there is no semantic reason to discriminate between Classes 2 and 3. Moreover, we need not find a discriminator, since there are no new words whose class we need to determine. The only distinction we need to make is that Class 1 only includes human nouns, in order to be able to differentiate between Class 1 and Class 2, which use the same singular marker.

Solution for practice problem 7.7. Arabic

Solution 7.7a Sentence 5. 'The monster beat [=did beat] the dog which he scared.'

Solution 7.7b

6. 'I created him.'
7. 'You$_{SG}$ brought the teacher who scared you$_{SG}$.'
8. 'The servant surprised the dog which the teacher beat [=did beat].'
9. 'I burnt the monster which beat [=did beat] the servant.'

Solution 7.7c

10. *'afzazta lxādima llaðī 'aṣamma lḥayawāna*
11. *'axraǧa lkalbu lmudarrisa llaðī 'aðallaka*
12. *'aḥraqtu lqāmūsa llaðī 'abda'tahu*

Rules:

- Word order: VSO; the relative clause is introduced by *llaðī* ('which'/'who'/'whom') and the word order inside it is identical (VSO).
- Noun: receives the suffixes *-u* (subject) or *-a* (object).

- Verb:
  - Verb stem is represented by three consonants (C_1-C_2-C_3), while the conjugation is done through transfixes (specific to Semitic languages). For example: 'to create' = *b—d—'*, 'to scare' = *f—z—z*, etc.
  - Subject is marked by the following transfixes:
    * 1SG: *'a*—C_1C_2—*a*—C_3—*tu*
    * 2SG: *'a*—C_1C_2—*a*—C_3—*ta*
    * 3SG: *'a*—C_1C_2—*a*—C_3—*a* (if $C_2 \neq C_3$)
      'a—C_1—*a*—C_2C_3—*a* (if $C_2 = C_3$)
  - Object is marked as a suffix to the verb: 2SG = *-ka*, 3SG = *-hu* only if it is not already expressed by noun.

Solution for practice problem 7.8. Welsh

Solution 7.8a

8. 'His brother has a beautiful queen.'
9. 'His witch has a wise princess.'
10. 'My princess has a good friend.'

Solution 7.8b

11. *Mae tywysoges benffol gan ei dad e.*
12. *Mae tad call gan ei dywysoges e.*
13. *Mae dewynes dalentog gan 'y mhlentyn i.*

Rules:

- Word order: *Mae* [O Adj.] *gan* S.
- The possessive surrounds the noun: 'his *X*' = *ei X e*, 'my *X*' = *'y X i*.
- Noun undergoes initial consonant mutation based on context:

No possessive	1SG poss.	3SG poss.
b	*m*	*f*
p	*mh*	*b*
d	*n*	*dd*
t	*nh*	*d*
g	*ng*	
c	*ngh*	*g*

Thus, we observe the following rules: for 1SG poss., voiced stops become nasals, preserving the place of articulation (*b* ⟶ *m*, *d* ⟶ *n*, *g* ⟶ *ng*), while voiceless stop become aspirated nasals with the same place of articulation (*p* ⟶ *mh*, *t* ⟶ *nh*, *c* ⟶ *ngh*).

For 3SG poss., we cannot deduce the transformation rule for voiced stops, but in the case of the voiceless ones, they become voiced (*p* ⟶ *b*, *t* ⟶ *d*, *c* ⟶ *g*).

The adjective undergoes an initial consonant mutation as well. In the masculine it will have a voiceless stop as the initial consonant, while if it is feminine, it will be voiced (e.g., 'beautiful': *teg* + 'brother', *deg* + 'princess').

Solution for practice problem 7.9. Tadaksahak

Solution 7.9a

17. 'The water was not drunk.'
18. 'I had the sheep watched.'
19. 'The salt will be swallowed.'
20. 'He is not taking the corpse.'

Solution 7.9b

21. *abzubuz aryen*
22. *aɣabʒiʒuwenket i*
23. *amanokal anenin aryen*
24. *cidi anetegmi*
25. *atesefred cidi*

Rules:

- Word order: SVO. If the subject is a pronoun, it is omitted;
- Verb: S–T–V–R;
  - S = Subject: *a-* = 3SG, *i-* = 3PL, *aɣa-* = 1SG;
  - T = Tense (combined with negation):

	Past	Present	Future
Affirmative	∅	*-b-*	*-te-*
Negative	*-ne-*	*-se-*	

  - V = Voice:
    * ∅ = active;
    * *-t-* = passive;
    * *-š-* / *-z-* / *-s-* / *-ʒ-* = causative ('to have someone do...'). If the stem contains any of these four sounds, the same sound is used here. Otherwise, *-s-* is used.
  - R = stem; the stem has two suppletive forms: one for active and another one for passive and causative.

Solution for practice problem 7.10. Sandawe

Solution 7.10a

15. 'Apparently, a cook_{M} skinned a rooster.'
16. 'A $\text{blacksmith}_{\text{F}}$ brought a cook_{F}.'
17. 'Apparently, bulls didn't hit a $\text{leopard}_{\text{F}}$.'
18. 'A cow herded hunters.'

Solution 7.10b

19. *thíméỳsò kókóʔà bálóʔǫ́wáá*
20. *ŋǃàméỳsù théká ǁʼěʔḛ́wáátshúgé*
21. *théká kòŋkórì bálóóyétshú*
22. *kʼámbà thíméỳsò xééʔíntshèégé*
23. *ǃʼìnéỳ ŋǃàméỳsògéà xééʔíŋ*

Rules:

- Human nouns receive the suffixes: ∅ (masc. SG), *-sù* (fem. SG), *-sò* (PL);
- Sentence structure:
  - Affirmative: S + O—(*gé*)—X_S + V—X_O;
  - Negative: S + O + V—X_O—Y_S—(*gé*).
- *-gé-* marks 'Apparently' (non-witnessed evidential);
- X_S and Y_S agree with the subject, while X_O agrees with the object, as follows:

	X_S	Y_S	X_O
SG masc.	*-à*	*-tshèé*	*-yé*
SG fem.	*-sà*	*-tshú*	*-sú*
PL human	*-ʔà*	*-tshó*	*-ʔín*[a]
PL non-human	*-ʔà*	*-tshó*	*-ʔwáá*[b]

[a] *-ʔín* ⟶ *-ʔíŋ* / _ # (alternatively, in affirmative sentences).
[b] *V́V́* + *-ʔwáá* ⟶ *V́ʔV̋wáá* and *V̀V́* + *-ʔwáá* ⟶ *V̌ʔV̋wáá*.

The morpheme *-ʔwáá* attracts tone change if V_2 has a high tone. In this case, V_2 will get devoiced and, if V_1 has low tone, it will become rising.

Solution for practice problem 7.11. Burushaski

Solution 7.11a

13. 'Those rams will fall.'
14. 'The gods will deceive the king.'
15. 'The spider will see this woman.'
16. 'Which woman will catch the frogs?'

Solution 7.11b

17. *ue biṭayo uɣurcan.*
18. *gušiŋance amic ɣaṣepišo uɣarkan?*
19. *thamišue guce hoolalašo uyeecan.*
20. *amin šugulue hilešo usarki?*
21. *ine hilese šuguli muyarani.*
22. *ɣeniṣe ine dasin musarko.*
23. *khine dasine šugulomuc uyeeco.*
24. *ṣiqare ise ɣurqun iyarani.*
25. *amis ̣jakun iwari?*

Rules:

- Word order: SOV, Modifier – Noun

- Modifiers:

	Human		Non-human	
	SG	PL	SG	PL
‘this’	*khine*	*khue*	*guse*	*guce*
‘that’	*ine*	*ue*	*ise*	*ice*
‘which’	*amin*		*amis*	*amic*

- Noun plural:
  1. Masculine (human) and non-human nouns:
     - if singular ends in *n*: *-n* ⟶ *-yo*;
     - if singular ends in *s*: *-s* ⟶ *-šo*;
     - if singular ends in another consonant: *C* ⟶ *-Cišo*;
     - if singular ends in a vowel: *-V* ⟶ *-Vmuc*;
  2. Feminine (human):
     - if singular ends in a vowel: *-V* ⟶ *-Vmuc*;
     - if singular ends in a consonant – nouns behave irregularly (they receive the suffix *-anc*, but some consonant alterations may occur). Nevertheless, the problem does not require us to infer any plural form from this category;
- Ergative marker: *-e* added after the plural marker; *o* ⟶ *u* / _ *e*;
- Verb: receives a prefix and a suffix. The prefix agrees with the absolutive argument of the verb, while the suffix agrees with the nominative argument of the verb, as follows:

	non-human / masc. human	fem. human	plural
prefix	*i-*	*mu-*	*u-*
suffix	*-i*	*-o*	*-an*

Further reading

Carnie, Andrew. 2011. *Modern syntax: A coursebook*. Cambridge: Cambridge University Press.

Cinque, Guglielmo & Richard S. Kayne (eds.). 2008. *The Oxford handbook of comparative syntax*. Oxford: Oxford University Press.

Coon, Jessica. 2013. *Aspects of split ergativity*. Oxford: Oxford University Press.

Coon, Jessica, Diane Massam & Lisa Demena Travis (eds.). 2017. *The Oxford handbook of ergativity*. Oxford: Oxford University Press.

Dixon, Robert M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

8 Semantics

8.1 Introduction

Semantics is the subfield of linguistics concerned with the study of meaning. Thus, semantics problems are not based on discovering the way certain words change their form (phonetics or phonology), get inflected or derived (morphology) or on the way in which words combine into sentences (syntax). This type of problem is strictly focused on the meaning of words and on the way two (or multiple) words can be combined to form a new word with a different meaning (e.g., in English we have words like *rainbow* which comes from *rain* + *bow*, thus the bow/arc/bent shape (in the sky) caused by rain). So, in this case, the focus is not on the combination process (morphologically, $X + Y \rightarrow XY$), but rather on the meaning that it has. Generally, semantics problems are chaos-and-order problems (the corpus is given in random order) and the corpus consists of words (or combinations of two to three words) which do not necessarily share any morphological feature, but rather a semantic one (belong to the same semantic field).

We need to remember that words that designate organs or body parts (*liver*, *heart*, *eye*, etc.) are the most "dangerous" ones, in the sense that their combinations often transcend their semantic field, taking on not entirely expected meanings (one of their most common uses is to express emotions or feelings, which can be connected to certain organs). For example, in Cameroon pidgin,[1] the word 'generous' is translated as *open han* ('open hand'), 'wickedness' = *blak hat* ('black heart'), 'hatred' = *bat hat* ('bad heart'), 'dizziness' = *blak ai* ('black eye'), 'poverty' = *drai han* ('dry hand'), and so on.[2] Therefore, we should pay extra attention when we encounter body parts or organs together with emotions or feelings in a semantics problem.

Furthermore, this type of problem requires a certain intuition to solve, since there is no absolute approach or solving method. Below we will present one solving method which can be used (the graph method).

[1] A pidgin is a mix of languages, a simplified way of communication, developed between two or more groups of people who do not share a common language. This way of communication is not spoken as a primary or native language. In this case, Cameroon pidgin is based on the English language.

[2] From a problem by Aleka Blackwell (NACLO 2014).

8.2 Graph method

A graph is a combination of points (nodes) connected by lines. In this method, the base words will be represented by the nodes, while their combinations will be represented by the lines. So, if we want to represent the structures *blak ai* and *blak hat* from above, the graph would be something like the one presented in Figure 8.1.

Figure 8.1: Sample graph for the structures *blak ai* and *blak hat*.

From this graph, we understand the following:

- the word *blak* combines with *ai* to form *blak ai*. Moreover, the direction of the arrow (from *blak* to *ai*) shows the order in which the words combine (so the resulting phrase is *blak ai* and not *ai blak*);
- similarly, *blak* combines with *hat* to form *blak hat*.

One thing we need to take into account is which of these words are actually given in the corpus. Based on our previous example, the dataset we analysed contain only the phrases *blak ai* and *blak hat*, so we need to distinguish the words that are given from those that are not (in this case, we marked them in bold and non-italic). In handwriting, it is easier if we underline or circle them.

In order to make things simpler, another thing we can do is not waste time writing the word combinations on top of the arrows, since the direction of the arrow already shows us the combination order (we can mark the fact that that word appears in our corpus through a horizontal line – as if we underline the phrase which we did not write anymore). Thus, the graph above (Figure 8.1) would become as shown in Figure 8.2.

Figure 8.2: Simpler representation of the graph shown in Figure 8.1.

Here, the line on top of the arrow shows us that the two compound words are given in the corpus.

In order to solve a linguistics problem using this method, we need to follow the next steps:

Step 1. Create a graph for all the words in the target language. For this step, it is preferred that we do not look at the English translations so we can focus solely on the word structure and not on the possible combinations of meaning.

Step 2. Create a partial graph for the English words. By partial, we mean that we do not necessarily need to include all the words. Of course, if we can include them all, it is even better, but sometimes we might not be completely certain how some words are connected with one another. Thus, it suffices that we construct a partial graph (it is important that this graph contains only combinations we can be sure of, not "likely" ones).

Step 3. Compare the partial graph of English words with the total graph of the words in the given language and see where they would match.

Step 4. Translate what we can and, knowing the shape of the graph, continue filling in the rest of the words.

Problem 8.1

Lango (Ksenia Gilyarova, IOL 2005)

Here are some words and phrases in Lango and their English translations *in random order*:

dyè ɔ̀t, *dyè tyèn*, *gìn*, *gìn wìc*, *ɲíg*, *ɲíg wàŋ*, *ɔ̀t cɛ̀m*, *wìc ɔ̀t*
'eyeball', 'grain', 'roof', 'garment', 'floor', 'restaurant', 'sole of foot', 'hat'

Problem 8.1a Determine the correct correspondences.

Problem 8.1b Translate into English: *cɛ̀m* and *dyè*.

Problem 8.1c Translate into Lango: 'window'.

Solution

Step 1. Create the graph (see Figure 8.3). We can notice that *ɔ̀t* appears three times, so we can start with it.

We continue to connect each word and obtain the final graphs in Figure 8.4. Note that for this problem there are two independent subgraphs.

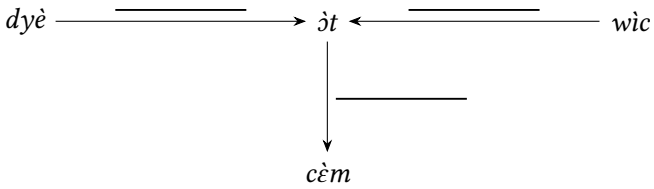


Figure 8.3: Graph corresponding to Step 1.

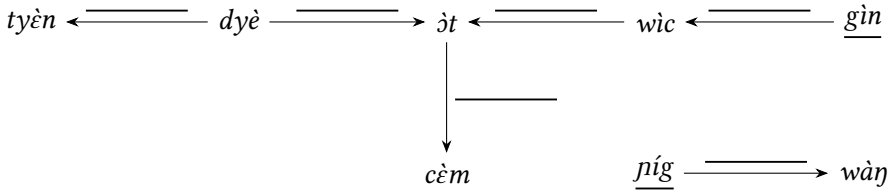


Figure 8.4: Complete graph of the words and phrases in Lango.

Notice that we have two separate subgraphs: the first one, which has a T-shape, and the second one which only has two nodes. Moreover, notice that the words *gìn* and *ɲíg* are underlined, meaning they also appear as single words in the corpus.

Step 2. It is time to try and form a partial graph using the English words. At first sight, we can certainly correlate the word pairs: 'roof' – 'floor' (top part and bottom part of the room/house), as well as 'hat' – 'roof' (the covering of the head and of the room/house). Since the word 'garment' is already given, we can consider 'hat' to be the 'head garment/the garment of the top part'. Thus, we can create the partial graph shown in Figure 8.5.

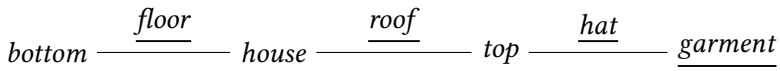


Figure 8.5: Partial graph of the English words.

We need to notice, in this case, that the connection between the nodes was not done using arrows, but lines, since the order of the constituents is not relevant.

Comparing the two graphs (Figures 8.4 and 8.5), notice that the partial graph contains a chain of four words (‘garment’ – ‘top’ – ‘house’ – ‘bottom’), and one of the end-nodes is underlined (meaning it is given in the corpus). Looking at the complete graph of Lango words (Figure 8.4), we know for sure that the partial graph of English words (Figure 8.5) cannot be part of the small subgraph, since it only has two nodes. In the big subgraph (the T-shaped one), only one node is underlined, namely *gìn*. Thus, we deduce that *gìn* = ‘garment’, and the four nodes (‘garment’ – ‘top’ – ‘house’ – ‘bottom’) can either be *gìn* – *wìc* – *ɔ̀t* – *dyè*, or *gìn* – *wìc* – *ɔ̀t* – *cèm*. Either way, the first three words are identical, so we deduce that *gìn* = ‘garment’, *gìn wìc* = ‘hat’, *wìc* = ‘top’, *wìc ɔ̀t* = ‘roof’, *ɔ̀t* = ‘house’. Following this, the proposed graph becomes the one shown in Figure 8.6.

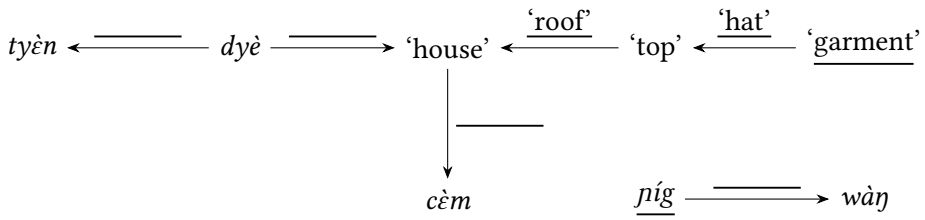


Figure 8.6: Partially solved graph.

Moreover, we know that one of the words *cèm* and *dyè* means ‘bottom’.

The remaining English words are: ‘floor’ (which we know represents ‘house + bottom’), ‘grain’, ‘eyeball’, ‘restaurant’, and ‘sole of foot’. We can already assume that ‘sole of foot’ is connected to ‘bottom’ (the bottom of the foot/leg/body). Thus, if *cèm* is ‘bottom’, we would not be able to connect it to ‘foot’, therefore *dyè* = ‘bottom’ and *tyèn* = ‘foot’. We can now modify the graph (see Figure 8.7).

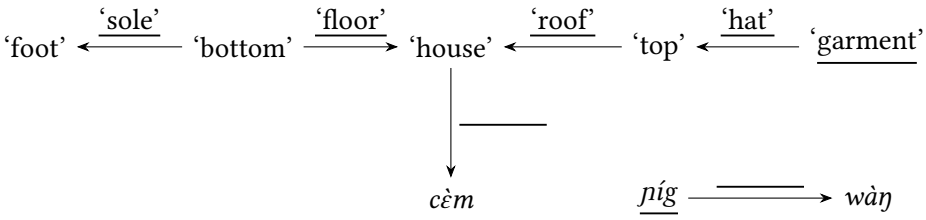


Figure 8.7: Partially solved graph, including ‘bottom’ and ‘foot’.

We are left with three words: ‘restaurant’, ‘eyeball’, and ‘grain’. Moreover, we know that one of them needs to be connected to ‘house’ (‘house’ + *cɛ̀m*), while the other two must be derived from one another (*ɲíg* and *ɲíg wàŋ*). Among the English words, the one which is most closely related semantically with ‘house’ is ‘restaurant’ (‘restaurant’ = ‘house’ + ‘food’), while ‘eyeball’ can be derived from ‘grain’ as in ‘eyeball’ = ‘grain’ + ‘eye’ (the grain of the eye).

Thus, we can make all the correspondences:

Solution

Problem 8.1a	*dyè ɔ̀t*	=	‘floor’	(bottom + house)
	dyè tyɛ̀n	=	‘sole of foot’	(bottom + foot)
	gìn	=	‘garment’	
	gìn wìc	=	‘hat’	(garment + top)
	ɲíg	=	‘grain’	
	ɲíg wàŋ	=	‘eyeball’	(grain + eye)
	ɔ̀t cɛ̀m	=	‘restaurant’	(house + food)
	wìc ɔ̀t	=	‘roof’	(top + house)

Problem 8.1b	*cɛ̀m*	=	‘food’
	dyè	=	‘bottom’

For task (c), we need to use the words we already have. Thus, we deduce that ‘window’ = eye of the house = ‘eye’ + ‘house’ (we deduce the word order from the phrase ‘eyeball’ = grain of the eye = ‘grain’ + ‘eye’, and not *‘eye’ + ‘grain’). Thus, ‘window’ = *wàŋ ɔ̀t*.

This is, however, an easy problem for which a graph is not necessarily needed, since one can observe that the only word which occurs in three different phrases is *ɔ̀t*, and the only three English translations which have something in common are ‘floor’, ‘roof’, and ‘restaurant’ (they are connected to a house/building). Nevertheless, the problem above offers an easy-to-understand example for the way in which graphs can be used to solve this type of problem.

We can now try to apply this method to solving a more complex problem.

Problem 8.2

Guaraní (Artur Corrêa Souza, RoLO 2021)

Here are some words and phrases in Guaraní and their English translations *in random order*:

1.	*jaxy*	A.	‘water’
2.	*jaxy-tata*	B.	‘brave’
3.	*jaxy endy*	C.	‘thumb’
4.	*kuã guaxu*	D.	‘liver, heart’
5.	*kuã regua*	E.	‘fire’
6.	*py’a*	F.	‘smoke’
7.	*py’a guaxu*	G.	‘pregnant’
8.	*tata*	H.	‘ring (jewellery)’
9.	*tata endy*	I.	‘moonlight’
10.	*tata rataxĩ*	J.	‘firelight’
11.	*ye guaxu*	K.	‘moon’
12.	*yvy rataxĩ*	L.	‘good soil’
13.	*yvy porã*	M.	‘dust’
14.	*yy*	N.	‘star’

Problem 8.2a Determine the correct correspondences.

Problem 8.2b Translate into English:

15.	*guaxu*	17.	*rataxĩ*	19.	*ye*
16.	*porã*	18.	*regua*		

Problem 8.2c Translate into Guaraní:

20.	‘calm, relaxed’	21.	‘fog’

Solution

First notice that in English we have words referring to organs (‘liver, heart’) and emotions (‘brave’ and, in task (c), ‘calm’). So we can expect that these two are

connected. Nevertheless, the first step is constructing the graphs for the Guaraní words (see Figure 8.8).

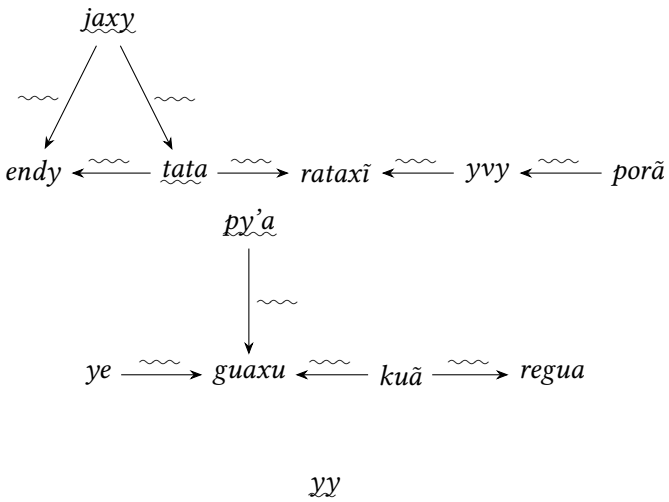


Figure 8.8: Complete graph of the words and phrases in Guaraní.

Note

In Appendix C I present a hand-drawn graph in order to show what such a graph might look like in reality, when solving a problem.

We notice that, in Guaraní, we have three independent subgraphs. For the partial graph of English words, we have the words: 'moon', 'fire', 'moonlight', and 'firelight'. These can be arranged in a graph as shown in Figure 8.9.

'fire' ~~~ 'light' ~~~ 'moon'

Figure 8.9: Partial graph of the English words 'moon', 'fire', 'moonlight', and 'firelight'.

This is an ideal partial graph since we have two base words (nodes), 'fire', and 'moon', which are found in the corpus and are both connected to the same word

('light'). According to the Guaraní graph in Figure 8.8, the only two nodes close to one another that are found in the corpus (are underlined) are *jaxy* and *tata*, and both of them are connected to the word *endy*. Thus, we deduce that *endy* = 'light', and {*jaxy*, *tata*} = {'fire', 'moon'}[3]. In order to determine which is which, we notice that *jaxy* is not connected to anything else, while *tata* is further connected to *rataxĩ*. In English, we have the word 'smoke' which is clearly connected to 'fire', so *tata* = 'fire' and *jaxy* = 'moon'. Moreover, from the graph, we notice that *jaxy* and *tata* combine with one another, thus, in English, we need to find a word formed by combining the words 'moon' and 'fire'. The only one which is semantically close to that is 'star' (= 'fire moon'). Adding this information, our graph will look like that in Figure 8.10.

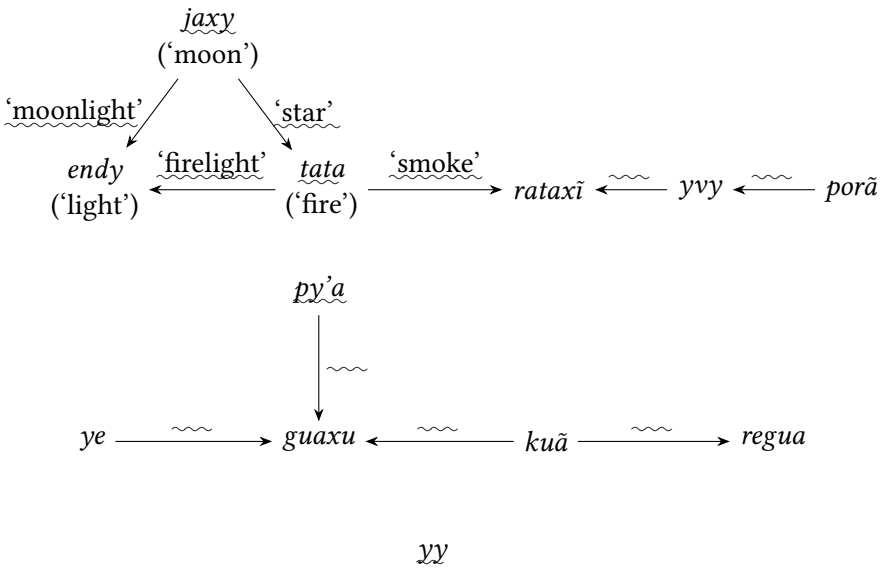


Figure 8.10: Partially solved graph.

As mentioned above, 'fire' is only combined with one more word and, in English, the only word that belongs to the same semantic field is 'smoke'. Thus, *tata rataxĩ* = 'smoke'. Nevertheless, we cannot immediately deduce the meaning of *rataxĩ* ('smoke' = 'fire' + *X*). However, looking at the English words, we notice the word 'dust' and, since this roughly relates to the same semantic area as 'smoke', they most likely have something in common (the word *X*). Thus, we get:

[3]By this notation, we mean that *jaxy* and *tata* correspond to 'fire' and 'moon', but we do not know which is which.

‘smoke’ = ‘fire’ + X and ‘dust’ = ? + X

Comparing again the remaining words, we notice we have the phrase ‘good soil’ (and, indeed, ‘smoke’ is to ‘fire’ as ‘dust’ is to ‘soil’). Therefore, X represents ‘particles/powder’ (smoke is a “powder” from the fire, while dust is a powder of soil). Thus, we can complete the top subgraph as shown in Figure 8.11.

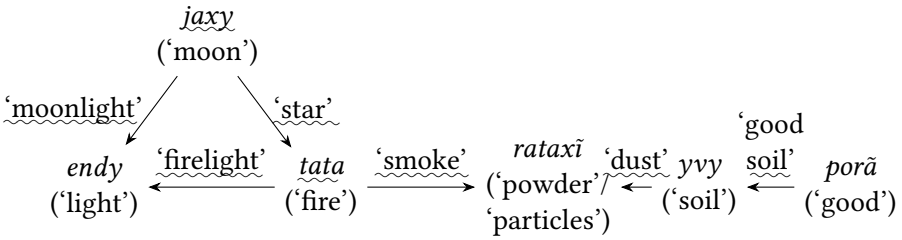


Figure 8.11: Completely solved top subgraph.

Since this subgraph is independent (not connected to the others in any way), we have reached a dead end and we need to build a new partial graph based on the English words. However, we have already made eight correspondences. The remaining words are:

4.	*kuã guaxu*	A.	‘water’
5.	*kuã regua*	B.	‘brave’
6.	*py’a*	C.	‘thumb’
7.	*py’a guaxu*	D.	‘liver, heart’
11.	*ye guaxu*	G.	‘pregnant’
14.	*yy*	H.	‘ring’

Among these, we can notice ‘ring’ and ‘thumb’ (both being related to the word ‘finger’, as in ‘thumb’ = ‘finger’ + ‘big’ and ‘ring’ = ‘finger’ + ‘jewellery/ ornament’). Only based on these two words, we can build the graph shown in Figure 8.12.

Figure 8.12: Partial graph.

Thus, ‘finger’ can correspond to either *kuã*, or *guaxu*. If it corresponded to *guaxu*, it needs to combine with another word among those given (i.e., ‘finger’

+ ‘water’, ‘finger’ + ‘brave’, ‘finger’ + ‘liver, heart’, ‘finger’ + ‘pregnant’), all of these combinations being highly unlikely (difficult to justify). Therefore, *kuã* = ‘finger’. Moreover, we notice that no other word seems to belong in the semantic field of the word ‘jewellery, ornament’, so, most likely this is the meaning of *regua*. The graph now looks like that shown in Figure 8.13.

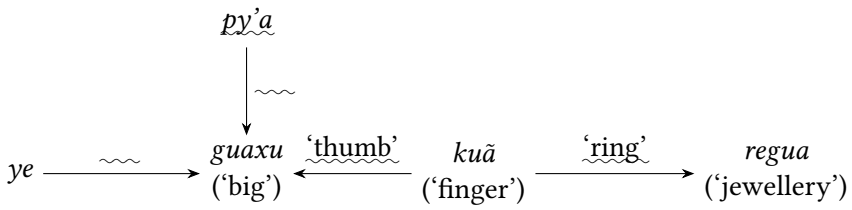


Figure 8.13: Completely solved subgraph.

We need to have in the corpus two more words which are formed from the combination of ‘big’ with other words (and one of the words it combines with – *py’a* – must also appear in the corpus). The first word we notice is ‘pregnant’ (we can consider that it is formed as ‘belly’ + ‘big’ or something similar). Moreover, we notice that we have ‘liver, heart’ and ‘brave’ among the remaining words. As mentioned above, words for emotions are often formed from words designating organs, so ‘brave’ = ‘big’ + ‘heart, liver’ is quite plausible. In this way, we also completed this subgraph and the only remaining word, *yy*, must mean ‘water’.

So we have the correspondences:

1.	*jaxy*	=	‘moon’	
2.	*jaxy-tata*	=	‘star’	(moon + fire)
3.	*jaxy endy*	=	‘moonlight’	
4.	*kuã guaxu*	=	‘thumb’	(finger + big)
5.	*kuã regua*	=	‘ring’	(finger + jewellery)
6.	*py’a*	=	‘liver, heart’	
7.	*py’a guaxu*	=	‘brave’	(liver, heart + big)
8.	*tata*	=	‘fire’	
9.	*tata endy*	=	‘firelight’	
10.	*tata rataxĩ*	=	‘smoke’	(fire + powder)
11.	*ye guaxu*	=	‘pregnant’	(belly + big)
12.	*yvy rataxĩ*	=	‘dust’	(soil + powder)
13.	*yvy porã*	=	‘good soil’	(soil + good)
14.	*yy*	=	‘water’	

In task (b), we are only asked to translate simple words (which correspond to the nodes of the graph), so this task is straightforward now: *guaxu* = 'big', *porã* = 'good', *rataxĩ* = 'powder', *regua* = 'ornament/jewellery', *ye* = 'belly'.

Remember

For this type of problem, there might be multiple acceptable answers, which is taken into account when grading. Thus, for the word *regua* (which must be deduced from the combination 'ring' = 'finger' + *regua*), there can be multiple interpretations: 'ornament', 'jewellery' etc., but it can also be considered to mean 'circle', 'surrounding', etc. (which is the actual meaning of the Guaraní word). Thus, all of these words would be equally acceptable.

In task (c), we are asked to translate the words 'fog' and 'calm'. The word 'fog' is easy to translate since it resembles 'dust' and 'smoke' (thus, 'fog' = 'water' + 'powder'), so its translation is *yy rataxĩ*. The word 'calm' can be compared with the word 'brave' (both referring to human qualities). Since 'brave' was formed from 'liver, heart', it is likely that 'calm' is too. Moreover, we notice that we also have another adjective: 'good'. Therefore, we can form the combination 'calm' = 'good' + 'liver, heart'. Thus, its translation is *py'a porã*.

Put all together, the answers are:

Problem 8.2a

1. K. 3. I. 5. H. 7. B. 9. J. 11. G. 13. L.
2. N. 4. C. 6. D. 8. E. 10. F. 12. M. 14. A.

Problem 8.2b

15. 'big'
16. 'good'
17. 'powder, particles'
18. 'circle'
19. 'belly, stomach'

Problem 8.2c

20. *py'a porã*
21. *yy rataxĩ*

8.3 Practice problems

Problem 8.3

Basque (Natalia Zaika, MSK 2012)

Here are some words in Basque and their English translations *in random order*:

> *igogailu*, *artzain*, *lantegi*, *lantalde*, *bizitegi*, *taldekide*, *erizain*, *garbigailu*, *ikastalde*, *bizikide*, *garbitegi*, *ikaskide*, *lankide*, *eritegi*, *artalde*
>
> 'classmate', 'flatmate', 'flock of sheep', 'crew', 'elevator', 'clinic', 'factory', 'nurse', 'home', 'shepherd', 'wash-house', 'colleague', 'washing machine', 'team member', 'class (of students)'

Problem 8.3a Determine the correct correspondences.

Problem 8.3b How is the word *artalde* different from the other Basque words?

Problem 8.3c Translate the word 'sheep-pen' into Basque, knowing that it has the same feature as the word *artalde*.

Problem 8.4

Turkish (Monojit Choudhury, PLO 2014)

Here are some words in Turkish and their English translations *in random order*:

> *gözlemci*, *döndürmek*, *gündöndü*, *gözlükcü*, *şarkıcı*, *çocukluk*, *gözlemek*, *pazar*, *pazartesi*, *cumartesi*, *güneşli*
>
> 'Saturday', 'Sunday', 'Monday', 'observer', 'singer', 'to observe', 'to rotate', 'sunny', 'sunflower', 'optician', 'childhood'

Problem 8.4a Determine the correct correspondences.

Problem 8.4b Translate into Turkish:

1. 'observation'
2. 'child'
3. 'the state of being a singer'
4. 'spectacles'
5. 'Friday'

Problem 8.5

Chinese (Roxana Dincă, RoLO 2015)

Here are some words and phrases in Chinese and their English translations *in random order*:

> *hóng*, *míngbái*, *báishì*, *huáng*, *hēishì*, *shìqing huángle*, *yǎn*, *yǎnhóng*, *hóngshì*, *hóngyán*, *hēihuà*, *báiyǎn*, *báihuà*, *hēibái fēnmíng*, *yán*, *hēibái*, *huángle*
>
> 'encoding', 'black-and-white', 'face', 'yellow', 'funeral', 'bankruptcy', 'to clarify', 'to dislike', 'young woman', 'failure', 'decoding', 'wedding', 'eye', 'black market', 'it's written in black and white', 'jealousy', 'red'

Problem 8.5a Determine the correct correspondences.

Problem 8.5b The word *báishì* can have two meanings in Chinese, although only one is reflected in the correspondences above. What is the other meaning?

Problem 8.6

Hausa (Paul Helmer, RoLO 2019)

Here are some phrases in Hausa and their English translations *in random order*:

1.	*bàakín rúwáa*	A.	'talkative boy'
2.	*bàbbán bàakín tsúntsúu*	B.	'white camel'
3.	*bàbbán yátsàa*	C.	'big beak'
4.	*bàbbán ràkúmín rúwáa*	D.	'thumb'
5.	*bàbbán yár sháanúu*	E.	'fruit'
6.	*bákín cíkìi*	F.	'fork'
7.	*bákín ràagóo*	G.	'pregnant woman'
8.	*cóokàlíi mài yátsàa*	H.	'sorrow'
9.	*dán ràagóo*	I.	'estuary'
10.	*dán mài bàbbán bàakíi*	J.	'lamb'
11.	*fárin ràkúmíi*	K.	'black sheep'
12.	*rúwán bíshíyàa*	L.	'(tree) sap'
13.	*yár mài bàbbán cíkìi*	M.	'tsunami'
14.	*yár bíshíyàa*	N.	'big heifer'

Problem 8.6a Determine the correct correspondences.

Problem 8.6b Translate into English:

15. *dán sháanúu*
16. *fárín cíkìi*
17. *yár mài bákin bàakíi*

Problem 8.6c Translate into Hausa:

18. ‘girl who has a spoon’
19. ‘crow’
20. ‘river’
21. ‘(tree) branch’

An ‘estuary’ is a wide part of the river, similar to a funnel. A ‘heifer’ is a young female cow.

Problem 8.7

Tetum (Aleksejs Peguševs, RoLO 2020)

Here are some words and phrases in Tetum and their English translations *in random order*:

1.	*ai boot*	A.	‘eyelid’
2.	*ai fuan boot*	B.	‘scapula’
3.	*ai fuan musan*	C.	‘leaf’
4.	*ai tahan*	D.	‘big fruit’
5.	*ibun kulit boot*	E.	‘big tree’
6.	*kbas*	F.	‘auricle’
7.	*kbas tahan*	G.	‘skin’
8.	*kulit*	H.	‘seed’
9.	*matan kulit*	I.	‘shoulder’
10.	*matan musan*	J.	‘eyeball’
11.	*tilun tahan*	K.	‘big lip’

Problem 8.7a Determine the correct correspondences.

Problem 8.7b Translate into English:

12. *matan fuan* 13. *ai fuan kulit* 14. *tilun boot*

One of the phrases has the same translation as one of the phrases 1–11.

Problem 8.7c Translate into Tetum:

15. 'mouth'
16. 'big eye'
17. '(tree) bark'
18. 'grain'

Problem 8.7d Two of the words above can be combined to construct a phrase meaning 'impolite person'. Which ones are these?

The 'scapula' (or shoulder blade) is the large flat bone that is part of the shoulder joint. The 'auricle' is the visible part of the ear.

Problem 8.8

Malagasy (Alexey Kretov, MSK 2011)

Here are some words and phrases in Malagasy and their English translations *in random order*:

1.	*mahandohalika*	A.	'grandson'
2.	*lohalika*	B.	'ankle'
3.	*zafin-kitrokely*	C.	'shoot of rice (departing from the stem)'
4.	*hafaladia*	D.	'up to the sole'
5.	*zafim-bary*	E.	'rice field'
6.	*kitrokely*	F.	'great-great-great-grandson'
7.	*zafim-paladia*	G.	'one who can get on his knees'
8.	*zafy*	H.	'great-great-great-great–grandson'
9.	*tanim-bary*	I.	'knee'
10.	*mahambozona*	J.	'one who can carry something on his neck'

Problem 8.8a Determine the correct correspondences.

Problem 8.8b Translate into English:

11. *tany* 12. *vozona* 13. *halohalika*

Problem 8.8c Translate into Malagasy:

14. 'great-great-grandson'
15. 'sole'
16. 'up to the ankle'

y = 'i' in 'pit'.

8.4 Solutions of practice problems

Solution for practice problem 8.3. Basque

Solution 8.3a

igogailu = 'elevator'
artzain = 'shepherd'
lantegi = 'factory'
lantalde = 'crew'
bizitegi = 'home'
taldekide = 'team member'
erizain = 'nurse'
garbigailu = 'washing machine'
ikastalde = 'class of (students)'
bizikide = 'flatmate'
garbitegi = 'wash-house'
ikaskide = 'classmate'
lankide = 'colleague'
eritegi = 'clinic'
artalde = 'flock of sheep'

Solution 8.3b *-t + t- ⟶ -t-* (if a morpheme which ends in *t* is joined to another that starts with *t*, one of the two *t*'s is dropped).

Solution 8.3c 'sheep-pen' = *artegi*

Rules: Each Basque word is composed of two morphemes (the first one shows the semantic field, while the other the category):

1st morpheme	2nd morpheme
igo- = 'to lift'	
lan- = 'to work'	*-zain* = 'worker'[a]
bizi- = 'to live'	*-tegi* = 'place'
art- = 'sheep'	*-talde* = 'collective'
eri- = 'sick'	*-kide* = 'member'
garbi- = 'to wash'	*-gailu* = 'machine'
ikas- = 'to learn'	

[a] In reality, a more accurate translation would be 'keeper'.

Solution for practice problem 8.4. Turkish

Solution 8.4a The morphemes have been separated by a hyphen (-) and their meaning is given, in order, between brackets:

göz-lem-ci	'observer'	(eye + abstract-noun + agent-marker)
göz-lük-cü	'optician'	(eye + state-of-being + agent-marker)
göz-le(m)-mek	'to observe'	(eye + abstract-noun + verb-marker)
döndür-mek	'to rotate'	(rotation + verb-marker)
gün(eş)-li	'sunny'	(sun + adjective-marker)
gün-döndü	'sunflower'	(sun + rotation)
şarkı-cı	'singer'	(song + agent-marker)
çocuk-luk	'childhood'	(child + state-of-being)
pazar	'Sunday'	
pazar-tesi	'Monday'	(Sunday + tomorrow)
cumar-tesi	'Saturday'	(Friday + tomorrow)

Note

Since Turkish is a language displaying vowel harmony, the form of the suffixes will differ depending on the word (*ci/cı/cü* or *lük/luk*).

Solution 8.4b

1. *gözlem*
2. *çocuk*
3. *şarikcilik*
4. *gözlük*
5. *cumar*[4]

Solution for practice problem 8.5. Chinese

Solution 8.5a

hóng = 'red'
míngbái = 'to clarify'
báishì = 'funeral'
huáng = 'yellow'
hēishì = 'black market'
shìqing huángle = 'bankruptcy'
yǎn = 'eye'
yǎnhóng = 'jealousy'
hóngshì = 'wedding'
hóngyán = 'young woman'
hēihuà = 'encoding'
báiyǎn = 'to dislike'
báihuà = 'decoding'
hēibái fēnmíng = 'it is written in black and white'
yán = 'face'
hēibái = 'black-and-white'
huángle = 'failure'

Solution 8.5b 'white market'

[4]In reality, it is *cuma*, but *cumar* is the answer as can be deduced from the given data.

Solution for practice problem 8.6. Hausa

Solution 8.6a

1.	I.	3.	D.	5.	N.	7.	K.	9.	J.	11.	B.	13.	G.
2.	C.	4.	M.	6.	H.	8.	F.	10.	A.	12.	L.	14.	E.

Solution 8.6b

15. 'calf' (boy + cow)

16. 'happiness' (white + stomach) – based on 'sorrow' = black + stomach

17. 'impolite/naughty girl' (girl + with + black + mouth)

Solution 8.6c

18. *yár mài cóokàlíi* (girl + with + spoon)

19. *bákín tsúntsúu* (bird + black)

20. *bàbbán rúwáa* (big + water)

21. *yátsàn bíshíyàa* (finger + tree)

Rules:

- The determiners come before the head noun;
- The words *yár* = 'female, woman' and *mài* = 'with' are invariable;
- All other words end in *-n* if they are not the head noun or they double the final vowel if they are the head noun. An alternative explanation is that they end in *-n*, unless they are phrase-final, in which case the *-n* is removed and the last vowel is doubled.

Solution for practice problem 8.7. Tetum

Solution 8.7a

1.	E.	3.	H.	5.	K.	7.	B.	9.	A.	11.	F.
2.	D.	4.	C.	6.	I.	8.	G.	10.	J.		

Solution 8.7b

12. 'eyeball' 13. '(fruit) peel' 14. 'big ear'

Solution 8.7c

15.	*ibun*	17.	*ai kulit*
16.	*matan boot*	18.	*musan*

Solution 8.7d *ibun boot* ('big mouth')

Solution for practice problem 8.8. Malagasy

Solution 8.8a

1. G.	3. F.	5. C.	7. H.	9. E.
2. I.	4. D.	6. B.	8. A.	10. J.

Solution 8.8b 11. 'field' 12. 'neck' 13. 'up to the knee'

Solution 8.8c 14. *zafin-dohalika* 15. *faladia* 16. *hakitrokely*

Explanations

When two words combine (A–B), there are some phonological changes occurring in both of them. Thus, the first word (the prefix) undergoes some alternations at the end. If it ends in *y*, it will become *iN* (where *N* is a nasal which assimilates to the place of articulation of the next consonant – *m* for *p/b* and *n* for *d/k*) – e.g., *zafi**n**-**k**itrokely* and *zafi**m**-**p**aladia*.

The following word (*B*) undergoes an initial consonant mutation, depending on whether the prefix ends in a vowel or a consonant, as follows:

First consonant of *B*	after vowel	after consonant
l	*l*	*d*
v		*b*
k		*k*
(?)	*f*	*p*

Since we are asked to find (?) (in task c), the only rule we can deduce is that there is no initial consonant mutation if the word before ends in a vowel.

Interestingly, the words for 'great-grandson', 'great-great-grandson', etc. are based on vertical position of body parts: the more distant the descendant, the lower the position of the body part used. The terms 'great-great-grandson', 'great-great-great-grandson', and 'great-great-great-great-grandson' are formed by compounding *zafy* = 'grandson' with 'knee', 'ankle', and 'sole', respectively.

Further reading

Koptjevskaja-Tamm, Maria & Henrik Liljegren. 2017. Semantic patterns from an areal perspective. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*. Cambridge: Cambridge University Press.

Kroeger, Paul. 2022. *Analyzing meaning: An introduction to semantics and pragmatics*. Berlin: Language Science Press.

ten Hacken, Pius (ed.). 2016. *The semantics of compounding*. Cambridge: Cambridge University Press.

Vanhove, Martine (ed.). 2008. *From polysemy to semantic change*. Amsterdam: John Benjamins Publishing Company.

9 Number systems

9.1 Introduction

In order to understand problems about number systems, we firstly need to understand the concept of base and the characteristics of such systems.

In Romanian, for example, the number 432 is written *patru sute treizeci și doi*. It can be segmented into morphemes as follows: *patru* ('four') *sute* ('hundreds') *trei* ('three')-*zeci* ('tens') *și* ('and') *doi* ('two'), and each morpheme has a well defined role. The morphemes *patru* ('four'), *trei* ('three'), *doi* ('two') represent the *digits*, the basic units, while the morphemes *sute* ('hundreds') and *zeci* ('tens') represent the *orders*, i.e., some greater base words. Finally, the morpheme *și* ('and') denotes addition. Since in this case the orders are 10 and 100 (10^1, 10^2), we are talking about a base-10 (or decimal) system.

Moreover, word order plays an important role, together with the way in which the multiplication and the addition are marked. Returning to the example *patru sute treizeci și doi*, we can deduce the following:

- the general order is from big to small (we first write the hundreds, then the tens and finally the units);
- the multiplier is placed before the order (thus, we write *patru sute* and not **sute patru*);
- the morpheme for 'tens' is written as part of the same word with its multiplier (*treizeci*, not **trei zeci*);
- the tens are separated from the units by the word *și*.

There is also a simpler method to show all of these rules, by writing the general structure of the number:

100X + 10Y + Z = X *sute* Y-*zeci și* Z

This rule combines all four rules from above, showing the sequence of the orders (hundreds, tens, units), the position of the multiplier (as a separate word before the hundreds and combined with the ending *zeci* for tens), as well as the morpheme *și* added between tens and units.

Although base 10 is the most common base and most languages use it, there are also other bases, among which the most common are base 6 and base 20. Less common, but occurring in a reasonable number of languages are bases 4, 5, 12, and 60, while bases 7, 9, 11, and 13 are extremely rare (at the moment, no known natural language uses them), so that, in linguistics problems, we can assume from the beginning that the system is highly unlikely to have the bases 7, 9, 11, 13.

A special category of bases is 22–28, which are found in languages whose number system is based on body parts, such as Oksapmin or Kaugel.

In order to better understand the concept of base, let us consider how the numbers 10, 25, and 100 would be written in a base-6 system. Since the base is 6, the orders are 6 (6^1), 36 (6^2), 216 (6^3), and so on. Therefore, the number 10 will be written as 6 + 4 (or 4 + 6, depending on the word order), number 25 will be written as $(4 \times 6) + 1$, while 100 will be written as $(2 \times 36) + (4 \times 6) + 4$.

So, in problems involving number systems, there are two important characteristics:

(1) word order (or direction): big-to-small or small-to-big or a mixture;

(2) the way in which addition and multiplication are marked (or whether they are not overtly marked).

In terms of (1), remember that it is not necessary that the word order is strictly ascending or descending, and there can also be exceptions. For example, in German, 432 is written *vierhundertzweiunddreißig* (*vier-hundert-zwei-und-drei-ßig* = 4-'hundreds'-2-'plus'-3-'tens'), so the order is hundreds-units-tens. In terms of (2), note that in French, for example, in the number 2,510 (*deux mille cinq cent dix* = 'two' - 'thousand' - 'five' - 'hundred' - 'ten'), both the addition and the multiplication are implicit and as such, if the digit is placed before the order, it gets multiplied, while addition occurs every time after the order (e.g., 2,510 can be written as $(2 \times 1000) + (5 \times 100) + 10$). Moreover, if the addition or multiplication is marked, this will be done using a rather pervasive morpheme, word, or group of words.

Problem 9.1

Quenya (Roxana Dincă, RoLO 2013)

Here are some Quenya numbers:

neldë	3	*enquë yucainen*	26
canta	4	*minë nelcainen*	31
lempë	5	*cancainen*	40
otso	7	*atta tolcainen*	82
tolto	8	*atta tolcainen tuxa*	182
nelcëa	13	*nertë nelcainen lemtuxa*	539
encëa	16		

Problem 9.1a Write in numerals:

a. *tolcëa*
b. *enquë cancainen*
c. *cancainen neltuxa*
d. *lempë tolcainen*
e. *tuxa*

Problem 9.1b Write in Quenya: 1, 70, 192, 385.

Solution

We notice that numbers 13 and 16 both end in the suffix *-cëa*, while numbers bigger than 26 have a different structure. Thus, comparing the words for 3 and 13, we can assume it is a base-10 language and that numbers $10 + X$ are written as *X-cëa*. We further notice that in order to form the number 13, only a part of the word for 3 is used (*nel*).

Comparing examples 82 and 182, they differ only by the word *tuxa* placed at the end; therefore, we can deduce that *tuxa* means 100 (which further confirms that the number system for this language is base 10). Moreover, looking at the number 539, we notice that the last word is *lemtuxa*, where *tuxa* = 100, and *lem* is the first part of the number 5. Thus, in this case as well, only a part of the stem of the unit is used. Furthermore, we notice that in the right column, the second word always ends in *cainen*. We can assume that this is the suffix which marks the tens ($10X$), whence we deduce that the word order in Quenya is units-tens-hundreds.

Only based on these observations, we can write the following rules:

$10 + X = X'$ *-cëa* $\qquad$ $100X + 10Y + Z = Z\ Y'$*-cainen* X'*-tuxa*

By X' and Y' we mean that a different, truncated, form of the word is used.

Based on these observations, we can create a table with the form of each digit in different contexts:

Digit	X	$10 + X$	$10X$	$100X$
1	*minë*			∅
2	*atta*		*yu-*	
3	*neldë*	*nel-*	*nel-*	
4	*canta*		*can-*	
5	*lempë*			*lem-*
6	*enquë*	*en-*		
7	*otso*			
8	*tolto*		*tol-*	
9	*nertë*			

The four columns represent the form in which the respective digit is used if it represents the units, if it appears with the suffix *-cëa* (meaning $10 + X$), if it appears with the suffix *-cainen* ($10X$), or if it appears with the suffix *-tuxa* ($100X$).

We notice that the same form of 3 appears both in the case of $10 + X$ and $10X$. Therefore, we deduce that both contexts use the same form. Moreover, we have no reason not to assume that the same form will also be used in the case of hundreds. In order to deduce how that form is constructed, we compare it with the full form in the first column (the units). We notice that the short form (which we designated by X' and Y') represents the first syllable of the unit. The only exception is the digit 2, where the form used for 20 is *yu-*. It appears that in Quenya 1 and 2 are irregular and have different forms in different contexts. This is not uncommon cross-linguistically.

Thus, we can solve the tasks and write the rules.

Rules: Digits are single words. In compound words, only the stem of the digit is used, which is represented by the first syllable (in the notation below, X' is the stem/first syllable of X). The digit 2 has a special form, *yu-*. Thus:

$10 + X = X'$*-cëa* $\qquad$ $100X + 10Y + Z = Z\ Y'$*-cainen* X'*-tuxa*

Problem 9.1a a. 18 b. 46 c. 340 d. 85 e. 100

Problem 9.1b 1 = *minë* 192 = *atta nercainen tuxa*
70 = *otcainen* 385 = *lempë tolcainen neltuxa*

In situations where the base is unknown, a simple method to get some additional information is to count how many morphemes there are. If in a particular language we count 11 digits, we expect the base to be, most likely, 10 or 12. Usually, this method is just an estimation, and the result should probably be taken with an error margin of ±2 because: (1) it is possible that we misidentified some of the digits, and (2) it is possible that the problem does not feature all the digits or even some digits might have different forms in different contexts. Moreover, it is extremely important that we count only the digits, not other morphemes (such as orders or addition/multiplication markers).

Problem 9.2

Embera Chami (Vlad A. Neacșu, PLO 2022)

Here are two equalities in Embera Chami:

(1) *umbea + huasoma kwimane = omme huasoma omme*
(2) *omme huasoma kwimane + huasoma abba = kwimane huasoma*

Problem 9.2a Write the equalities above with numerals.

Problem 9.2b Write in Embera Chami: 1, 5, 17, 23.

Solution

At first sight, it might seem a very difficult problem without an obvious starting point and with very little information given. Nevertheless, if we check the structure of the numbers, we notice that there are two types: single words or numbers like X *huasoma* Y. We can assume that the second type will represent bigger numbers, and that *huasoma* is the base, these numbers representing $X \times huasoma + Y$ (or $Y \times huasoma + X$). Moreover, we notice that the same word can represent both X and Y, therefore X and Y are the slots where the digits are placed. Based on these rules, we can try to count the number of digits that occur in the problem, and we notice that there are only four (*umbea*, *kwimane*, *omme*, *abba*). Moreover,

they have a constant form (there are no changes or added or deleted morphemes). We can then assume that the base is 5 and that *huasoma* = 5.

The next thing we need to do is figure out the order of the constituents, i.e., figure out whether X *huasoma* Y means $5X + Y$ or $5Y + X$. Looking at equality (1), we have two cases:

a. *huasoma* $X = 5X$ b. *huasoma* $X = 5 + X$

Assuming case a. is true, eq. (1) becomes:

$$umbea + 5 \times kwimane = omme + 5 \times omme$$

This seems unlikely, since we know that *umbea* is, most likely, a digit. Thus, if *huasoma kwimane* meant $5X$, then *umbea* + *huasoma kwimane* should be equal to *umbea huasoma kwimane* (i.e., *umbea* $+5\times$*kwimane*). Generally, the carryover[1] is a strong strategy to discover certain digits.

Thus, case b. must be correct, and now we know that *huasoma* $X = 5 + X$, so we can deduce that X *huasoma* $= 5X$ and, extrapolating, X *huasoma* $Y = 5X + Y$.

Moreover, from eq. (1), we see that the number resulting from the addition has the multiplier *omme*. Since this is the result of a sum between a unit and a number like base + X, the result can either also be base + X (if there is no carryover) or 2×base + X (if there is carryover). Since we know that there is carryover (the number on the right has a multiplier, since it has the structure $5X + Y$), we deduce that *omme* = 2.

If we denote the remaining three digits by U, K and A (corresponding to their first letter), we can rewrite the equalities as follows:

(1) $U + (5 + K) = 12$

(2) $(10 + K) + (5 + A) = 5K$

Rearranging equality (1), we get: $U + K = 7$.

Knowing that U and K are digits smaller than 5 (the base), U and K can only correspond to 3 and 4, not necessarily in this order. Thus, A can only be 1 since it is the only remaining digit. Therefore, *abba* = 1.

Replacing this in eq. (2) gives us: $10 + K + 6 = 5K \Leftrightarrow 4K = 16$. So $K = 4$ and, subsequently, $U = 3$.

Thus, the rules are:

[1] We use the term *carryover* for the following situation: when you add 17 and 19, you add the units first (7 and 9 to get 16) and "carry over" the 1 (from 16) to the tens.

- 1 = *abba*, 2 = *omme*, 3 = *umbea*, 4 = *kwimane*, 5 = *huasoma*;
- $5X + Y = X$ *huasoma* Y.

Solution 9.2a (1) $3 + 9 = 12$ (2) $14 + 6 = 20$

Solution 9.2b

$1 =$ *abba*

$5 =$ *huasoma*

$17 = 3 \times 5 + 2 =$ *umbea huasoma omme*

$23 = 4 \times 5 + 3 =$ *kwimane huasoma umbea*

Discussion (not part of the solution)

For this problem, notice that counting the digits and understanding the structure of the numbers is extremely important for number problems. Moreover, carry-overs offer valuable information for problems in which equalities are given, and they can usually be used to infer the digits 2 and 3.

Moreover, it is important to mention that, in order to be able to solve the problem, we need to use the fact that the problem is self-sufficient, otherwise we cannot deduce that the base is 5. Let us consider a base B and the following values for the digits: *omme* $= 2$, *kwimane* $= 4$, *abba* $= B - 4$, *umbea* $= B - 2$. The equalities become:

(1) $(B{-}2) + B + 4 = 2B + 2$

(2) $2B + 4 + (B + (B{-}4)) = 4B$

In this case, the two equalities hold, no matter the base (e.g., for base 13 we would get the equalities: (1) $11 + 17 = 28$ and (2) $30 + 22 = 52$). Nevertheless, task (b) asks for the translation of 1, so *abba* needs to be 1 (we know that *umbea* is bigger than *abba*, since $B - 4 < B - 2$), and if *abba* $= 1 = B - 4 \Rightarrow B = 5$.

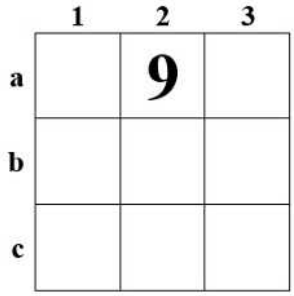
This is the complete thought process based upon which we deduce the base is 5. However, generally, we expect that the problem features all (or almost all) digits, enabling us to estimate the base based on counting the digits.

Problem 9.3

Yup'ik (Kai Low Rui Hao, UKLO 2017)

Yup'ik people have an interesting concept when it comes to counting – the words for the numbers can be broken down into meaningful parts which may be related to body parts. For example, the word for 5, *talliman*, means 'arm' and the word for 6, *arvinlegen*, means 'cross over', as you need to change hands to go on counting.

The Yup'ik people often include geometry in the border patterns of their traditional garments, called "parkas". One such pattern comes from a 3×3 square, as represented below. This is a magic square, constructed by placing the digits 1 to 9 within the cells such that the sum of all the digits in every row, column, and diagonal is the same.



To help you fill in the magic square, the following clues are given in Yup'ik. *Hint:* 294 in Yup'ik is *yuinaat qula cetaman qula cetaman.*

- Rows:
    a. *Yuinaat yuinaq cetaman qula malruk*
    b. *Yuinaat akimiaq malruk akimiaq malruk*
    c. *Yuinaat yuinaak malruk akimiaq atauciq*
- Columns:
    1. *Yuinaat yuinaq atauciq akimiaq pingayun*
    2. *Yuinaat yuinaak malruk yuinaat malrunglegen qula atauciq*
    3. *Yuinaat qula pingayun akimiaq atauciq*

Problem 9.3a Fill in the numbers missing from the magic square above. One digit is already given (a2 = 9).

Problem 9.3b Write in Yup'ik the number given in the diagonal from top left (the number formed by the digits a1-b2-c3).

Solution

1. Solving the magic square

This is in reality the easiest part. It is known that in a magic square the middle number must be 5 (you can attempt a mathematical proof, it is rather easy), hence b2 = 5. Moreover, it is known that the sum on every row, column and diagonal must be 15. Since on the middle column we already have 9 and 5, we deduce that c2 = 1.

On the first row, we already have the digit 9, so the sum of the other two digits must be 6. We have three possibilities: 5 and 1 (impossible, since 5 is already used), 3 and 3 (impossible, since we cannot repeat digits), or 4 and 2 (this is therefore the only possible option). Therefore, the first row can be either 492 or 294. Since in the introduction we are given the Yup'ik name for 294, which does not appear in the crossword clues (hence, it doesn't appear in the square), we deduce that the first row must be 492, so a1 = 4, a3 = 2.

Since we are told that the sum on the diagonals is also constant (so, 15), we can easily deduce that c1 = 8 and c3 = 6, which makes filling in the rest of the square trivial. In the end we get:

	1	2	3
a	4	9	2
b	3	5	7
c	8	1	6

2. Solving the number problem

Once the square is filled in, we can extract all the information in a table, transforming the problem into a classic one, in which we are given some numbers spelled out in Yup'ik:

276 *yuinaat qula pingayun akimiaq atauciq*
294 *yuinaat qula cetaman qula cetaman*
357 *yuinaat akimiaq malruk akimiaq malruk*
438 *yuinaat yuinaq atauciq akimiaq pingayun*
492 *yuinaat yuinaq cetaman qula malruk*

816 *yuinaat yuinaak malruk akimiaq atauciq*
951 *yuinaat yuinaak malruk yuinaat malrunglegen qula atauciq*

The first important observation is based on the last word in every number. We have four types of numbers: ending in *atauciq* (276, 816, 951), ending in *malruk* (357, 492), ending in *cetaman* (294), and ending in *pingayun* (438). Looking closely, one notices that these numbers can be grouped based on their remainder when divided by 5 (i.e., modulo 5). Thus, we can deduce that:

atauciq = 1, *malruk* = 2, *pingayun* = 3, *cetaman* = 4

Replacing these numbers, we get:

276 *yuinaat qula 3 akimiaq 1*
294 *yuinaat qula 4 qula 4*
357 *yuinaat akimiaq 2 akimiaq 2*
438 *yuinaat yuinaq 1 akimiaq 3*
492 *yuinaat yuinaq 4 qula 2*
816 *yuinaat yuinaak 2 akimiaq 1*
951 *yuinaat yuinaak 2 yuinaat malrunglegen qula 1*

Since we assumed that the last number is added, we can simply subtract it (from the number representation) and delete it (from the spelled-out numbers). We are left with:

275 *yuinaat qula 3 akimiaq*
290 *yuinaat qula 4 qula*
355 *yuinaat akimiaq 2 akimiaq*
435 *yuinaat yuinaq 1 akimiaq*
490 *yuinaat yuinaq 4 qula*
815 *yuinaat yuinaak 2 akimiaq*
950 *yuinaat yuinaak 2 yuinaat malrunglegen qula*

Now we can easily notice that the numbers that end in 5 have the last word *akimiaq*, while those ending in 0 have the last word *qula*. Moreover, in the introduction we are told that 5 = *talliman*, so *akimiaq* cannot be 5 as well. We can therefore assume that *akimiaq* = 15, and *qula* = 10. Subtracting these numbers, we are left with:

260 *yuinaat 10 3*
280 *yuinaat 10 4*
340 *yuinaat 15 2*
420 *yuinaat yuinaq 1 3*
480 *yuinaat yuinaq 4*
800 *yuinaat yuinaak 2*
940 *yuinaat yuinaak 2 yuinaat malrunglegen*

Comparing the numbers 260 and 280, we notice that they differ only by 3 vs. 4, while their numerical difference is 20. Thus, we can deduce that *yuinaat* is 20 and it represents a multiplier. Therefore, 280 is written as '20×' '10' '4'. Since $280 = 20 \times 14$, we deduce that the two numbers after '20×' are firstly added together and then multiplied by 20. This is, *yuinaat* $X\ Y = 20(X + Y)$.

Looking at 800, we notice it contains *yuinaak*, instead of *yuinaq*. Since we already know that *yuinaat* = '20×', replacing it we obtain: 800 = '20×' *yuinaak* '2'. So *yuinaak* also means '20×' (basically, it shows that the following number also needs to be multiplied instead of added). Therefore, 940 is written as '20×20× 2 20× ' *malrunglegen*, from which we deduce that *malrunglegen* = 7, i.e., $940 = (20 \times 20 \times 2) + (20 \times 7)$.

Based on these, we can write all the rules and solve the tasks.

Solution 9.3a

	1	2	3
a	4	9	2
b	3	5	7
c	8	1	6

Solution 9.3b $456 = 20 \times (20 + 2) + 15 + 1$ = *yuinaat yuinaq malruk akimiaq atauciq*

Rules: Numbers are written in base 20. Numbers smaller than 20 are written as $10 + A$ or $15 + A$ (where A is 1, 2, 3, or 4). Numbers smaller than 400, i.e., $20A + B$, are written as *yuinaat* $A\ B$, where A and B are between 1 and 19.

Base words are:

1	*atauciq*	6	*arvinlegen* (given in intro)
2	*malruk*	7	*malrunglegen*
3	*pingayun*	10	*qula*
4	*cetaman*	15	*akimiaq*
5	*talliman* (given in intro)	20	*yuinaq* (‘20+’), *yuinaat* (‘20×’)

Addition is implicit and the constituent order is from big to small. For numbers bigger than 800, a new structure is added at the beginning – *yuinaat yuinaak X*, meaning $400X$, and the rest is written as above. For example, $951 = 800 + 151 = (20 \times 20 \times 2) + (20 \times 7) + 11$ = ‘(20×) (20×) (2) (20×) (7) (10) (1)’ = *yuinaat yuinaak malruk yuinaat malrunglegen qula atauciq.*

9.2 Overcounting

Problem 9.4

Umbu-Ungu (Ksenia Gilyarova, IOL 2012)

	Umbu-Ungu
10	*rureponga talu*
15	*malapunga yepoko*
20	*supu*
21	*tokapunga telu*
27	*alapunga yepoko*
30	*polangipunga talu*

	Umbu-Ungu
35	*tokapu rureponga yepoko*
40	*tokapu malapu*
48	*tokapu talu*
50	*tokapu alapunga talu*
69	*tokapu talu tokapunga telu*
79	*tokapu talu polangipunga yepoko*
97	*tokapu yepoko alapunga telu*

telu < yepoko

Problem 9.4a Write in numerals:

a. *tokapu polanigpu*

b. *tokapu talu rureponga telu*

c. *tokapu yepoko malapunga talu*

d. *tokapu yepoko polangipunga telu*

Problem 9.4b Write in Umbu-Ungu: 13, 66, 72, 76, 95.

Solution

Initial observations:

- *supu* is a single word, so most likely is a multiple of the base; therefore the base can be 4, 5, 10, or 20;
- 10 is not a single word, so it is unlikely that the base is 5, 10 or 20. Therefore, it seems to be a base-4 language;
- *tokapu* occurs in the number 35 (but it does not occur in 30), so it most likely means 32. 31, 33 and 34 do not seem to make sense as single words, since they are extremely unlikely as bases, and it can't be 35 either since there are other words following it. Moreover, 32 confirms the hypothesis that the language is base-4, since it is a multiple of 4;
- General structure: (*tokapu*) (X) (Y*-pu(nga)*) (Z).

Based on this structure, we can count the digits. These appear as X or Z (but we notice that they do not occur as Y, which means that Y*-pu(nga)* is a single word). There are only three words appearing in X and Z positions: *talu*, *telu*, *yepoko*.

Moreover, we notice that 48 is *tokapu talu*. Most likely, *talu* is a multiplier for *tokapu*, and, since the numbers smaller than 48 do not contain this structure, we deduce that *talu* = 2. Indeed, for numbers 35 and 40, *tokapu* occurs without a multiplier (1 is implicit), so the first multiplier that ought to appear is 2. Based on the same logic, *yepoko* must mean 3 and, knowing that *telu* < *yepoko*, we get that *telu* is 1.

Replacing these numbers in the given data, we get:

	Umbu-Ungu
10	*rureponga* 2
15	*malapunga* 3
20	*supu*
21	*tokapunga* 1
27	*alapunga* 3
30	*polangipunga* 2

	Umbu-Ungu
35	*tokapu rureponga* 3
40	*tokapu malapu*
48	*tokapu* 2
50	*tokapu alapunga* 2
69	*tokapu* 2*tokapunga* 1
79	*tokapu* 2 *polangipunga* 3
97	*tokapu* 3 *alapunga* 1

Based on the number 48, since we assumed that 2 is a multiplier, we deduce that *tokapu* = 24. Moreover, based on the first column, by subtraction, we obtain the numbers: *rureponga* = 8, *malapunga* = 12, *supu* = 20, *tokapunga* = 20, *alapunga* = 24, *polangipunga* = 28.

However, we notice that we have two different words for 20. Nevertheless, all words than end in *-punga* also have a units digit (1, 2 or 3), so it is possible that each word has two forms, one for when it appears alone and the other that appears when a units digit is added. Thus, based on the words we know already, 35 is written as '24 8 3' (and, indeed, 35 = 24 + 8 + 3).

An interesting thing happens with the number 40. Knowing that *tokapu* = 24, it must be that *malapu* is 16 (but we also know that *malapunga* = 12). Thus, we deduce the following rule: each multiple of 4 is a single or base word (ending in *-pu*). When a units digit (1, 2, or 3) is added, the following multiple of 4 is used to which the suffix *-nga* is added. Therefore, 24 = *tokapu* and 28 = *alapu*, but 25 = *alapunga telu* (basically, (28 − 4) + 1), 26 = *alapunga talu* (28 − 4 + 2) and 27 = *alapunga yepoko* (28 − 4 + 3).

This is a rather common phenomenon called OVERCOUNTING, in which the numbers are regarded as *going towards...*. Thus, 27 can be translated literally as 'three (units) towards 28' (meaning that it is three units past 24). Overcounting occasionally occurs Indo-European languages as well (e.g., in German, the clock time 7.30 is read as *halb acht* (meaning 'half eight') – which is to say, half an hour has passed *towards 8 o'clock*).

A last observation concerns the numbers 48, 50 and 69. We notice that both 48 and 69 use the structure *tokapu talu* (meaning 48), but 50 does not, so we deduce that 50 is written as 24 + (28 − 4) + 2. This can also be considered a type of overcounting. Normally, when "units" get as big as the order, we carryover and increase the multiplier by a unit (e.g., in English after *twenty-nine*, we do not say **twenty-ten*, but rather *thirty*). In this language, however, the change of the order

only occurs at 32 (although the base is 24). An indication pointing towards this is that, although we have single words for all multiples of 4, we do not have a word for 8 (thus, we cannot write the numbers 5, 6 and 7). For this reason, instead of writing $24X + 6$, we actually write $24(X - 1) + 30$.

Based on all these observations, we can solve the tasks and write the solution:

Solution 9.4a
a. $24 + 32 = 56$
b. $24 \times 2 + (12 - 4) + 1 = 57$
c. $24 \times 3 + (16 - 4) + 2 = 86$
d. $24 \times 3 + (32 - 4) + 1 = 101$

Note

In an official solution, it suffices to write just the number (the final result). Nevertheless, writing the structure formed by each morpheme is a safety net to prevent careless mistakes. The same goes for task (b).

Solution 9.4b
$13 = 12 + 1 = (16 - 4) + 1 =$ *malapunga telu*
$66 = 24 \times 2 + 16 + 2 = (24 \times 2) + (20 - 4) + 2 =$ *tokapu talu supunga talu*
$72 = 24 \times 3 =$ *tokapu yepoko*
$76 = 24 \times 2 + 28 =$ *tokapu telu alapu*
$95 = 24 \times 3 + 20 + 3 = (24 \times 3) + (24 - 4) + 3 =$ *tokapu yepoko tokapunga yepoko*

Rules:

- Base words (X): 1 = *telu*, 2 = *talu*, 3 = *yepoko*
- Base words (Y): 12 = *rurepo*, 16 = *malapu*, 20 = *supu*, 24 = *tokapu*, 28 = *alapu*, 32 = *polangipu*
- Addition is implicit.
- Numbers from 9 to 31 are written as: Y-*nga* $X = (Y - 4) + X$.

- Numbers greater than 32, having the structure $24A + B$, are written as *tokapu A B*, where $A = \{2, 3\}$, and B is between 9 and 32 (except for 24, in which case it is directly written as $24(X + 1)$ – in the data, 48 is not written as **tokapu tokapu*, but rather *tokapu talu*).

Problem 9.5

Huli (Bill Huang, UKLO 2016)

The perfect squares from 1 to 100 are written in Huli below, *in random order*:

A. *ngui ki, ngui tebone-gonaga waragaria*

B. *mbira*

C. *ngui dau, ngui waragane-gonaga waragaria*

D. *nguira-ni pira*

E. *nguira-ni mbira*

F. *dira*

G. *maria*

H. *ngui tebo, ngui mane-gonaga maria*

I. *ngui ma, ngui dauni-gonaga maria*

J. *ngui waraga, ngui kane-gonaga pira*

Problem 9.5a For each of them, write its corresponding value.

Problem 9.5b Here are four consecutive numbers written in Huli, in ascending order:

a. *ngui ka, ngui haline-gonaga bearia*

b. *ngui ka, ngui haline-gonaga hombearia*

c. *ngui ka, ngui haline-gonaga haleria*

d. *ngui ka, ngui haline-gonaga deria*

Write their corresponding values.

Problem 9.5c Write in Huli: 2, 4, 6, 7, 22, 44, 66, 77, 88, 173.

Solution

The first step is figuring out the structure of Huli numbers. We have three types of structures: single words (*dira*, *maria*, *mbira*), structures like *nguira-ni X* and structures like *ngui X*, *ngui Y-gonaga Z*. At first sight, we would expect the numbers represented by single words to be the smallest (digits) – taking this with a grain of salt, since some of them could also represent orders, e.g., 100; numbers like *nguira-ni X* are the second smallest ones (we can probably assimilate them with the type $base + X$), while the last category represents the biggest numbers ($base \times A + B$) – although we still do not know why there are three digits in these structures and not only two.

Once the structures are identified, we know exactly where the digits are placed in these structures, so we can try counting them in order to get an estimate of the base. We get the morphemes: *ki*, *tebone*, *waragaria*, *mbira*, *dau*, *waragane*, *pira*, *dira*, *maria*, *tebo*, *mane*, *ma*, *dauni*, *waraga*, *kane*. Nevertheless, we notice that digits can have different forms (since we find the triplets *ma* – *mane* – *maria* and *waraga* – *waragane* – *waragaria*, each of them following the same pattern: *X* – *X-ne* – *X-ria*), and furthermore, those without any suffix appear only after *ngui*, those with the suffix *-ne* appear only with the ending *-gonaga*, while those with the suffix *-ra/-ria* appear only at the very end, after *nguira-ni* or if they are single structures. Therefore, we can assume that they denote the same digit, and each digit has three different forms, depending on the context. Thus, we are left with 10 morphemes: *ki*, *mbi(ra)*, *dau*, *pi(ra)*, *di(ra)*, *dau(ni)*, *ka(ne)*, *tebo*, *waraga*, *ma*, so we could expect this language to be base-10. Nevertheless, if we look at task (b), we notice four more morphemes occur: *bea(ria)*, *hombea(ria)*, *hale(ria)*, *de(ria)*. Therefore, the total number of digits is 14. Since base 13 is extremely unlikely, as well as base 14, it is most likely one of the bases 12, 15 or 16 (among which, base 15 is the most likely one since we have discovered 14 morphemes).

Furthermore, since the base is bigger than 12, we certainly know that 1, 4, and 9 are digits (so they will be represented by a single word). Therefore, based on the previous observation according to which $X <$ *nguira-na X* $<$ *ngui X*, *ngui Y-gonaga Z*, we can already split the numbers into categories. Thus:

> {*mbira*, *dira*, *maria*} must correspond to {1, 4, 9} – the use of curly brackets shows that we are not sure about the exact order,

and

> {*nguira-ni mbira*, *nguira-ni pira*} = {16, 25}.

Since *mbira* appears in both structures, we can try obtaining (by subtraction) the value of *nguira-ni*, which, most likely, represents the base.

We can do this by considering all the possible cases, as follows, and calculating the difference between them:

		mbira		
		1	4	9
nguira-ni	16	15	12	7
mbira	25	24	21	16

Thus, the values for *nguira-ni*, and, implicitly, for the base are 7, 12, 15, 16, 21, 24. Since we know we have roughly 14 digits, we can exclude the bases 7, 21 and 24. Moreover, since 16 appears in the corpus (and it is not a single word), it is unlikely that it will be the base (if *nguira-ni* = 16, it follows that either *mbira* or *pira* is 0, which is impossible). So, we are left with the bases 12 and 15.

		mbira		
		1	4	9
nguira-ni	16	15	12	
mbira	25			

Moreover, we notice that these two bases are possible only if *nguira-ni mbira* is 16. Thus, we already have the first two correspondences: *nguira-ni mbira* = 16, *nguira-ni pira* = 25. Moreover, if *nguira-ni* is 12, then *pira* must be 13, which is highly unlikely, since it would be greater than the base. Therefore, *nguira-ni* = 15 (we therefore talk about a base-15 system), *mbira* = 1, *pira* = 10.

Now we can analyse the more complex numbers (which we know most likely will be written as $15X + Y$). The first step is now writing the remaining squares like this (in order to make comparisons easier):

$$36 = 15 \times 2 + 6 \qquad 64 = 15 \times 4 + 4 \qquad 100 = 15 \times 6 + 10$$

$$49 = 15 \times 3 + 4 \qquad 81 = 15 \times 5 + 6$$

Knowing that *pira* = 10, and 4 can only be *maria* or *dira*, we notice that these two occur multiple times at the end of some numbers, so we can group them like this:

$36 = 15 \times 2 + 6$	*ngui ki, ngui tebone-gonaga waragaria*
$81 = 15 \times 5 + 6$	*ngui dau, ngui waragane-gonaga waragaria*
$49 = 15 \times 3 + 4$	*ngui tebo, ngui mane-gonaga maria*
$64 = 15 \times 4 + 4$	*ngui ma, ngui dauni-gonaga maria*
$100 = 15 \times 6 + 10$	*ngui waraga, ngui kane-gonaga pira*

The numbers from the same cell are not necessarily ordered (i.e., we know that 36 and 81 are represented by the two phrases on the right, but we don't know which is which). Moreover, we deduce that *maria* = 4 (and we are left with *dira* = 9, since it is the only unassigned single word in the data) and *waragaria* = 6. Moreover, knowing that the numbers change their form *X* – *X-ne* – *X-ria*, we can check if any of these numbers occur in any other position. At this stage, we can already replace them in the data. Last but not least, we can delete the units in order to simplify the table.

15×2	*ngui ki, ngui tebone-gonaga*
15×5	*ngui dau, ngui 6 -gonaga*
15×3	*ngui tebo, ngui 4 -gonaga*
15×4	*ngui 4 , ngui dauni-gonaga*
15×6	*ngui 6 , ngui kane-gonaga*

Based on the last example (100), we notice that the first digit, after *ngui*, represents the multiplier of the base. Therefore, we deduce that *tebo* = 3, and we can make the correspondences in the second row:

15×2	*ngui ki, ngui tebone-gonaga*
15×5	*ngui dau, ngui 6 -gonaga*
15×3	*ngui 3 , ngui 4 -gonaga*
15×4	*ngui 4 , ngui dauni-gonaga*
15×6	*ngui 6 , ngui kane-gonaga*

Furthermore, the number after *ngui* and the units are enough to calculate any number, so what could be the role of *ngui Y-gonaga*? Firstly, since it does not offer any supplementary information (it is redundant, since the value of the number can be calculated without it and only based on *ngui X* and *Z*), it is probably somehow connected to one of the other digits and, since it also starts with *ngui*, it is most likely related to *X*. Looking at the example 15×3, we notice that *Y* is 1

unit greater than X. Therefore, we deduce that $15X + Z$ is written in Huli as *ngui* X, *ngui* $(X + 1)$-*gonaga* Z, and we can then determine all the correspondences:

Solution 9.5a

A – 36	C – 81	E – 16	G – 4	I – 64
B – 1	D – 25	F – 9	H – 49	J – 100

Moreover, we can make a table with the three different forms of the digits we have. Since this is base 15, we expect to have *units* from 1 to 14:

Digits	units (X)	*ngui* X	*ngui* X-*gonaga*
1	*mbira*		
2		*ki*	
3		*tebo*	*tebone*
4	*maria*	*ma*	*mane*
5		*dau*	*dauni*
6	*waragaria*	*waraga*	*waragane*
7		*ka*	*kane*
8			*haline*
9	*dira*		
10	*pira*		
11			
12			
13			
14			

Solution 9.5b In task (b), we are given four extra consecutive numbers, and all of their unit digits are new (they do not occur anywhere else in the problem). Looking at the table above, the only four consecutive digits that do not occur anywhere else are 11–14, and, knowing that the numbers are in ascending order, we deduce that: *ngui ka, ngui haline-gonaga bearia* $= 7 \times 15 + 11 = 116$ and the other numbers are 117, 118, and 119 respectively.

Now we can fill in the table with the additional information:

Digits	units (X)	*ngui X*	*ngui X-gonaga*
1	*mbira*		
2		*ki*	
3		*tebo*	*tebone*
4	*maria*	*ma*	*mane*
5		*dau*	*dauni*
6	*waragaria*	*waraga*	*waragane*
7		*ka*	*kane*
8			*haline*
9	*dira*		
10	*pira*		
11	*bearia*		
12	*hombearia*		
13	*haleria*		
14	*deria*		

The last thing we need to do is analyse the way in which the three forms are constructed. We notice that the second column (*ngui X*) is the base form and, in order to obtain the first column, we add the suffixes *-ra/-ira*, while the suffixes *-ne/-ni* are used to form the third column.

Analysing the occurrence of the suffix *-ra*, we realise that it is only used if the base form ends in *i*. So we can write the phonological rule: *-ria* $\longrightarrow$ *-ra* / *i* _ . For the last column, things are a bit more complicated since there is only one instance in which *ni* appears (with *dau*). Since we have only one example, it is hard to define a precise phonological rule. One option is to consider that *ni* appears after *u*, in which case we would talk about an assimilation of the vowel height, since both *i* and *u* are high vowels.

Thus, we can solve task (c).

Solution 9.5c

2 = *kiria*

4 = *maria*

6 = *waragaria*

7 = *karia*

22 = *nguira-ni karia*

44 = *ngui ki, ngui tebone-gonaga deria*

66 = *ngui ma, ngui dauni-gonaga waragaria*

77 = *ngui dau, ngui waragane-gonaga kiria*

88 = *ngui dau, ngui waragane-gonaga haleria*

173 = *ngui bea, ngui hombeane-gonaga halira*

Rules:

- *nguira-ni* $X_2 = 15 + X$
- *ngui* X_1, *ngui* $(X_3 + 1)$-*gonaga* $Y_2 = 15X + Y$
- Subscripts 1, 2 and 3 denote the form of the digit:
  - Form 2 – base form;
  - Form 1 – add the suffix *-ria* (*-ria* ⟶ *-ra* / *i* _);
  - Form 3 – add the suffix *-ne* (*-ne* ⟶ *-ni* / *{i, u}* _);

It is interesting to note the phenomenon in which we write both *ngui* X and *ngui* $(X + 1)$-*gonaga*, although the second part is redundant. Basically, we can consider this structure to mean Z units after *ngui* X, towards *ngui* $(X+1)$-*gonaga*. For example, the number 49 (*ngui tebo, ngui mane-gonaga maria*) can be read as 'four (units) away from the third group of 15s towards the fourth (group of 15s)'. This is another example of overcounting in which the units are enclosed between two consecutive multiples of the base.

9.3 Subtractive systems

We have seen so far that the basic operations in every number system are addition and multiplication. There is, however, a special type of number system, subtractive systems, in which, besides addition and multiplication, subtraction also plays an important role. For example, in Problem 9.4, the numbers were formed as $M+1$, $M+2$, $M+3$, N (where M and N are orders – we ignore the fact that M was derived from N). In subtractive systems, we can talk about numbers formed as $\{M, M+1, M+2, N-2, N-1, N\}$ or $\{M, M+1, M+2, M+3, N-1, N\}$. Thus, 19 can, for example, be written as $20-1$ instead of $16+3$. It is important to not confuse subtractive systems with overcounting. In the case of overcounting, no subtraction is involved, but rather the order we use is one unit higher (since it is counted towards it), but still, the units are added. In subtractive systems, the units (or some of them) are subtracted from the order.

Currently, no purely subtractive system is known (in which addition does not play some role), and all subtractive systems use both addition and subtraction. Most subtractive systems only employ subtraction for 1 (therefore, for example, in such a base-10 subtractive system, the number 38 is written as $30 + 8$, but the number 39 is written as $40 - 1$).

Problem 9.6

Yoruba (Harold Somers, UKLO 2020)

Here are some numbers in Yoruba:

èji	2	*ẹ́ẹ́rìndilogóji*	36
ẹ̀rin	4	*ẹ̀rìndogóji*	44
àrun	5	*àádorin*	70
ẹ̀rinlá	14	*ẹ́ẹ́tàdilogórin*	77
eéjìdilogun	18	*ẹ̀tàdogórin*	83

Problem 9.6a Write in numerals:

a. *àádota*
b. *àrùndogórin*
c. *aárùndilogórin*
d. *ẹ̀tàdogórun*
e. *òkándilogóji*

Problem 9.6b Write in Yoruba: 12, 45, 57, 90, 99.

The marks above the vowels denote tones; *e* and *ẹ* are distinct vowels.

Solution

Firstly, we can notice the pair 4 and 14, in which the only difference is the suffix *-lá*. We can therefore deduce that $10 + X = X$*-lá*, so this system is most likely base-10.

Next, we have the pair 36 and 44, in which the only difference is *-il-*. Moreover, the first part of the word is highly similar to the word for 4 (in the case of 44 it is actually identical, while in the case of 36 it is slightly mutated), and the

same thing goes for the pair 77 and 83. We notice that a common aspect of these numbers is that they are symmetrical with respect to the closest tens (thus, 36 and 44 are symmetrical with respect to 40, i.e., $36 = 40 - 4$ and $44 = 40 + 4$, while 77 and 83 are symmetrical with respect to 80, i.e. ±3). This can make us think of a subtractive system. Moreover, the last morpheme of 36 and 44 is *óji* (which resembles *èji* = 2), while the last morpheme of 77 and 83 is *órin* (~*ẹ̀rin*). Since this number cannot refer to the units (we know already that the units are 4 and 3, respectively), it most likely represents the tens. Thus, 40 = *óji*, and 80 = *órin*. Therefore, we realise it is actually a base-20 system, and the structure of the numbers is:

- X-*lá* $= 10 + X$
- U-*dog*-$Z = 20Z + U$
- U-*dilog*-$Z = 20Z - U$

Moreover, we notice that 70 has a special form, also derived from *órin* = 80, so, most probably, it represents $80 - 10$.

Thus, we notice that the digits have different forms depending on the context in which they appear (single units, $10 + X$, added units, subtracted units, $20X$, $20X - 10$). Moreover, we already noticed that $10 + X$ uses the same form of X as the single unit, so we can combine these two forms.

Combining everything into a table, we get:

	Unit / $10 + X$	+ units	− units	$20X$	$20X - 10$
1				*un*	
2	*èji*		*eéjì*	*óji*	
3		*ẹ̀tà*	*ẹ́ẹ́tà*		
4	*ẹ̀rin*	*ẹ̀rìn*	*ẹ́ẹ́rìn*	*órin*	*àádorin*
5	*àrun*				

Since we have all the different forms of the digit 4, we can see the transformations that take place. In order to form the added units, the second vowel receives a grave accent; to form the subtracted units, the first vowel is doubled (of which the first is toneless, while the second has an acute accent), and the final vowel gets a grave accent. In order to form $20X$, the first vowel becomes *ó*, while for the formation of $20X - 10$, the first vowel is replaced by *àádo-*.

The only exception is the number 20 (20 × 1), but, as explained before, we can expect that digit 1 might have some irregular forms.

Another simpler way to write the rules is to notice that each digit has the structure *V̀₁CV₂(n)*. We can easily derive the rest of the forms: + units = *V̀₁CV̀₂(n)*, − units = *V₁V́₁CV̀₂(n)*, 20*X* = *ó CV₂(n)*, 20*X* − 10 = *àádo CV₂(n)*, once again, mentioning that for 1 the form 20*X* is irregular (*un*).

Thus, we can solve all the tasks:

Problem 9.6a a. 50 b. 85 c. 75 d. 103 e. 39

Note

In order to figure out the meaning of *òkán* in task (e), we need to look at the table above (which we previously filled in with the additional information that we discovered in task (a)) and we notice that 1 is the only digit for which we do not know the subtracted units form (column 3 from the table above). So we can conclude that *òkán* is the subtracted form of 1. Moreover, we see again that its form is irregular.

Problem 9.6b

12 = *èjilá*
90 = *àádorun*
57 = *ẹ́ẹ́tàdilogóta*
45 = *àrùndogóji*
99 = *òkándilogórun*

9.4 Body-part counting systems

In this type of system, the names of the digits are derived from the names of body parts. Thus, 1 = ‘little finger’, 2 = ‘ring finger’, 3 = ‘middle finger’, 4 = ‘index finger’, 5 = ‘thumb’, 6 = ‘hand’, 7 = ‘elbow’, 8 = ‘arm’, 9 = ‘shoulder’, 10 = ‘ear’, 11 = ‘head’. Depending on the language, the number of body parts which are used can vary (some languages also include terms for ‘wrist’, ‘forearm’, ‘arm’, ‘eye’, etc.). We mentioned in the beginning of this chapter that the base of these systems is usually between 22 and 28. The interesting thing about these systems starts as soon as we reach the word for ‘head’ (or ‘forehead’, ‘nose’, etc.), which

represents the centre. From here on, the words are repeated in reverse order, with an additional morpheme/word which means 'opposite/other'. Thus, continuing the series above, 12 = 'opposite ear', 13 = 'opposite shoulder', 14 = 'opposite arm', 15 = 'opposite elbow', 16 = 'opposite hand', 17 = 'opposite thumb', and so on.

At first sight, this particularity, although interesting, does not seem to pose any problems, but let us consider the following examples:

(1) Let us first consider a classic number system (not based on body parts) – for example, Japanese – and let us consider two pairs of numbers:
san – *juusan* and *go* – *juugo*

In the case of these numbers, we have pairs following the pattern *X* – *juu-X*. So, if we are told that *san* = 3 and *juusan* = 13, we could instantly deduce that *juu-* = 10 (since 13 − 3 = 10), and, if, additionally, we are told that *go* = 5, we will immediately claim that *juugo* = 15.

(2) Let us now consider a body-part-based system – e.g., the Kombai system, used in New Guinea – and let us analyse the following numbers:
woro = 4, *imofo woro* = 20, *go* = 6, *imofo go* = ?

If we followed the same method as in the previous case, we would again notice the pattern *X* – *imofo X*, and based on the pair 4 – 20, we would deduce that *imofo* = 16, so we would say that *imofo go* = 22. Nevertheless, in reality, *imofo go* = 18, since the Kombai number system is based on body parts, centred on 12. Therefore, we notice that for these systems, the order is not obtained by subtracting two numbers like *X* and *imofo X*, but rather by *adding* them.

For Japanese: *juusan* − *san* = 10 = *juugo* − *go*, but for Kombai *woro* + *imofo woro* = *go* + *imofo go*. This is the main issue (and difficulty) with this type of system: the fact that it can easily be confused with a classic system, but, in fact, the base is obtained by adding two similar numbers rather than subtracting them.

A separate discussion concerns what one considers to be the base of such a system. We return to the example of Kombai, where we mentioned that the system is "centred on 12", but *imofo go* + *go* = 24. For linguistics problems, it is preferred that the base is considered as the double of the centre, because, in this way, we will have the relation *imofo X* = 24 − *X*. If we considered the base to be 12, the relation between the two numbers would become: $X = 12 - a$, while *imofo* $X = 12 + a$, which is much harder to use in practice.

9.5 Time

Problems related to time (reading the clock, calendar, etc.) can be considered a subtype of number systems since these problems will almost always feature numbers. It is rather common for the names of the days of the week and the months of the year to be formed based on numbers ('Monday' = day 1, 'March' = month 3). This phenomenon is obvious in Chinese, where, starting from the numbers 1 = 一, 2 = 二, 3 = 三, 4 = 四, we can write the days of the week and the months as follows: 'Monday' = 星期一, 'Tuesday' = 星期二, 'Wednesday' = 星期三, 'Thursday' = 星期四; 'January' = 一月, 'February' = 二月, 'March' = 三月, 'April' = 四月.

Although these problems are relatively rare at the level of international linguistics competitions, they can feature some interesting phenomena, as shown in the following two problems.

Problem 9.7

Czech (Mirjam Fried, Princeton)

Here are some examples of how to tell the time in Czech:

za pět minut osm	'five minutes to eight'
za deset minut osm	'ten minutes to eight'
čtvrt na osm	'quarter past seven'
za sedm minut osm	'seven minutes to eight'
za osm minut čtvrt na osm	'seven minutes past seven'
za deset minut čtvrt na sedm	'five minutes past six'
půl osmé	'half past seven'
půl deváté	'half past eight'
za deset minut půl šesté	'twenty minutes to five'
čtvrt na deset	'quarter past nine'

Problem 9.7a Translate into Czech:

1. 'twenty-three minutes past five'
2. 'ten minutes to nine'

Solution

We notice three types of structures:

- *čtvrt na ...* = ‘quarter past ...’
- *půl ...-é* = ‘half past ...’
- *za ... minut (čtvrt na / půl) ...* – otherwise

It is obvious that *minut* = ‘minute(s)’, so if we compare the first two examples, we obtain: *osm* = 8, *deset* = 10 and *pět* = 5. Moreover, we deduce that *za X minut Y* = ‘X minutes to Y’.

We notice that *osm* occurs in the phrase ‘half past seven’, but not in ‘half past eight’. This indicates that Czech also features overcounting when it comes to telling the time (similar to German). Therefore, *půl X-é* = ‘half past $(X - 1)$’, from which we deduce that *devát* = 9.

Moreover, looking at the example ‘quarter past seven’, in which the word *osm* = 8 occurs again, we deduce that overcounting is also applied for quarters, so *čtvrt na X* = ‘quarter past $(X - 1)$’ (or ‘quarter towards X’).

The only structures we are left to analyse are those following the pattern *za X minut ... na Y*. We can separate them from the rest and replace the words we already know:

za osm minut čtvrt na osm	‘seven minutes past seven’
za deset minut čtvrt na sedm	‘five minutes past six’
za deset minut půl šesté	‘twenty minutes past five’

We already know that the hour is placed in the end, so we can directly replace the structures *čtvrt na X* and *půl šesté*. Moreover, we know all the words, so we notice that ‘seven minutes past seven’ is written as *za* ‘8 minutes quarter past seven’, and ‘five minutes past six’ is written as *za* ‘10 minutes quarter past six’. We therefore notice that the Czech system is based on overcounting. Therefore, ‘seven minutes past seven’ is translated as ‘8 minutes to [quarter towards 8]’ or ‘8 minutes to [quarter past 7]’.

The same thing is noticed in the structure *za deset minut půl šesté* which is literally translated as ‘10 minutes to [half towards 6]’ or ‘10 minutes to [half past 5]’.

So we now know all the possible structures so we can write the rules and solve the tasks.

Rules:

- *půl X-é* = ‘half past $(X - 1)$’
- *čtvrt na X* = ‘quarter past $(X - 1)$’
- *za X minut Y* = ‘X minutes to Y’, where Y can be a full hour (o’clock) or any of the two structures above.

Problem 9.7a 1. ‘23 past 5’ = ‘7 minutes to [half past five]’ = ‘7 minutes to [half towards 6]’ = *za sedm minut půl šesté*

2. ‘10 past 9’ = ‘5 minutes to [quarter past nine]’ = ‘5 minutes to [quarter towards 10]’ = *za pět minut čtvrt na deset*

In this case, overcounting is used for all the structures, but, as mentioned above, there are languages in which only certain structures use overcounting. For example, in German, overcounting is only used in the half-past constructions – *halb X* = ‘half past $(X - 1)$’.

Problem 9.8

Swahili (Nilai Sarda, PLO 2014)

Here are some Swahili phrases and their English translations *in random order*:

1.	*jumamosi, saa moja usiku*	A.	‘Sunday, 1.00 AM’
2.	*jumapili, saa tatu na robu asubuhi*	B.	‘Sunday, 7.30 AM’
3.	*jumamosi, saa saba usiku*	C.	‘Sunday, 9.15 AM’
4.	*jumamosi, saa mbili na robu usiku*	D.	‘Tuesday, 12.15 PM’
5.	*jumanne, saa tano na nusu usiku*	E.	‘Tuesday, 11.30 PM’
6.	*jumanne, saa sita na robu asubuhi*	F.	‘Saturday, 10.30 AM’
7.	*jumamosi, saa nne na nusu asubuhi*	G.	‘Saturday, 7.00 PM’
8.	*jumapili, saa moja na nusu asubuhi*	H.	‘Saturday, 8.15 PM’

Problem 9.8a Determine the correct correspondences.

Problem 9.8b Translate into English:

9. *jumatano, saa moja na robu asubuhi*
10. *jumapili, saa nne na nusu asubuhi*

Problem 9.8c Translate into Swahili:

11. 'Monday, 12.15 AM' 12. 'Monday, 10.00 PM'

Solution

The first step is noticing the structure of the phrases. All of them start with a word separated by a comma, which has the prefix *juma-*. This most likely represents the day (it is unlikely that all that variety of times would use the same one-word structure). Therefore, the structure after the comma must be the time. For this, we notice two types of structures:

saa X usiku and *saa X na {robu/nusu} {usiku/asubuhi}*

We can assume that the simplest structure refers to the o'clock times, which is also reinforced by the fact that we only have two such structures in both Swahili and English. Thus:

jumamosi, saa moja usiku	'Sunday, 1.00 AM'
jumamosi, saa saba usiku	'Saturday, 7.00 PM'

Curiously, in Swahili, the same name seems to be used to express different days in English. We will have to explain why that is in due course.

Among the remaining examples, only one more contains the hour 7 (although it is AM, instead of PM), and only one other example starts with either *saa saba* or *saa moja* (since we do not know the correspondence between the two). Since phrase 8 contains *saa moja*, we deduce that it corresponds to translation B. Moreover, we deduce that *saa moja* means 7 o'clock, so we can make three correspondences:

1.	*jumamosi, saa moja usiku*	G.	'Saturday, 7.00 PM'
3.	*jumamosi, saa saba usiku*	A.	'Sunday, 1.00 AM'
8.	*jumapili, saa moja na nusu asubuhi*	B.	'Sunday, 7.30 AM'

Based on example 8, we can infer that *na nusu* means 'half past'. Moreover, most likely, *usiku* and *asubuhi* correspond, in some way, to the notion of AM and PM, but it is not a one-to-one correspondence, since *usiku* is used for both 7 PM and 1 AM.

We only have two more examples which contain *na nusu* (5 and 7), and the English times correspond to 10.30 AM and 11.30 PM. Since *usiku* is used for both 7 PM and 1 AM, we can assume that it will also be used for 11.30 PM, so that *usiku* suggests the idea of 'evening/night'.

We get two more correspondences:

1.	*jumamosi, saa moja usiku*	G.	'Saturday, 7.00 PM'
3.	*jumamosi, saa saba usiku*	A.	'Sunday, 1.00 AM'
8.	*jumapili, saa moja na nusu asubuhi*	B.	'Sunday, 7.30 AM'
5.	*jumanne, saa tano na nusu usiku*	E.	'Tuesday, 11.30 PM'
7.	*jumamosi, saa nne na nusu asubuhi*	F.	'Saturday, 10.30 AM'

The last three correspondences can be easily made. Only one of the phrases contains the word *usiku*, and among the times 8.15 PM, 9.15 AM, 12.15 PM, the only one that refers to evening time is 8.15 PM. Therefore, 4=H. Among the remaining two phrases, one starts with *jumanne*, which seems to mean 'Tuesday'. Therefore, we have the correspondences:

1.	*jumamosi, saa moja usiku*	G.	'Saturday, 7.00 PM'
3.	*jumamosi, saa saba usiku*	A.	'Sunday, 1.00 AM'
8.	*jumapili, saa moja na nusu asubuhi*	B.	'Sunday, 7.30 AM'
5.	*jumanne, saa tano na nusu usiku*	E.	'Tuesday, 11.30 PM'
7.	*jumamosi, saa nne na nusu asubuhi*	F.	'Saturday, 10.30 AM'
2.	*jumapili, saa tatu na robu asubuhi*	C.	'Sunday, 9.15 AM'
4.	*jumamosi, saa mbili na robu usiku*	H.	'Saturday, 8.15 PM'

To sum up, we notice that *usiku* is used between 7 PM and 1.00 AM (it corresponds to 'evening/night'), while *asubuhi* is used between 7.30 AM and 12.15 PM (it corresponds to 'morning/day').

One of the issues we noticed in the beginning was that the Swahili days of the week do not match the English ones. So let us separate (and order chronologically) these dates:

6.	*jumanne, saa sita na robu asubuhi*	D.	'Tuesday, 12.15 PM'
5.	*jumanne, saa tano na nusu usiku*	E.	'Tuesday, 11.30 PM'
7.	*jumamosi, saa nne na nusu asubuhi*	F.	'Saturday, 10.30 AM'
1.	*jumamosi, saa moja usiku*	G.	'Saturday, 7.00 PM'
4.	*jumamosi, saa mbili na robu usiku*	H.	'Saturday, 8.15 PM'
3.	*jumamosi, saa saba usiku*	A.	'Sunday, 1.00 AM'
8.	*jumapili, saa moja na nusu asubuhi*	B.	'Sunday, 7.30 AM'
2.	*jumapili, saa tatu na robu asubuhi*	C.	'Sunday, 9.15 AM'

We notice that *jumanne* always corresponds to 'Tuesday', but *jumamosi* includes both 'Saturday' and 'Sunday'. Nevertheless, if we look closely, we notice that the Sunday phrase (which in Swahili is translated as 'Saturday') has the time

1.00 AM. Therefore, we deduce that, in Swahili, the day does not change at midnight, but rather when *usiku* ends. So, according to the data given, the day starts and ends when evening becomes morning, some time after 1.00 AM and before 7.30 AM.

We can now compare the usage of numbers for day and hour. The only morpheme that is repeated is *nne* (meaning 'Tuesday' and 10). Nevertheless, if we analyse the data closely, we also notice other pairs like *pili* – *mbili* ('Sunday', 8) and *mosi* – *moja* ('Saturday', 7). Thus, it seems like the Swahili week begins on Saturday, and the Xth day of the week is translated using the number $6 + X$ (i.e., first day, Saturday, is translated using 7; day 2, Sunday, using 8, etc.).

Based on this, we can solve the tasks (b) and (c).

Problem 9.8b 9. 'Wednesday, 7.15 AM'

jumatano is formed from *tano* = 11, so it corresponds to 'Wednesday'; *saa moja na robu asubuhi* = 7.15 AM.

10. 'Sunday, 10.30 AM'

Problem 9.8c 11. *jumapili, saa sita na robu usiku*

Firstly, since the time is 12.15 AM, we know that in Swahili the day before will be used, hence, 'Sunday'.

12. *jumatatu, saa nne usiku*

Rules:

1. General structure: [*Day of the week*], [*Time*]
2. [*Time*]: *saa* [X] [Y] [Z]
   X = hour
   Y: *na robu* = 'quarter past', *na nusu* = 'half past'
   Z: *usiku* = 'night-time', *asubuhi* = 'day-time'
3.

Hour	Day of the week	
7 *moja*	*jumamosi*	'Saturday'
8 *mbili*	*jumapili*	'Sunday'
9 *tatu*		'Monday'
10 *nne*	*jumanne*	'Tuesday'
11 *tano*		'Wednesday'
12 *sita*		'Thursday'
1 *saba*		'Friday'

Discussion (not part of the solution)

In reality, the Swahili system is much more interesting and simpler than it may seem. In countries where Swahili is spoken, the sun rises at around 7 AM so the people consider 7 AM to be the first hour of the day (8 AM is the second hour, etc.), and the sunrise is also the moment in which the day changes. The words *usiku* and *asubuhi* also refer strictly to sunrise and sunset (*usiku* = between 7 PM and 6.59 AM = 'after sunset', and *asubuhi* = between 7 AM and 6.59 PM = 'after sunrise').

Thus, the days of the week are named: day 1, day 2 etc. (first day is Saturday) and the table becomes:

Hour	Day of the week	
1 *moja*	*jumamosi*	'Saturday'
2 *mbili*	*jumapili*	'Sunday'
3 *tatu*		'Monday'
4 *nne*	*jumanne*	'Tuesday'
5 *tano*		'Wednesday'
6 *sita*		'Thursday'
7 *saba*		'Friday'

saa $X = (X + 6)$ 'o'clock'

This system of counting hours based on the sunrise is rather common in Equatorial Africa, where the day length is relatively constant, and a similar system is also used in Ethiopia. What is important to remember is that every language is used to express a culture and that culture might be completely different from yours. Therefore, when approaching a linguistics problem, it is important to keep an open mind and not expect all things to work in the way you are used to (e.g., the day changes at midnight, the numbers are in base 10, the week has 7 days, etc.).

9.6 Practice problems

Problem 9.9

Danish (Michael Swan, UKLO 2012)

Here are some numbers written in Danish:

tre	3	*tredive*	30
fire	4	*fyrre*	40
fem	5	*syvoghalvtreds*	57
seks	6	*tres*	60
syv	7	*otteoghalvfjerds*	78
tyve	20	*firs*	80

Problem 9.9a Write in numerals:

a. *treogtyve*
b. *seksoghalvtreds*
c. *fireogtres*
d. *femoghalvfjerds*
e. *syvoghalvfems*

Problem 9.9b Write in Danish: 8, 27, 36, 65, 98.

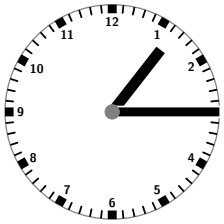
Problem 9.10

Estonian (Babette Verhoeven-Newsome, UKLO 2014)

The following expressions show how to tell the time in Estonian:

Kell on üks

Kell on kaks



Veerand kaks

Pool neli

Kolmveerand üksteist

Viis minutit üks läbi

Here are some numbers in Estonian:

6 = *kuus* 7 = *seitse* 8 = *kaheksa* 9 = *üheksa* 10 = *kümme*

Problem 9.10a What do the following Estonian time expressions mean? Write with numbers:

a. *Kakskümmend viis minutit üheksa läbi*

b. *Veerand neli*

c. *Pool kolm*

d. *Kolmveerand kaksteist*

e. *Kolmkümmend viis minutit kuus läbi*

Problem 9.10b Write the following times in Estonian:

f. 8:45 g. 4:15 h. 11:30 i. 7:05 j. 12:30

Problem 9.11

Waorani (Dragomir R. Radev, UKLO 2012)

Here are some equalities written in Waorani. Each sequence printed in bold represents one number from 1 to 10.

(1) *mẽña mẽña mẽña mẽña + mẽña go mẽña = ãẽmãẽmpoke go aroke × 2*

(2) $aroke^2 + mẽña^2 = ãẽmãẽmpoke$

(3) $ãẽmãẽmpoke\ go\ aroke^2 = mẽña\ go\ mẽña \times ãẽmãẽmpoke\ mẽña\ go\ mẽña$

(4) *mẽña × ãẽmãẽmpoke = tipãẽmpoke*

Problem 9.11a Write in Waorani the numbers from 4 to 10.

Problem 9.12

Selkup (Svetlana Burlak, MSK 1991)

Here are some numbers in Selkup and their values *in random order*:

> *somplylasar ɛj šitty, muktyssar ɛj ukkyr, sompylasar ɛj sompyla, šittysar, ukkyr ca muktyssar, šitty ca tɛ̄sar, sompylasar ɛj sel'cy, ukkyr ca tōn*
> 20, 38, 52, 55, 57, 59, 61, 99

Problem 9.12a Determine the correct correspondences.

Problem 9.12b Write in Selkup: 41, 48, 77, 98.

A bar above a vowel denotes length.

Problem 9.13

Vambon (Alexander Piperski, Elementy)

Below are given the following equalities written in Vambon *in random order*:

$3 \times 1 = 3$ $3 \times 2 = 6$ ... $3 \times 9 = 27$

takhem × *ambalop* = *emkelop*
takhem × *hitulop* = *silutop*
takhem × *javet* = *emsanop*
takhem × *kumuk* = *emalin*
takhem × *mben* = *emben*
takhem × *muyop* = *emhitulop*
takhem × *sanop* = *takhem*
takhem × *sanopkunip* = *kumuk*
takhem × *takhem* = *javet*

Problem 9.13a Write in numerals:

a. (*emnggokmit* + *nggokmit*) ÷ *sanopkunip* = *kalit*
b. *emambalop* − *emjavet* = *hitulop*

Problem 9.13b Write in Vambon:

c. $20 + 22 - 26 = 16$ d. $12 + 13 = 25$

Problem 9.13c You are given the following Vambon words:

kalit	'nose'	*kelop*	'eye'
kumuk	'wrist'	*muyop*	'elbow'
nggokmit	'neck'	*sanopkunip*	'ring finger'
silutop	'ear'		

Which body parts do the following words refer to?

e. *ambalop* g. *malin* i. *sanop*
f. *javet* h. *mben*

Problem 9.13d What does the prefix *em-* mean in Vambon?

Problem 9.14

Alamblak (Roxana Dincă, RoLO 2013)

Here are some Alamblak numbers and their numerical values *in random order*:

a. *yima hosfirpati tir hosfirpat*
b. *yima yohtti tir hosfi rpat*
c. *yima hosfi hosf*
d. *yima hosfi tir hosf*
e. *yima yohtti tir yohtti rpat*
f. *yima hosfirpati tir hosfi hosfirpat*
g. *yima hosfirpati tir hosfirpati hosfirpat*
h. *yima yohtti tir hosfirpati rpat*
i. *yima hosfihosfi tir yohtti hosfihosf*
j. *yima hosfi tir hosfi hosf*

26, 31, 36, 42, 50, 52, 73, 75, 78, 89

Moreover, it is known that:

1 = *rpat*	2 = *hosf*	3 = *hosfirpat*	4 = *hosfihosf*
5 = *tir yohtt*	6 = *tir yohtti rpat*	11 = *tir hosfi rpat*	

Problem 9.14a Determine the correct correspondences.

Problem 9.14b Write in numerals:

k. *yima hosfirpati hosfihosf + yima yohtti tir hosf =*
= yima hosfihosfi tir hosfi hosfihosf

l. *tir yohtti hosf + tir hosfi hosf = tir hosfirpati hosfihosf*

Problem 9.14c Write in Alamblak: 21, 48, 83.

Problem 9.15

Chabu (Danylo Mysak, UkrLO 2018)

Below are given the first four multiples of the number *efi tʃumtʃum eku bab eku iŋki*, written in Chabu in ascending order (if the number is X, the four numbers below represent 2X, 3X, 4X, and 5X, respectively):

bab ef eku efi tʃumtʃum eku iŋki

ink ufe kor eku bab eku bab

ink ufe kor eku bab ef eku bab

bab ufe kor

Problem 9.15a Write in Chabu all the divisors of the number *bab eku iŋki ufe kor* (including 1). If you consider some of them can be written in different ways, write all the possibilities.

Problem 9.16

Tifal (Svetlana Burlak & Peter Arkadiev, MSK 2017)

Here are some equalities written in Tifal. It is known that none of the numbers in the problem are greater than 30.

asumano × *aleeb* = *bokob*
asumano × *ataling* = *tadang*
bokob × *ataling* = *ataling madi*
bokob × *asumano* = *nakal madi*
asumano × *feet* = *feet madi*
ataling × *ataling* = *tadang madi*
asumano + *ataling* = *feet*
feet + *miit* = *feet madi*
tadang + *ataling* = *tadang madi*

Problem 9.16a Write in numerals:

(1) *beeti* + *nakal* = *beeti madi*
(2) *bokob* + *maakob* = *feet*
(3) *awok* × *awok* = *asumano madi*

Problem 9.16b Write in Tifal the results of the following equalities:

(4) *tadang* + *miit* =
(5) *ataling madi* − *aleeb* =

Problem 9.17

Mansi (Ivan Derzhanski, IOL 2005)

Here are some numbers in Mansi (written in Latin script):

ńollow	8
atxujplow	15
atlow nopɒl ontɔllow	49
atlow	50
ontɔlsāt ontɔllow	99
xōtsātn xōtlow nopɒl at	555
ontɔllowsāt	900
ontɔllowsāt ńollowxujplow	918

Problem 9.17a Write in numerals:

a. *atsātn at*
b. *ńolsāt nopɒl xōt*
c. *ontɔllowsātn ontɔllowxujplow*

Problem 9.17b Write in Mansi: 58, 80, 716.

9.7 Solutions of practice problems

Solution for practice problem 9.9. Danish

Solution 9.9a a. 23 b. 56 c. 64 d. 75 e. 97

Solution 9.9b 8 = *otte* 36 = *seksogtredive* 98 = *otteoghalvfems*
27 = *syvogtyve* 65 = *femogtres*

Rules: The Danish system is based on *scores* (20s) rather than *tens*, thus: *tres* (3) $\longrightarrow$*treds* ($20 \times 3 = 60$); *fire* (4) $\longrightarrow$*firs* ($20 \times 4 = 80$). Note that there are special words for 20, 30, and 40.

The particle *halv-* attached before means -10 ('halfway towards'), while noting that the word might undergo some phonological changes (*tres* – *treds*, *firs* – *fjerds*): *treds* (60) $\longrightarrow$*halvtreds* ($60 - 10 = 50$).

The numbers are formed following the structure *U ogS* (*U* represents units and *S* scores; *og* means 'and').

Solution for practice problem 9.10. Estonian

Solution 9.10a a. 9:25 b. 3:15 c. 2:30 d. 11:45 e. 6:35

Solution 9.10b
f. *kolmveerand üheksa*
g. *veerand viis*
h. *pool kaksteist*
i. *viis minutit seitse läbi*
j. *pool üks*

Rules:

- X:00 = *kell on* X
- X:15 = *veerand* (X+1)
- X:30 = *pool* (X+1)
- X:45 = *kolmveerand* (X+1)
- Else, X:Y = Y *minutit* X *läbi*
- $10 + X$ = X-*teist*
- $10X + Y$ = X-*kümmend* Y

Solution for practice problem 9.11. Waorani

Solution 9.11a

4 *mẽña go mẽña*

5 *ãẽmãẽmpoke*

6 *ãẽmãẽmpoke go aroke*

7 *ãẽmãẽmpoke go mẽña*

8 *mẽña mẽña mẽña mẽña*

9 *ãẽmãẽmpoke mẽña go mẽña*

10 *tipãẽmpoke*

Solution for practice problem 9.12. Selkup

Solution 9.12a

somplylasar ɛj šitty	= 52	*muktyssar ɛj ukkyr*	= 61
sompylasar ɛj sompyla	= 55	*šittysar*	= 20
ukkyr ca muktyssar	= 59	*šitty ca tēsar*	= 38
sompylasar ɛj sel'cy	= 57	*ukkyr ca tōn*	= 99

Solution 9.12b 41 = *tēsar ɛj ukkyr* 48 = *šitty ca sompylasar*
77 = *sel'cysar ɛj sel'cy* 98 = *šitty ca tōn*

Rules:

- Base-10 subtractive system.
- $10X = X$-*sar* 100 = *tōn*
- $10X + Y = X$-*sar ɛj* Y $Y = (1, 7)$
- $10X + Y = (10 - Y)$ *ca* $(X + 1)$-*sar* $Y = (8, 9)$

Solution for practice problem 9.13. Vambon

Body-part-based system, centred on 14.

Solution 9.13a a. $(17 + 11) \div 2 = 14$ b. $23 - 19 = 4$

Solution 9.13b c. *emuyop + emkumuk − emsanopkunip = emsilutop*
d. *silutop + kelop = emtakhem*

Solution 9.13c e. 'thumb' g. 'shoulder' i. 'little finger'
f. 'arm' h. 'forearm'

Solution 9.13d The prefix *em-* means ‘other / opposite’. Note: *em* ⟶ *e* / _ *m*.

Solution for practice problem 9.14. Alamblak

Solution 9.14a

a. 71	c. 42	e. 26	g. 78	i. 89
b. 31	d. 50	f. 73	h. 36	j. 52

Solution 9.14b k. 64 + 30 = 97 l. 7 + 12 = 19

Solution 9.14c 21 = *yima yohtti rpat*

48 = *yima hosfi tir yohtti hosfirpat*

83 = *yima hosfihosfi hosfirpat*

Rules:

- General structure: *yima X-i tir Y-i Z* = $20X + 5Y + Z$
- The digit 1 has two different forms: *rpat* (if it is Z), *yohtti* (for X or Y)
- Thus, multiplication is implied, while addition is marked by the suffix *-i*

Solution for practice problem 9.15. Chabu

Solution 9.15a

1.	*iŋki / ink*	10	*bab ef*
2.	*bab*	12	*bab ef eku bab*
3.	*bab eku iŋki*	15	*bab ef eku efi tʃumtʃum*
4.	*bab eku bab*	20	*ink ufe kor*
5.	*efi tʃumtʃum*	30	*ink ufe kor eku bab ef*
6.	*efi tʃumtʃum eku iŋki*		

Rules:

1. General structure:
   - 2[+X] = *bab* [*eku X*] ($X < 3$)
   - 5[+X] = *efi tʃumtʃum* [*eku X*] ($X < 5$)

- $10[+X]$ = *bab ef* [*eku X*] $(X < 10)$
- $20Y[+X]$ = *Y ufe kor* [*eku X*] $(X < 20)$

2. *Y* is written even if it is 1 (20 = *ink ufe kor*).

3. The digit 1 has two forms: *ink* (if it is a multiplier) or *iŋki* (if it is added). In task (a), 1 has two alternative spellings, since we do not know which one to choose if it appears as a single word.

Solution for practice problem 9.16. Tifal

Solution 9.16a (1) $9 + 10 = 19$

(2) $6 + 1 = 7$

(3) $5 \times 5 = 25$

Solution 9.16b (4) *tadang* + *miit* = *aleeb madi*

(5) *ataling madi* − *aleeb* = *bokob madi*

Rules: Body-part-based system, centred on 14. *X madi* = 28−*X*.

Solution for practice problem 9.17. Mansi

Solution 9.17a a. 405 b. 76 c. 819

Solution 9.17b 58 = *xōtlow nopɔl ńollow*

80 = *ńolsāt*

716 = *ńollowsātn xōtxujplow*

Rules: System based largely on overcounting. Tens are formed from units: adding the suffix *-low* (for 50, 60) or replacing the suffix *-low* with *-sāt* (for 80, 90).

$10 + X$ = *X-xujplow* $100X$ = *X-sāt*

$10X + Y = 10(X + 1)$ *nopɔl Y* $100X + Y = (X + 1)$*-sātn Y*

$90 + X$ = *ontɔlsāt X* $900 + X$ = *ontɔllowsāt X*

Further reading

Chrisomalis, Stephen. 2010. *Numerical notation: A comparative history*. Cambridge: Cambridge University Press.

Hurford, James R. 2011. *The linguistic theory of numerals*. Cambridge: Cambridge University Press.

Ifrah, George. 2000. *The universal history of numbers: From prehistory to the invention of the computer*. New York: John Wiley & Sons Inc.

Zaslavsky, Claudia. 1999. *Africa counts: Number and pattern in African cultures*. Chicago: Chicago Review Press.

10 Other types of problems

This last chapter focuses on problems that do not fit into the previous categories. For these problems there is not any well-defined theoretical background, since they can be considered to be quite unique, meaning that there will not be two or more problems based on the same idea. The only exceptions to this are the problems based on orientation systems and kinship systems, so this chapter will only focus on these two types. The theoretical background is relatively small, so we will instead present some general information that can be further applied to other problems.

10.1 Problems based on orientation systems

The purpose of orientation system problems is to identify the way in which a specific language expresses directions (relative positions of objects, such as in 'in front', 'behind', 'to the left', 'to the right', etc. or directions towards something: 'go ahead', 'turn left', 'turn right', etc.). These problems are typically easy to recognise since they usually contain the image of a map or a similar diagram or picture.

Typologically speaking, orientation systems can be classified into two categories: *absolute* or *relative* referential systems. In absolute referential systems, the directions are relative to one or more fixed points (e.g., cardinal directions or geographical locations). Probably the best-known language which uses an absolute orientation system is Guugu Yimithirr, spoken in the Hope Vale region, northern Queensland, Australia. This language uses cardinal directions (north, south, east, west) for every single context related to position or direction. Thus, the speakers of this language do not talk about their 'left' or 'right leg', but rather their 'west leg' (meaning the right leg, if the speaker faces south or the left leg if the speaker faces north), 'north leg', etc.

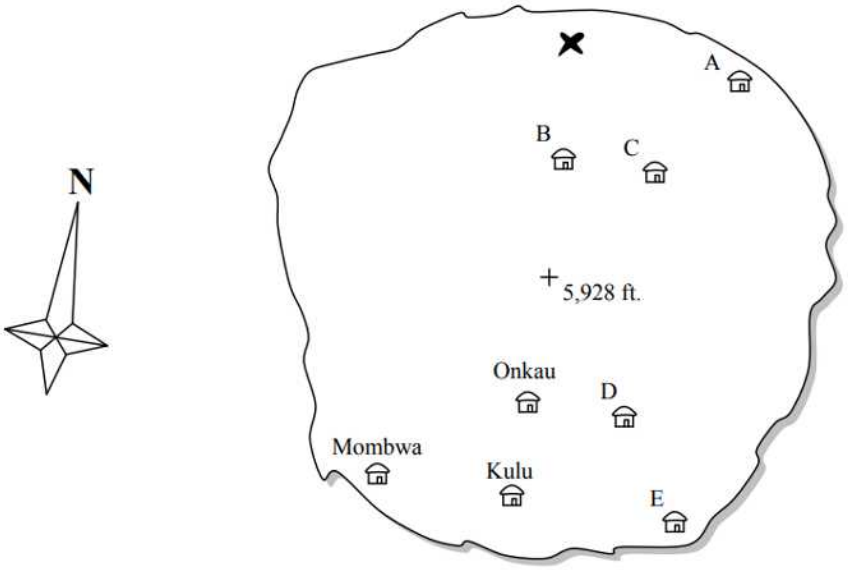
A special category of absolute referential systems, which is also the one most commonly appearing in linguistics problems, is that in which the reference system is based on the topography of the area. Usually, these words refer to directions such as 'upstream', 'downstream', 'uphill', 'downhill', 'towards the forest', 'towards the shore', etc.

Problem 10.1

Manam (Patrick Littell, NACLO 2008)

Manam Pile is a Malayo-Polynesian language spoken on Manam Island off the coast of Papua New Guinea. Manam is one of the most active volcanoes in the world. Below, a Manam islander describes the relative locations of the houses shown on the map.

1. *Onkau pera kana auta ieno, Kulu pera kana ilau ieno.*
2. *Mombwa pera kana ata ieno, Kulu pera kana awa ieno.*
3. *Tola pera kana auta ieno, Sala pera kana ilau ieno.*
4. *Sulung pera kana awa ieno, Tola pera kana ata ieno.*
5. *Sala pera kana awa ieno, Mombwa pera kana ata ieno.*
6. *Pita pera kana ilau ieno, Sulung pera kana auta ieno.*
7. *Sala pera kana awa ilau ieno, Onkau pera kana ata auta ieno.*
8. *Butokang pera kana awa auta ieno, Pita pera kana ata ilau ieno.*



Problem 10.1a Onkau's, Mombwa's and Kulu's houses have already been located on the map above. Who lives in the other five houses (A-E)?

Problem 10.1b Arongo is building his new house in the location marked with an X on the map. In three Manam Pile sentences like the ones on the previous page, describe the location of Arongo's house in relation to the three closest houses (A-C).

Solution

The first step is identifying the structure of the sentences. This is:

H_1 pera kana X_1 ieno, H_2 pera kana X_2 ieno.

where H_1 and H_2 refer to the names of the houses. Therefore, X_1 and X_2 must represent the words for directions. We notice that these can be: *ata*, *auta*, *awa*, *ilau*. Moreover, based on the examples 7 and 8, we notice that they can also combine with one another: *ata ilau*, *awa auta*, *awa ilau*, *ata auta*. Therefore, we deduce that there are two main directions or axes: *ata–awa* and *ilau–auta*, which can combine together (similarly to how the directions north–south and east–west can combine to form directions like north-east, north-west, etc.). Furthermore, we notice that X_1 is always the opposite of X_2 (similar to the English sentences 'A is due east, and B due west' or 'A is north-east, and B south-west').

Based on this information, we can rewrite the eight sentences above in a simpler way:

1. 'Onkau is to the' *auta* 'of Kulu.'
2. 'Mombwa is to the' *ata* 'of Kulu.'
3. 'Tola is to the' *auta* 'of Sala.'
4. 'Sulung is to the' *awa* 'of Tola.'
5. 'Sala is to the' *awa* 'of Mombwa.'
6. 'Pita is to the' *ilau* 'of Sulung.'
7. 'Sala is to the' *awa ilau* 'of Onkau.'
8. 'Butokang is to the' *awa auta* 'of Pita.'

The first hypothesis that comes to mind is that indeed, *ata*, *awa*, *ilau*, and *auta* represent the 'north', 'south', 'east', and 'west'. Thus, based on sentence 1, we deduce that *auta* = 'north', and based on 2 we deduce that *ata* = 'west'. Moreover, knowing already that the two main directions are *ata* – *awa* and *ilau* – *auta*, we can also deduce the other two cardinal points, mainly *awa* = 'east' and *ilau* = 'south'.

Thus, from sentence 7, we find out that Sala's house is to the south-east of Onkau's, so Sala's house can only be E, which is also confirmed by sentence 5, which mentions that Sala's house is to the east of Mombwa's. Next, from sentence 3, we deduce that Tola's house is to the north of Sala's, so Tola's house can only be A or C, but, since sentence 4 says that Sulang's house is to the east of Tola's, Tola's house needs to be C and Sulung's A (if Tola's house were A, there is no other house to the east of it). From sentence 6, Pita's house is to the south of Sulung's, so it is most likely D. Finally, according to sentence 8, Butokang's house is to the north-east of Pita's, but, at the same time, we know that Butokang's house must be B, since it is the only house left. Thus, we reach a contradiction which points to the fact that, most likely, the four words are not cardinal points.

If we carefully read the introduction, we learn that the island is in fact a volcano. Taking this into account, we can interpret the first sentence to mean that Onkau's house is *higher* up the mountain than Kulu's. Since we excluded the possibility that this word (*auta*) means 'north', perhaps it means 'towards the top' or 'further away from the shore'. Thus, its pair *ilau* will mean 'towards the shore'. We can conclude that the first directional axis is based on altitude (from the sea towards the top of the volcano).

From sentence 2, we understand that we need to find one more direction which points laterally on the map and we already figured out that it cannot be 'east' and 'west'. Thus, perhaps the words actually represent 'left' and 'right', when looking towards the top of the volcano. In other words, we can imagine the island to be a clock whose centre is the top of the volcano and the relative positions of the houses are not described by means of east–west but rather clockwise–anticlockwise. Thus, from sentence 2 we deduce that *ata* = 'clockwise' ('to the left, facing the volcano') and, consequently, *awa* = 'anticlockwise' ('to the right, facing the volcano').

Consequently, in sentence 7 we are told that Sala's house is towards the shore and to the right of Onkau's house, so, again, Sala's house can only correspond to E (since D is approximately at the same altitude with Onkau's, so it is not towards the shore). This is also confirmed by sentence 5 which states that Sala's house is anticlockwise with respect to Mombwa's (nothing about the relative vertical position is mentioned, so it is on the same level).

From sentence 3, Tola's house is higher than Sala's, so Tola's house = D. Next, from sentence 4, Sulung's house is anticlockwise of Tola's (to the right, facing the volcano), so Sulung's house = C; from sentence 6, Pita's house is lower than Sulung's (towards the shore), so Pita's house = A. Finally, we are left with B which must be Butokang's house, as confirmed by sentence 8, which states that Butokang's house is higher and to the right of Pita's.

Now we can solve task (b). The house marked with X will be 'towards the shore' with respect to B, 'to the right' of A and 'towards the shore and to the right' of C.

Therefore, we can solve all tasks.

Solution 10.1a A = Pita, B = Butokang, C = Sulung, D = Tola, E = Sala

Solution 10.1b *Arongo pera kana awa ieno, Pita pera kana ata ieno.*
Arongo pera kana awa ilau ieno, Sulung pera kana ata auta ieno.
Arongo pera kana ilau ieno, Butokang pera kana auta ieno.

As previously mentioned, with this type of problem, we need to pay attention to the topography of the area and consider the landforms around and how they can be used to indicate directions. In this case, since the language is spoken on an island, it is very plausible that the sea is one of the points of reference. Thus, each point can be closer to the sea or more towards the centre of the island (in this case the middle of the island is actually a volcano so equivalent directions can also be uphill and downhill). Since in the introduction we are told that the island has an active volcano, we expect it to be relevant in solving the problem. Considering that a volcano can be assimilated to a cone, we can consider the second direction to be circular (clockwise or anticlockwise).

10.2 Kinship problems

The core purpose of this type of problem is that different languages use different types of terms to refer to different family relations. This type of problem is very easy to recognise since it will refer to a family tree. In the corpus, some sentences are given in which the relations of some family members with other members of the family are described in the target language.

There are six main types of kinship systems, but before discussing them, it is important to get acquainted with the ways of representing a kinship diagram (a family tree). Figure 10.1 shows a basic kinship diagram. Triangles represent men and circles represent women. A double line (an equals sign, =) denotes marriage

and simple lines represent blood relations. Figure 10.1 tells us that 1 and 3 are two women (they are represented by circles) who are sisters (marked by a simple line), and each of them is married (1 married to 2 and 3 to 4). The family formed by 1 and 2 has two children (vertical line) – a girl (5) and a boy (6), just like the family formed by 3 and 4. Thus, based on this diagram, we can characterise each person with respect to the other, e.g., 6 is the son of 1 and the nephew of 4, 2 is the husband of 1 and the brother-in-law of 3, etc. Generally, linguistics problems will also include a legend which explains the meaning of each symbol, but the notations mentioned above are the standard ones. Moreover, from case to case, some problems might also include the age of the person, since some languages differentiate certain kinship terms based on age (e.g., in Chinese there are different words for 'younger sister' – 妹妹 *mèimei*, 'older sister' – 姐姐 *jiějie*, 'younger brother' – 弟弟 *dìdi*, and 'older brother' – 哥哥 *gēge*).

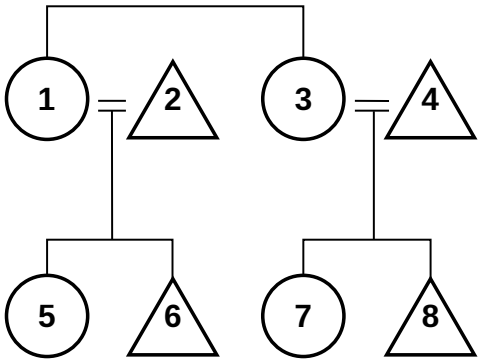


Figure 10.1: A typical kinship diagram.

In 1949 the anthropologist G.P. Murdock identified six basic patterns of kinship terminology systems, which are now generally accepted. Of course, certain languages can display systems different from them, but the six types below are the most common ones.

In order to represent these systems, a diagram like the one in Figure 10.2 is used. A shaded triangle or circle, if included, represents the person from whose point of view the tree is presented, called the *Ego* in genealogy.

The simplest kinship system is called *Hawaiian*, which has only three basic kinship terms: 'mother', 'father' and 'sibling'. Thus, persons A, D and E are all called 'father', persons B, C, F are all called 'mother' and all siblings and cousins are called the same ('sibling'). In this case (the Hawaiian system), we write A = D = E, B = C = F and G = H = I = J = K = L = M = N = O = P. This representation does not take into account the variations based on age so, solely based on this

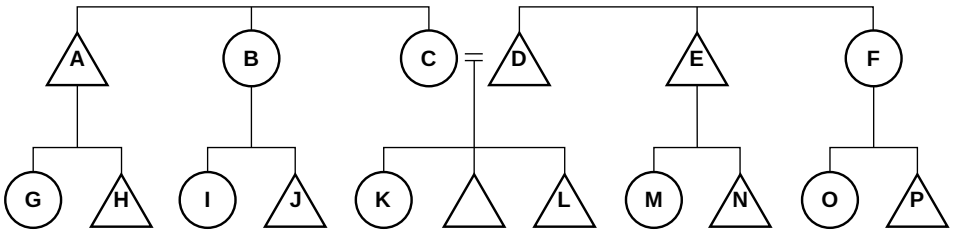


Figure 10.2: Template kinship diagram for describing the different types of kinship systems.

description, we cannot know that in the Hawaiian language (which displays the Hawaiian kinship term) there is a difference between younger and older sibling. Generally, in linguistics problems, if the age does not play a relevant role, it will not be included in the diagram (i.e., if a diagram includes the age of the persons, age will certainly play a role).

The next system is called *Eskimo*[1] kinship (or *Inuit*) one, which is also the system used in English. In this system we differentiate C ('mother'), D ('father'), B = F ('aunt'), A = E ('uncle'), K = L ('siblings') and G = H = I = J = M = N = O = P ('cousins'). Although the six main kinship systems we describe here represent general patterns, these patterns can be slightly modified from one language to another. A case worth mentioning is that of Romanian which, although it is considered to use an Eskimo kinship term, it has two different words for 'cousin' based on the gender, and differentiates between G = I = M = O (*verișoară*, 'female cousin') and H = J = N = P (*verișor*, 'male cousin'). This is true of many European languages, and in some (e.g. German) the difference *Kusine–Vetter* is not just a gender suffix.

The next system we talk about is called *Sudanese* kinship, one example of which is Turkish, which we use here to illustrate. In this type of system there is a separate term for each of the persons A-F (*dayı* = 'mother's brother', *amca* = 'father's brother', *teyze* = 'mother's sister', *hala* = 'father's sister', *anne* = 'mother', *baba* = 'father') and a different term for each pair of cousins: I and J are called 'maternal parallel cousins' – *maternal* refers to the fact that they are on the mother's side. The term *parallel* is used because the blood relation is from the same-sex persons (i.e., same-sex siblings) - mother and mother's sister (two women). In the same way, O and P are called 'paternal parallel cousins', from the father's side and, more exactly, from the father's brother (parallel since it is the father's

[1]The term dates from 1949 and is still used even though the word *Eskimo* is now disdained as being derogatory.

brother (same sex as the father), not sister). The other two categories are called 'maternal cross cousins' and 'paternal cross cousins', where *cross* refers to the fact that the blood relation is of opposite-sex persons (mother's brother and father's sister).

This distinction between *parallel* and *cross* cousins is rather common and so relevant that in the next kinship system, called *Iroquois* kinship, B = C ('mother') and D = E ('father'). Here, the same-sex siblings of the parents (i.e., mother's sister and father's brother) are also considered to be 'parents'. On the other hand, the opposite-sex siblings of the parents are those called 'uncle' (A) and 'aunt' (F). For this reason, G = H = O = P ('cousins') – since they are the children of the aunt and uncle –, but I = J = K = L = M = N ('siblings') – since they are the children of the mother and father. The next two systems are derived from this system.

The *Crow* kinship system starts from the Iroquois system, with the only change occurring for the persons O and P (the children of the father's sister). They are called 'aunt' (if it is a girl, so O) or 'father' (P). Thus, in this system, D = E = P ('father') and F = O ('aunt'). The rest of the persons follow the Iroquois system (A = 'uncle', B = C = 'mother', G = H = 'cousins', I = J = K = L = M = N = 'siblings').

The last system, called *Omaha* kinship, is the opposite of Crow kinship. In this system, a special role is reserved for the children of the mother's brother (instead of father's sister, as in the Crow system). Like the Crow system, the same-sex child is also called 'uncle' or 'aunt', depending on their (and Ego's) sex (same sex as Ego), while the opposite-sex child is called 'mother' or 'father'. Therefore, A = H = 'uncle', B = C = G = 'mother', D = E = 'father', F = 'aunt', I = J = K = L = M = N = 'sibling' and O = P = 'cousins'.

Although the systems are each named after a language, there are other languages which follow each system: for example English has 'Eskimo kinship', Bulgarian has 'Sudanese kinship'.

A comparative representation of these six types of system is represented below:

<table>
<tr><th>Person</th><th>Hawaiian</th><th>Eskimo</th><th>Sudanese</th><th>Iroquois</th><th>Crow</th><th>Omaha</th></tr>
<tr><td>A</td><td>'father'</td><td>'uncle'</td><td>'mother's brother'</td><td colspan="3">'uncle'</td></tr>
<tr><td>B</td><td rowspan="2">'mother'</td><td>'aunt'</td><td>'mother's sister'</td><td colspan="3" rowspan="2">'mother'</td></tr>
<tr><td>C</td><td colspan="2">'mother'</td></tr>
<tr><td>D</td><td rowspan="2">'father'</td><td colspan="2">'father'</td><td colspan="3" rowspan="2">'father'</td></tr>
<tr><td>E</td><td>'uncle'</td><td>'father's brother'</td></tr>
<tr><td>F</td><td>'mother'</td><td>'aunt'</td><td>'father's sister'</td><td colspan="3">'aunt'</td></tr>
</table>

<table>
<tr><td>Person</td><td>Hawaiian</td><td>Eskimo</td><td>Sudanese</td><td>Iroquois</td><td>Crow</td><td>Omaha</td></tr>
<tr><td>G</td><td rowspan="10">‘sibling’</td><td rowspan="3">‘cousin’</td><td rowspan="2">‘maternal CC’[2]</td><td colspan="2" rowspan="2">‘cousin’</td><td>‘mother’</td></tr>
<tr><td>H</td><td>‘uncle’</td></tr>
<tr><td>I</td><td rowspan="2">‘maternal PC’[2]</td><td colspan="3" rowspan="4">‘sibling’</td></tr>
<tr><td>J</td></tr>
<tr><td>K</td><td colspan="2" rowspan="2">‘sibling’</td></tr>
<tr><td>L</td></tr>
<tr><td>M</td><td rowspan="4">‘cousin’</td><td rowspan="2">‘paternal PC’[2]</td></tr>
<tr><td>N</td></tr>
<tr><td>O</td><td rowspan="2">‘paternal CC’[2]</td><td rowspan="2">‘cousin’</td><td>‘aunt’</td><td rowspan="2">‘cousin’</td></tr>
<tr><td>P</td><td>‘father’</td></tr>
</table>

Problem 10.2

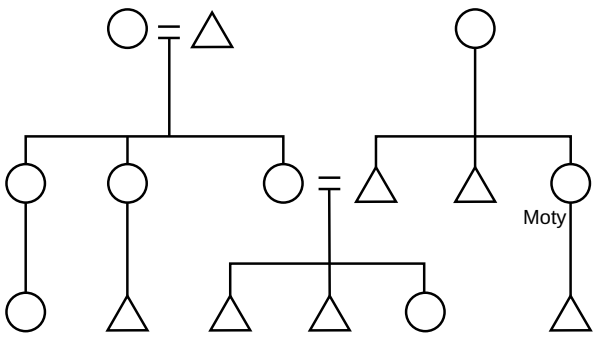
Arawak (Michał Boroń, original)

Below you can find part of an Arawak family tree. Three family members – two men and a woman (not necessarily in this order) – describe their family:

1. *De to Fatan. Onikhan to dajo. Mithakotoan ken Kolhen to dakhithonon. Tholhady to dato. Ematonoa to dathi.*

2. *De to Sobole. Bokoa to dakhithi. Kolhen ken Moty to dajonon. Balhose ken Konoko to dathinon. Onikhan ken Ylhydaba to dakythynon. Fatan ken Mithakotoan to dajaboathonon. Sareke to dajorodatho. Ematonoa to dadokothi.*

3. *De to Balhose. Ylhydaba to dajo. Kolhen to daretho. Sobole ken Bokoa to daithinon. Konoko to dakhithi. Moty to dajorodatho. Mithakotoan ken Fatan to darebiathonon. Sareke to dato.*



[2]CC = ‘cross cousin’, PC = ‘parallel cousin’.

Problem 10.2a Supply the tree with names. If multiple options are possible, provide them all.

Problem 10.2b Three more people (from the same family) describe their family:

4. *De to Ajonym. Fatan ken Kolhen to* __(1)__. *Onikhan to dakythy. Mithakotoan to dajo.* __(2)__ *to dadokothi.*
5. *De to* __(3)__. *Balhose ken Konoko to daithinon. Holholho ken Sobole ken* __(4)__ *to dalykynthinon. Moty to dato.* __(5)__ *to dalykyntho.*
6. *De to Kolhen.* __(6)__ *to dajo.* __(7)__ *to darethi. Ematonoa to* __(8)__. *Sareke to dato. Sobole ken Bokoa to* __(9)__. *Konoko to* __(10)__.

Fill each gap with *exactly one word.*

In the diagram above, triangles represent men and circles represent women. Horizontal lines represent siblings and vertical lines children. The equals sign denotes marriage.

Solution

Firstly, we look at the structure of the sentences. Each sentence, except for the first, follows the pattern Name *to* X. So we know that X represents the kinship term. We deduce that *ko* = 'is/are', and the first sentence most likely means 'I am X', therefore *de* = 'I'. Moreover, we notice that if in the rest of the sentences there are more names co-occurring, they are separated by *ken*, so this word most likely means 'and'. Last but not least, we notice that every time that a sentence contains more than one name, the kinship term ends in *non*, so we deduce that the suffix *-non* marks the plural.

Based on this we can make a table in which we show the relations between the persons mentioned in the data:

	Fatan	Sobole	Balhose
Fatan		*dajaboatho*	*darebiatho*
Onikhan	*dajo*	*dakythy*	

	Fatan	Sobole	Balhose
Mithakotoan	*dakhitho*	*dajaboatho*	*darebiatho*
Kolhen	*dakhitho*	*dajo*	*daretho*
Tholhady	*dato*		
Ematonoa	*dathi*	*dadokothi*	
Sobole			*daithi*
Bokoa		*dakhithi*	*daithi*
Moty		*dajo*	*dajorodatho*
Balhose		*dathi*	
Konoko		*dathi*	*dakhithi*
Ylhydaba		*dakythy*	*dajo*
Sareke		*dajorodatho*	*dato*

Moreover, we notice that two names are missing: *Ajonym* and *Holholho* (both of them found in task (b)).

Generally, the first step in solving this type of problem, once the table above is made, is assuming that if two persons have the same relation with a third, then those two persons belong to the same generation (e.g., if A and B are *m* to person X – or X is *m* to persons A and B – then, most likely, A and B are part of the same generation, i.e., they are on the same level of the tree). In this case, we notice from the diagram that we have three generations: the top one, which has three members, the middle one (siz members) and the bottom one (six members).

Starting with Fatan, we notice that they are the same relation (*dakhitho*) to both Mithakotoan and Kolhen, so we can assume that these two belong to the same generation. Similarly, we deduce that Fatan and Mithakotoan belong to the same generation (since they both are *dajaboatho* to Sobole). Knowing that Mithakotoan and Kolhen as well as Mithakotoan and Fatan belong to the same generation, we can deduce that all three of them are part of the same generation.

Using the same thought process, we get the following pairs of persons belonging to the same generation: Kolhen – Moty (they are *dajo* for Sobole), Balhose – Konoko (*dathi* for Sobole), Fatan – Mithakotoan (*darebiatho* for Balhose), Sobole – Bokoa (*daithi* for Balhose), Onikhan – Ylhydaba (*dakythy* for Sobole).

Collecting all the information, we get:

Group 1: Mithakotoan – Kolhen – Fatan – Moty

Group 2: Balhose – Konoko

Group 3: Sobole – Bokoa

Group 4: Onikhan – Ylhydaba

We need to take into account that two (or more) of these groups can combine to form a generation. Moreover, we know for sure that Group 1 is part of the middle generation since it contains Moty.

Furthermore, we can use task (b) to see if there are any other names that cooccur. In example 5, sentence 3, we notice that Holholho and Sobole appear together. At first sight, it can seem insignificant since we have no information about Holholho in the corpus. However this information is actually extremely important. If we add Holholho to Group 3, then this group will contain three persons. Since Group 1 has four persons and Group 3 has three persons, we certainly know that these two cannot represent the same generation (since there is no generation with seven members), so we deduce that Group 1 and Group 3 belong to different generations.

Let us look now at the kinship term *dajo*. It appears between Sobole (Group 3) and Kolhen (Group 1), so we know that this term spans across generations. The same relation appears between Fatan (Group 1) and Onikhan (Group 4). Based on this, we deduce that Group 4 represents the last separate generation (it does not combine with either Group 1 or Group 3). This might not seem obvious at first, but we know that the relation between a person from Group 3 and a person from Group 1 is the same as the relation between a person from Group 1 and one from Group 4 (we can write this succinctly as G3–G1 = G1–G4). Moreover, we already know that G3 and G1 belong to different generations, so, if G4 belonged to the same generation as G1, then the second relation (the word *dajo* which refers to Kolhen from G1 and Onikhan from G4) would be within the same generation, while this is not true for the first relation (the word *dajo* which refers to G3–G1). If G4 was part of the same generation as G3, then we would have a reciprocal relation (the relation of X to Y uses the same name as the relation of Y to X – but across different generations), which is highly unlikely. The only option left is, therefore, that G4 belongs to a separate generation.

Let us focus now on the relation *dajo*, which appears between Onikhan (G4) and Fatan (G1), as well as between Ylhydaba (G4) and Balhose (G2) (i.e., G4–G1 = G4–G2). From here, we deduce that G2 and G1 belong to the same generation.

Now we can classify the three generations:

Generation A (Gen. A): Mithakotoan, Kolhen, Fatan, Moty, Balhose, Konoko

Generation B (Gen. B): Sobole, Bokoa, Holholho

Generation C (Gen. C): Onikhan, Ylhydaba

For simplicity, we can remake the table, but this time replacing the name with the generation.

	A	B	A
A		*dajaboatho*	*darebiatho*
C	*dajo*	*dakythy*	
A	*dakhitho*	*dajaboatho*	*darebiatho*
A	*dakhitho*	*dajo*	*daretho*
Tholhady	*dato*		
Ematonoa	*dathi*	*dadokothi*	
B			*daithi*
B		*dakhithi*	*daithi*
A		*dajo*	*dajorodatho*
A		*dathi*	
A		*dathi*	*dakhithi*
C		*dakythy*	*dajo*
Sareke		*dajorodatho*	*dato*

The first observation is that both Tholhady and Sareke have the relation *dato* with respect to Gen. A, so they both belong to the same generation (B or C, since Gen. A already has six members, so it is complete).

The term *dajorodatho* is established between two persons from Gen. A, so it is a kind of relation established within the same generation. Since Sareke uses the same relation with a person from Gen. B, we deduce that Sareke also belongs to Gen. B.

For Ematonoa, we look at the term *dathi*. We have the equivalence: A–B = Ematonoa–A. Since we said we exclude transitive relations, Ematonoa must belong to Gen. C. Thus, the generations become:

Gen. A: Mithakotoan, Kolhen, Fatan, Moty, Balhose, Konoko

Gen. B: Sobole, Bokoa, Holholho, Tholhady, Sareke

Gen. C: Onikhan, Ylhydaba, Ematonoa

There is only one person unassigned, Ajonym, who surely belongs to Gen. B (since it is the only incomplete generation in terms of the number of members). Moreover, we know the correspondence between the letters A–C and the top/

middle/bottom generations. Gen. C is surely the top one, since it has only three members, while Gen. A is the middle one since it contains Moty. Therefore, Gen. B remains and must be the bottom one. The final generations are:

Top generation: Onikhan, Ylhydaba, Ematonoa

Middle generation: Mithakotoan, Kolhen, Fatan, Moty, Balhose, Konoko

Bottom generation: Sobole, Bokoa, Holholho, Tholhady, Sareke, Ajonym

We expect that the kinship terms established between the middle and the top generation to be 'mother' and 'father' (the other option is something like 'wife's sister's mother', which is rather too complex). Thus, we notice that for Fatan, Onikhan and Ematonoa are *dathi* and *dajo*, respectively; we can assume that they represent 'mother' and 'father'. Since there is only one married couple in the top generation, it must be Onikhan-Ematonoa. To differentiate between 'mother' and 'father', we notice that *dajo* also appears with Ylhydaba, who belongs to the top generation. Since the only person left in the top generation is a woman, we deduce that *dajo* = 'mother' and *dathi* = 'father'. Moreover, we can assign all the names to the top generation (from left to right: Onikhan, Ematonoa, Ylhydaba).

We notice that for Sobole, both Onikhan and Ylhydaba are *dakythy*. Since Sobole is part of the bottom generation, we can easily deduce that *dakythy* = 'grandmother'. Moreover, we deduce that Sobole must be one of the children in the middle of the tree (the group of three siblings), since they are the only persons whose grandparents are both Onikhan and Ylhydaba.

Let us now look at Sobole's statements. Since we already know the words for 'mother' and 'father', we notice that in the generation above, Sobole has two persons who they refer to as *dajo* 'mother' (Kolhen and Moty), two who they call *dathi* 'father' (Konoko and Balhose) and two who they refer to as *dajaboatho* (Fatan and Mithakotoan). Among the two persons called 'mother', we can surely expect that one of them but not both is their biological mother. Since we already know where Moty is placed, we deduce that Kolhen is the (biological) mother of Sobole (the married woman). Moreover, certainly the two men will be those called 'father' (Konoko and Balhose), and the two remaining women on the right will be Fatan and Mithakotoan. Furthermore, the relation between Fatan and Mithakotoan is *dakhitho*, so *dakhitho* = 'sister (of a woman)'. Similarly, the relation between Konoko and Balhose is *dakhithi*, so *dakhithi* = 'brother (of a man)'. Next, we notice the similarity between the two terms, which differ only in terms of their last vowel (*-i* or *-o*), which makes us assume that the difference in gender is marked by the last vowel (thus, the stem *dakhith-* means 'same-sex sibling'; if

it ends in the vowel *-i*, meaning masculine, it will refer to same-sex brothers i.e., brother of a male, while if it receives the suffix *-o*, it means same-sex sister, i.e., a woman's sister).

We also notice that Bokoa and Sobole are *dakhithi*, so the two are siblings. Therefore, Sobole and Bokoa are the two men from the bottom generation. Moreover, the two men are *daithi* for Balhose, so *daithi* = 'son'. We can assume that Balhose is the biological father of the two children (unlike Konoko). This is also confirmed by the fact that Kolhen is *daretho* for Balhose, and this kinship term does not occur again, so it means that *daretho* = 'wife'.

Sareke is *dato* for Balhose, and Sareke is part of the bottom generation. The only person, in relation with Balhose, which we have not described, is their daughter, so Sareke is the daughter of Balhose (the sister of Sobole and Bokoa), and *dato* = 'daughter'.

Furthermore, since Tholhady is the daughter of Fatan (knowing that Fatan and Mithakotoan are the two women in the middle generation, and only one of them has a daughter), we can determine all the correspondences of the middle generation. The names are, from left to right: Fatan, Mithakotoan, Kolhen, Balhose, Konoko, Moty.

The only persons left to be assigned to the diagram are Ajomyn and Holholho. Checking task (b), sentence 4, Ajonym says that Onikhan is their *dakythy* 'grandmother', so Ajonym is the son of Mithakotoan, while Holholho is the son of Moty.

The only thing left undetermined is the difference between Sobole and Bokoa. Since we have no information which could help differentiate between the two (and since task (a) says that there are multiple options in some cases), we deduce that the correct assignment of names is as presented in Figure 10.3 (with the mention that Sobole and Bokoa can be swapped).

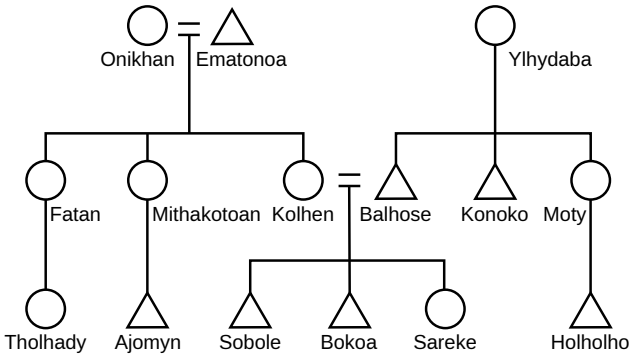


Figure 10.3: The name assignment for Problem 10.2.

Based on this, we can easily deduce the remaining kinship terms, and make a table with all the words we know (since we saw that the masculine-feminine pairs are similar, we organise them in different columns):

Kinship term	Male	Female
'child'	*daithi*	*dato*
'parent' or 'father's sibling'	*dathi*	*dajo*
'grandparent'	*dadokothi*	*dakythy*
'grandchild'	*dalykynthi*	*dalykyntho*
'same-sex sibling'	*dakhithi*	*dakhitho*
'opposite-sex sibling'		*dajorodatho*
'spouse'	*darethi*	*daretho*
'spouse's same-sex sibling'	*darebiathi*	*darebiatho*
'mother's sibling'		*dajaboatho*

We notice again that some kinship terms can switch between masculine and feminine by changing the last vowel from *-i* (masculine) to *-o* (feminine).

Based on this, we can solve task (b):

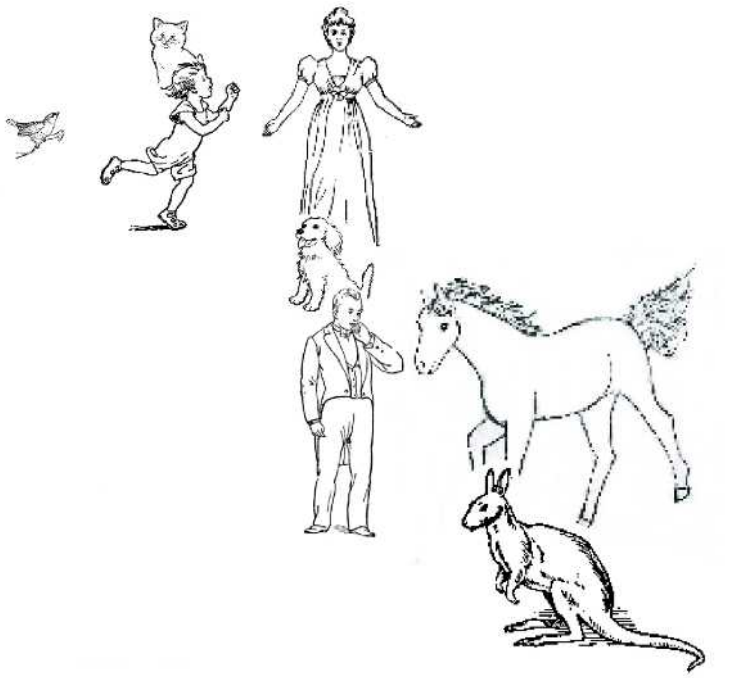
(1) *dajaboathonon*
(2) *Ematonoa*
(3) *Ylhydaba*
(4) *Bokoa*
(5) *Sareke*
(6) *Onikhan*
(7) *Balhose*
(8) *dathi*
(9) *daithinon*
(10) *darebiathi*

10.3 Practice problems

Problem 10.3

Bardi (Catherine Sheard, UKLO 2012)

In the picture below, note that both you and the speaker are facing the paper. The bird is to the left of everything else and the kangaroo is to the right of everything else. The cat is behind everything else and the kangaroo is in front of everything else.



Here are some Bardi sentences describing the scene:

i) *Aamba bornkony yaawardon.*

ii) *Baawa joorroonggony garrabalgoon.*

iii) *Boorroo alaboor yaawardon.*

iv) *Iila alaboor ooranygoon.*

v) *Iila baybirrony aambon.*

vi) *Minyaw baybirrony baawon.*

vii) *Oorany joorroonggony baawon.*

viii) *Yaawarda bornkony aambon.*

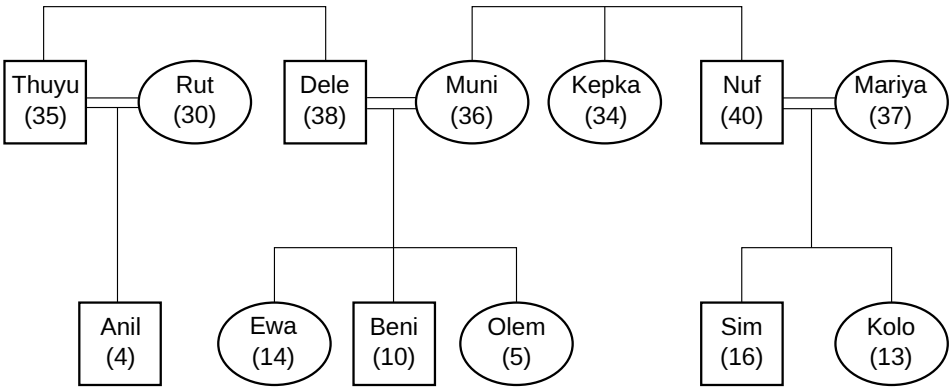
Problem 10.3a Based on these, determine the following correspondences:

1.	*Aarlgoodony*	A.	‘bird’
2.	*Aamba*	B.	‘child’
3.	*Alaboor*	C.	‘cat’
4.	*Baawa*	D.	‘dog’
5.	*Baybirrony*	E.	‘horse’
6.	*Boorroo*	F.	‘kangaroo’
7.	*Bornkony*	G.	‘man’
8.	*Garrabal*	H.	‘woman’
9.	*Iila*	I.	‘next to’
10.	*Joorroonggony*	J.	‘behind’
11.	*Minyaw*	K.	‘in front of’
12.	*Oorany*	L.	‘to the left of’
13.	*Yaawarda*	M.	‘to the right of’

Problem 10.4

Kharia (Barbora Dohnalová, ČLO 2021)

Below is shown the kinship tree of a Kharia family in which circles represent women and squares represent men. The age of each person is written under their name.



Each member of this family says something about their family in Kharia:

a. *Bhaiiŋaʔ ɲimi Thuyu.*

b. *Didiiŋaʔ ɲimi Muni. Ɗonkuiiŋaʔ ɲimi Mariya.*

c. *Sowiŋaʔ ɲimi Nuh.*

d. *Didikiiŋaʔ ɲimiki Olem oɖoyoʔ no Ewa.*

e. *Konon bahiniŋaʔ ɲimi Olem. Didikiiŋaʔ ɲimiki Ewa oɖoyoʔ no Kolo.*

f. *Ɗonkuiiŋaʔ ɲimi Muni. Dad=iŋaʔ ɲimi Dele.*

g. *Sowiŋaʔ ɲimi Thuyu. Beʔʧiŋaʔ ɲimi Anil.*

h. *Dadakiiŋaʔ ɲimiki Beni oɖoyoʔ no Sim.*

i. *Beʔʧiŋaʔ ɲimi Beni. Ɗonkuiiŋaʔ ɲimi Mariya.*

j. *Bhaikiiŋaʔ ɲimiki Beni oɖoyoʔ no Anil. Apaiŋaʔ ɲimi Dele.*

k. *Konon bahinkiiŋaʔ ɲimiki Kolo, Ewa oɖoyoʔ no Olem.*

l. *Konon bahinkiiŋaʔ ɲimiki Muni oɖoyoʔ no Kepka.*

m. *Apaiŋaʔ ɲimi Nuh. Dad=iŋaʔ ɲimi Sim.*

Problem 10.4a Assign each of the sentences above to the person who uttered it.

Problem 10.4b Fill in the blanks:

i. Muni: "__(1)__*ɲimi Dele.*"

ii. Kepka: "__(2)__*ɲimi Nuh.*"

iii. Ewa: "*Konon bahinkiiŋaʔ ɲimiki* __(3)__."

iv. Anil: "*Dad=iŋaʔ ɲimi* __(4)__."

v. Beni: "__(5)__*ɲimi Dele.*"

Problem 10.4c A few years later, Kepka has a son named Caitu. Fill in the blanks:

vi. Sim: "__(6)__*ɲimi Caitu.*"

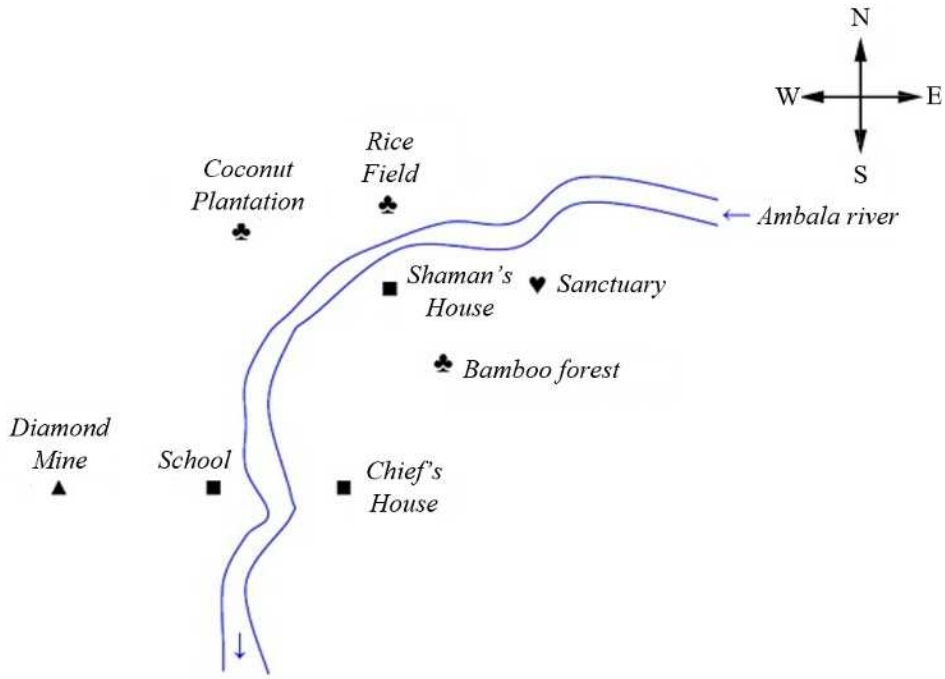
vii. Kepka: "__(7)__*ɲimi Caitu.*"

viii. Caitu: "__(8)__*ɲimiki Ewa, Olem oɖoyoʔ no Kolo.*"

Problem 10.5

Embaloh (Ksenia Gilyarova, MSK 2006)

A tourist travels to a village on the course of the river Benoit Martinus Ambala (Kalimantan Island, Indonesia), in order to learn the Embaloh language. He will live at the Chief's House (see map). On the first day, the Chief takes his guest outside, points towards the north, south, east and west and says "urait, kalaut, anait, suali". The tourist wrote down in his own dictionary: *urait* = 'north', *kalaut* = 'south', *anait* = 'east', *suali* = 'west'.



The next day, the tourist wanted to visit the Sanctuary – the place where all the important tribal ceremonies take place. He took his dictionary and compass, but he left his map at home. Exiting the Chief's House, he started walking north and reached the Shaman's House. He asked the Shaman: "How can I get to the Sanctuary?" "Keep going *urait*," replied the Shaman. "So, I should head north" thought the tourist, checking his dictionary. He crossed the river, but then he got lost in a Rice Field, so he decided to return to the Shaman. A plantation worker guided him: "Towards the Shaman's House, go *suali*." "That means west?! Weird!" thought the tourist, but nevertheless he headed west. However, the river didn't come up and the rice fields were slowly replaced by coconut plantations and the tourist realised he was completely lost. "Whatever it is," a worker on the coconut

plantation started comforting the tourist, "keep going *kalaut*. You will get to the School and the teacher will explain everything."

Checking his dictionary, he headed south and he indeed reached the school. "Chief's House *anait*?" the tourist asked the teacher. "No, *anait* Diamond Mine. Chief's House *suali*" he replied. The tourist, humbled, headed west and found himself at the Diamond Mine. Extremely angry, he asked: "How can I finally reach the Chief's House or at least the School?" "Chief's House *suali*, but School *andoor*." Unfortunately, this last word does not appear in the tourist's dictionary.

Problem 10.5a Explain why the tourist got lost and explain how the orientation system of this tribe works, as well as what each direction means.

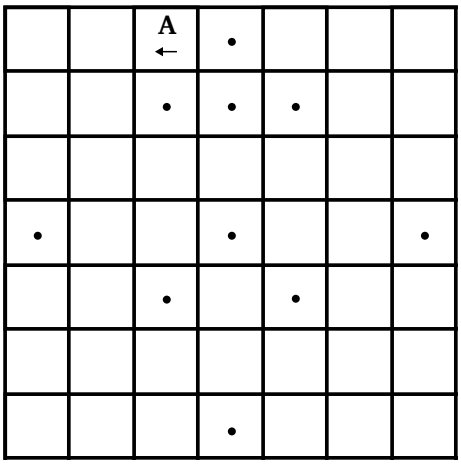
Problem 10.5b Describe in Embolah the directions:

1. from the Sanctuary to the Shaman's House;
2. from the Bamboo Forest to the Shaman's House;
3. from the Shaman's House to the Rice Field.

Problem 10.6

Hungarian (Adam Hesterberg, UKLO 2014)

The picture below represents a field divided into 49 squares (7x7), aligned with north at the top and east on the right. In some of the squares there are rocks, indicated by a black circle ●.



There are four Hungarian friends - A, B, C and D – standing in the field, each in a different square not containing a rock, and each facing in one of the four cardinal directions (north, south, east, west). Each person makes some statements describing the position of the rocks. For instance, A's first statement means 'To the east (behind me) there is one stone'. References to directions are to be understood as describing a single line in the field: 'due east', 'directly behind me', and so on. No directions describe a more complex spatial relationship.

A says: *Délre két kő van.*
Keletere (mögöttem) egy kő van.
Jobbra nincs kő.

B says: *Délre (balra) nincs kő.*
Északra egy kő van.
Mögöttem két kő van.

C says: *Északra (előttem) nincs kő.*
Nyugatra egy kő van.
Jobbra két kő van.

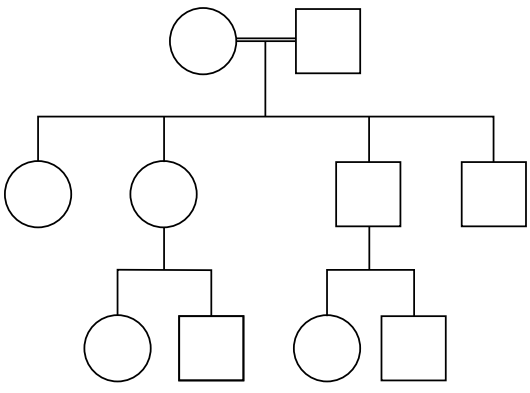
D says: *Nyugatra (jobbra) két kő van.*
Északra egy kő van.
Balra nincs kő.

Problem 10.6a Which square is occupied by each of B, C and D? Draw an arrow (like the one under A) to show the direction they are facing.

Problem 10.7

Tabaq (Dan-Mircea Mirea, RoLO 2017)

Two linguists, Dr. David Lovelang and Dr. Matt Hateword were studying the language spoken by the Tabaq people in South Sudan. While Dr. Lovelang was focused on the phonology of the language, Dr. Hateword was concerned with the way their kinship system works. To further his studies, he chose ten members of a big family and asked them to say a name first and then use the kinship term they'd use to describe that person. He carefully wrote everything down in a notebook and drew the following diagram:



Shortly after, Dr. Hateword had to give up on his research, leaving all his scribbles as well as the blank diagram to his colleague. At first, Dr. Lovelang had no clue as to how he could fill in the diagram, but, after a closer look at the information in the notebook, he managed to fill it in.

Here are the scribbles from the notebook:

Rowa:	*íɻ̱è ɻ̱éèr* *Minni, ɻ̱ɔ̀ɔ̀ɖ̱ɔ̀* *Nadwah, íɻ̱è-n-ɻ̱ɔ̀ɔ̀ɖ̱ɔ̀-ɻ̱éèr*	Sadiq:	*Salva, ɻ̱ɔ̀ɔ̀ɖ̱ɔ̀* *Abir, màà* *Sihan, íɻ̱è ɻ̱éèr*
Kuwa:	*Salva, ɔ́ɻ̱é-kɔ̀ɻ̱ɔ̀* *Abdalla, ɻ̱ɔ̀ɔ̀ɖ̱ɔ̀* *Rowa, ɻ̱ɔ̀ɔ̀ɖ̱ɔ̀-ɻ̱éèr*	Minni:	*Rowa, màà* *Sadiq, ɻ̱íì* *Abir, wɔ́ɔ́*
Salva:	*Abir, wɔ́ɔ́* *Malak, áɲá-n-ɻ̱ɔ̀ɔ̀ɖ̱ɔ̀* *Nadwah, íɻ̱è ɻ̱éèr*	Nadwah:	*Kuwa, wɔ́ɔ́* *Abdalla, fáàfá* *Rowa, áɲá*
Sihan:	*Sadiq, íɻ̱è* *Minni, íɻ̱è-n-ɻ̱ɔ̀ɔ̀ɖ̱ɔ̀-kɔ̀ɻ̱ɔ̀* *Kuwa, áfá*	Abir:	*Malak, ɔ́ɻ̱é* *Nadwah, ɔ́ɻ̱é* *Sadiq, ɻ̱ɔ̀ɔ̀ɖ̱ɔ̀-kɔ̀ɻ̱ɔ̀*
Malak:	*Minni, íɻ̱è* *Sihan, màà* *Abdalla, ɻ̱íì*	Abdalla:	*Kuwa, áfá* *Malak, íɻ̱è-n-ɻ̱ɔ̀ɔ̀ɖ̱ɔ̀* *Salva, íɻ̱è-n-ɻ̱ɔ̀ɔ̀ɖ̱ɔ̀-kɔ̀ɻ̱ɔ̀*

While filling in the diagram, Dr. Lovelang noticed that, in Tabaq, certain kinship terms can be expressed using two different terms, and one of them is derived from the other. Moreover, he noticed somewhere in the notebook the following information: *Sadiq is a man.* and *Rowa has children.*

Problem 10.7a Fill in the diagram above with the names of the ten family members (in the diagram above circles represents women and squares men).

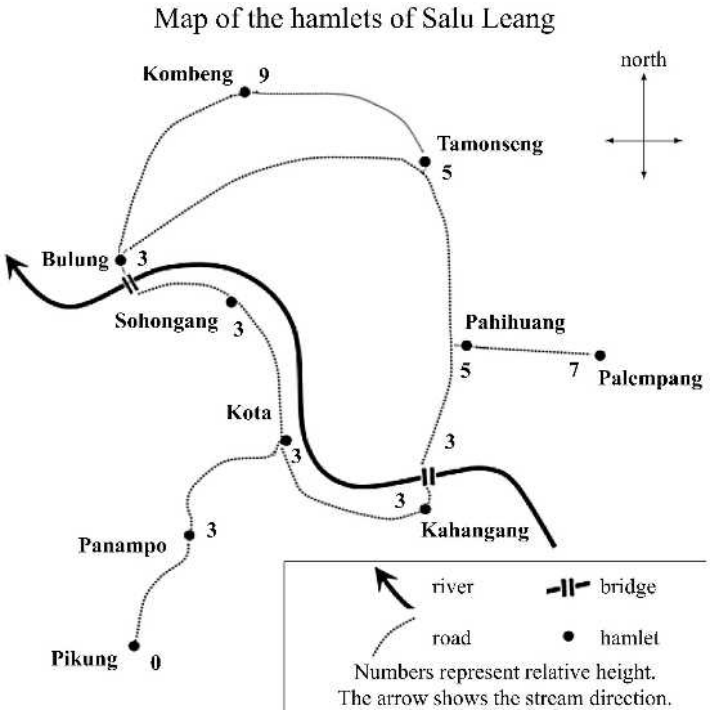
Problem 10.7b Write in Tabaq all the possible kinship terms that denote the relation between the following persons:

a. Sadiq to Salva
b. Abir to Nadwah
c. Salva to Sihan
d. Nadwah to Minni
e. Malak to Kuwa
f. Abdalla to Rowa
g. Abdalla to Salva

Problem 10.8

Aralle-Tabulahan (Ksenia Gilyarova, IOL 2016)

A linguist came to Salu Leang (Sulawesi) to study the Aralle-Tabulahan language. He visited various hamlets of Salu Leang (see the map below) and asked local residents: *Umba laungngola?* 'Where are you going?'



Below are the answers he got. There are gaps in some of them.

- In Kahangang hamlet:
  - *Lamaoä' bete' di Bulung.*
  - *Lamaoä' sau di Kota.*
  - *Lamaoä' __(1)__ di Palempang.*
- In Kombeng hamlet:
  - *Lamaoä' pano di Pahihuang.*
  - *Lamaoä' tama di Sohongang.*

  - *Lamaoä' naung di Tamonseng.*
  - *Lamaoä' __(2)__ di Palempang.*
- In Kota hamlet:
  - *Lamaoä' dai' di Kombeng.*
  - *Lamaoä' dai' di Palempang.*
  - *Lamaoä' naung di Pikung.*
  - *Lamaoä' __(3)__ di Bulung.*
  - *Lamaoä' __(4)__ di Sohongang.*
- In Palempang hamlet:
  - *Lamaoä' bete' di Kahangang.*
  - *Lamaoä' dai' di Kombeng.*
  - *Lamaoä' pano di Panampo.*
  - *Lamaoä' sau di Sohongang.*
  - *Lamaoä' __(5)__ di Bulung.*
  - *Lamaoä' __(6)__ di Kota.*
  - *Lamaoä' __(7)__ di Pahihuang.*
- In Pahihuang hamlet:
  - *Lamaoä' naung di Bulung.*
  - *Lamaoä' naung di Pikung.*
- In Bulung hamlet:
  - *Lamaoä' pano di Pahihuang.*
  - *Lamaoä' pano di Panampo.*
  - *Lamaoä' __(8)__ di Kota.*
  - *Lamaoä' __(9)__ di Pikung.*
- In Panampo hamlet:
  - *Lamaoä' tama di Kahangang.*
  - *Lamaoä' pano di Tamonseng.*
  - *Lamaoä' __(10)__ di Kota.*
- In Pikung hamlet:
  - *Lamaoä' pano di Kota.*

  - *Lamaoä' dai' di Pahihuang.*
  - *Lamaoä' sau di Sohongang.*
  - *Lamaoä' __(11)__ di Bulung.*
  - *Lamaoä' __(12)__ di Kahangang.*
  - *Lamaoä' __(13)__ di Panampo.*
- In Sohongang hamlet:
  - *Lamaoä' bete' di Bulung.*
  - *Lamaoä' tama di Kahangang.*
  - *Lamaoä' tama di Kota.*
  - *Lamaoä' dai' di Pahihuang.*
- In Tamonseng hamlet:
  - *Lamaoä' pano di Pahihuang.*
  - *Lamaoä' pano di Panampo*
  - *Lamaoä' __(14)__ di Kahangang.*
  - *Lamaoä' __(15)__ di Palempang.*

Problem 10.8a Fill in the blanks.

Problem 10.9

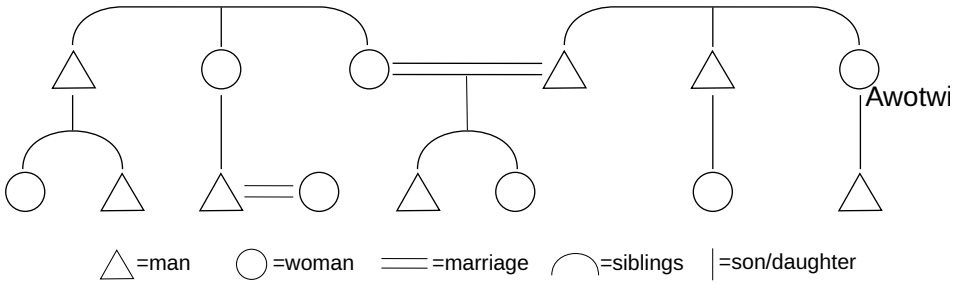
Akan (Ksenia Gilyarova, IOL 2018)

Below three Akan men who belong to one family introduce themselves and some members of their family (see the family tree):

1. *Yɛfrɛ me Enu. Yɛfrɛ me banom Thema ne Yaw ne Ama. Yɛfrɛ me yere Kunto. Yɛfrɛ me nuanom Awotwi ne Nsia. Yɛfrɛ me wɔfaase Berko.*

2. *Yɛfrɛ me Kofi. Yɛfrɛ me nua Esi. Yɛfrɛ me agya Ofori. Yɛfrɛ me sewaanom Dubaku ne Kunto. Yɛfrɛ me sewaabanom Yaw ne Ama ne Kobina.*

3. *Yɛfrɛ me Kobina. Yɛfrɛ me ɛnanom Dubaku ne Kunto. Yɛfrɛ me nuanom Yaw ne Ama. Yɛfrɛ me wɔfa Ofori. Yɛfrɛ me yere Efua.*



Problem 10.9a Supply the family tree with names.

Problem 10.9b Here are some more statements by two other men from this family:

4. *Yɛfrɛ me Yaw. Yɛfrɛ me ɛnanom* __(1)__. *Yɛfrɛ me* __(2)__*Nsia ne* __(3)__. *Yɛfrɛ me nuanom Thema ne* __(4)__. *Yɛfrɛ me* __(5)__*Awotwi. Yɛfrɛ me* __(6)__*Ofori. Yɛfrɛ me* __(7)__*Esi ne* __(8)__. *Yɛfrɛ me* __(9)__*Berko.*
5. *Yɛfrɛ me* __(10)__. *Yɛfrɛ me banom Kofi ne* __(11)__. *Yɛfrɛ me* __(12)__*Yaw ne* __(13)__. *Yɛfrɛ me* __(14)__*Kunto ne* __(15)__.

Fill in the gaps. Some gaps contain more than one word.

10.4 Solutions of practice problems

Solution for practice problem 10.3. Bardi

Solution 10.3a 1-l, 2-g, 3-k, 4-b, 5-j, 6-f, 7-i, 8-a, 9-d, 10-m, 11-c, 12-h, 13-e.

Solution for practice problem 10.4. Kharia

Solution 10.4a

a. Dele
b. Kepka
c. Mariya
d. Anil
e. Beni
f. Thuyu
g. Rut
h. Olem
i. Muni
j. Ewa
k. Sim
l. Nuh
m. Kolo

Solution 10.4b

1. *Sowiŋaʔ*
2. *Dad=iŋaʔ*
3. *Kolo oɖoyoʔ no Olem*
4. *Beni*
5. *Apaiŋaʔ*

Solution 10.4c

6. *Bhaiiŋaʔ*
7. *Beʔʈiŋaʔ*
8. *Didikiiŋaʔ*

Rules: Sentence structure: [Kinship term]–(PL)–*iŋaʔ ɲimi*–(PL) [Persons]

1. PL = *-ki-* = plural marker (if there is more than one person)
2. Persons: Their names; if it's more than one person, *oɖoyoʔ no* = 'and' is added before the last one.
3. Kinship terms:
   - *bhai* = 'younger brother'
   - *dad=* = 'older brother' (plural: *dadaki*)
   - *konon bahin* = 'younger sister'
   - *didi* = 'older sister'
   - *beʔʈ* = 'son'
   - *apa* = 'father'
   - *sow* = 'husband'
   - *đonkui* = 'brother's wife'

Solution for practice problem 10.5. Embaloh

Solution 10.5a On the first day, when he was told the four directions, he automatically assumed that they represented cardinal directions. In reality, they are based on the topography of the area and represent the relative positions with respect to the river. As such:

andoor = 'towards (closer to) the river'

anait = 'away from the river'

urait = 'upstream'

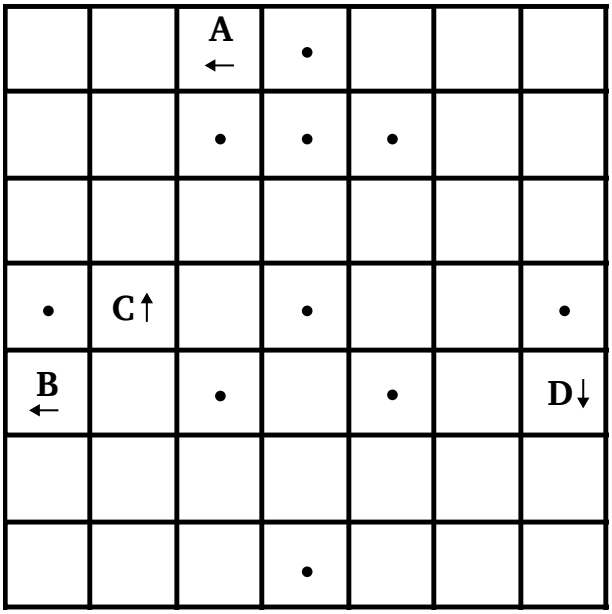
kalaut = 'downstream'

suali = 'across the river'

Solution 10.5b

1. *kalaut*
2. *andoor*
3. *suali*

Solution for practice problem 10.6. Hungarian



előttem = ‘front’	*északra* = ‘north’	*nincs* = ‘0’
mögöttem = ‘back’	*délre* = ‘south’	*egy* = ‘1’
balra = ‘left’	*nyugatra* = ‘west’	*két* = ‘2’
jobbra = ‘right’	*keletere* = ‘east’	

Solution for practice problem 10.7. Tabaq

Solution 10.7a The names in the diagram are, from left to right:

Top generation: Abir, Kuwa

Middle generation: Sihan, Rowa, Sadiq, Abdalla

Bottom generation: Malak, Minni, Nadwah, Salva

Solution 10.7b

a. *áfá*

b. *wɔ́ɔ́*

c. *íɻ̪è-n-ɻ̪ɔ̀ɔ̀ɖ̪ʊ̀* or *íɻ̪è-n-ɻ̪ɔ̀ɔ̀ɖ̪ʊ̀-kɔ̀ɻ̪ʊ̀*

d. *tíì-n-ɻ̪ɔ̀ɔ̀ɖ̪ʊ̀* or *tíì-n-ɻ̪ɔ̀ɔ̀ɖ̪ʊ̀-ɻ̪éɛ̀r*

e. *ʊ́ɻ̪é* or *ʊ́ɻ̪é-ɻ̪éɛ̀r*

f. *íṯè* or *íṯè-kɔ̀ṯʊ̀*

g. *fáàfá*

Rules: The kinship terms are:

*íṯè** = ‘sibling’	*áɲá* = ‘father’s sister’
*ṯɔ̀ɔ̀ḏʊ** = ‘child’	*màà* = ‘mother, mother’s sister’
*ʊ́ṯé** = ‘grandchild’	*fáàfá* = ‘father’s brother’
*íṯè-n-ṯɔ̀ɔ̀ḏʊ̀** = ‘nephew, niece’	*tíì* = ‘mother’s brother’
wɔ́ɔ́ = ‘grandparent’	*áfá* = ‘father’

The words marked with * can receive the suffixes *-ṯéèr* and *-kɔ̀ṯʊ̀*, in order to mark the gender (feminine and masculine, respectively).

The structure *X-n-ṯɔ̀ɔ̀ḏʊ* is translated as ‘X’s child’ (thus, a nephew/niece is actually translated as the ‘child of the sibling’). The words *màà*, *áɲá*, *tíì*, and *fáàfá* can be combined with *-n-ṯɔ̀ɔ̀ḏʊ̀* to express ‘cousin’.

Solution for practice problem 10.8. Aralle-Tabulahan

Solution 10.8a

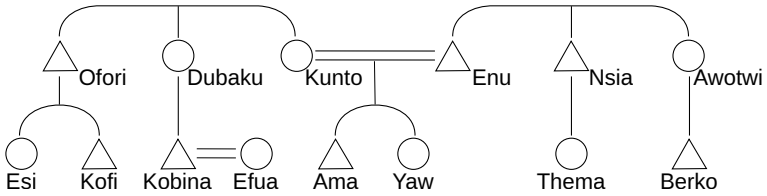
1. *dai’*	5. *naung*	9. *naung*	13. *dai’*
2. *dai’*	6. *sau*	10. *pano*	14. *bete’*
3. *bete’*	7. *naung*	11. *bete’*	15. *dai’*
4. *sau*	8. *tama*	12. *tama*	

Rules: Basic directions are:

bete’ = ‘across the river’	*pano* = ‘on a flat road’
tama = ‘upstream’	*dai’* = ‘upwards’
sau = ‘downstream’	*naung* = ‘downwards’

Solution for practice problem 10.9. Akan

Solution 10.9a



Solution 10.9b

(1) *Dubaku ne Kunto*
(2) *agyanom*
(3) *Enu*
(4) *Ama ne Kobina*
(5) *sewaa*
(6) *wɔfa*
(7) *wɔfabanom*
(8) *Kofi*
(9) *sewaaba*
(10) *Ofori*
(11) *Esi*
(12) *wɔfaasenom*
(13) *Ama ne Kobina*
(14) *nuanom*
(15) *Dubako*

Rules:

- *Yɛfrɛ me N.* = 'My name is N.'
- *Yɛfrɛ me R N.* = 'My R's name is N.' (*R* = kinship term)
- *X ne Y* = 'X and Y'
- *-nom* = PL
- Kinship terms:
  - *ɛna* = 'mother, mother's sister'
  - *wɔfaase* = 'sister's child'
  - *sewaa* = 'father's sister'
  - *sewaaba* = 'father's sister's child'
  - *nua* = 'sibling, parallel cousin'
  - *agya* = 'father, father's brother'
  - *ba* = 'child, brother's child'
  - *wɔfa* = 'mother's brother'
  - *wɔfaba* = 'mother's brother's child'
  - *yere* = 'wife'

Further reading

Dousset, Laurent. 2012. *Australian aboriginal kinship: An introductory handbook with particular emphasis on the Western Desert*. Marseilles: Pacific-Credo Publications.

Fortescue, Michael. 2011. *Orientation systems of the North Pacific Rim*. Chicago: The University of Chicago Press.

Levinson, Stephen C. 2003. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.

Parkin, Robert. 2021. *How kinship systems change*. New York: Berghahn Books.

Appendix A: Data about the languages featured in the problems: Family, number of native speakers, region

Acehnese
Problem no: 5.8 *Native speakers:* 3.37 mil.
Family: Austronesian *Region:* Indonesia

Afrihili
Problem no: 5.12 *Native speakers:* 0
Family: Constructed *Region:* –

Ainu
Problem no: 6.13 *Native speakers:* 2
Family: Isolated *Region:* Hokkaido (Japan)

Akan
Problem no: 10.9 *Native speakers:* 11 mil.
Family: Niger-Congo *Region:* Ghana, Ivory Coast, Togo

Alabama
Problem no: 5.8 *Native speakers:* 370
Family: Muskogean *Region:* Texas (USA)

Alamblak
Problem no: 9.14 *Native speakers:* 1,500
Family: Sepik *Region:* Papua New Guinea

Arabic
Problem no: 2.9, 4.9, 7.7 *Native speakers:* 350 mil.
Family: Afroasiatic *Region:* Asia, Africa

Aralle-Tabulahan
Problem no: 10.8 *Native speakers:* 12,000
Family: Austronesian *Region:* West Sulawesi (Indonesia)

Arawak
Problem no: 10.2 *Native speakers:* 2,510
Family: Arawakan *Region:* South America

Armenian
Problem no: 2.5 *Native speakers:* 6.7 mil.
Family: Indo-European *Region:* Armenia

Bardi
Problem no: 10.3 *Native speakers:* 4
Family: Nyulnyulan *Region:* Australia

Bari
Problem no: 4.13, 5.8 *Native speakers:* 750,000
Family: Nilo-Saharan *Region:* South Sudan

Basque
Problem no: 8.3 *Native speakers:* 750,000
Family: Isolated *Region:* France, Spain

Beja
Problem no: 7.4 *Native speakers:* 1-2 mil.
Family: Afroasiatic *Region:* Sudan, Eritrea, Egypt

Bulgarian
Problem no: 5.4 *Native speakers:* 8 mil.
Family: Indo-European *Region:* Bulgaria

Burushaski
Problem no: 7.11 *Native speakers:* 112,000
Family: Isolated *Region:* Pakistan

Chabu
Problem no: 9.15 *Native speakers:* 400
Family: Isolated *Region:* Ethiopia

Chickasaw
Problem no: 3.8 *Native speakers:* 75
Family: Muskogean *Region:* South Oklahoma (USA)

Chinese
Problem no: 8.5
Family: Sino-Tibetan
Native speakers: 1.2 bil.
Region: China

Chuvash
Problem no: 3.4
Family: Turkic
Native speakers: 1 mil.
Region: Russia

Cree
Problem no: 6.12
Family: Algic
Native speakers: 96,000
Region: Canada, USA

Cushillococa Ticuna
Problem no: 4.14
Family: Isolated
Native speakers: 63,000
Region: Brazil, Columbia, Peru

Czech
Problem no: 9.7
Family: Indo-European
Native speakers: 14 mil.
Region: Czechia

Dabida
Problem no: 6.2
Family: Niger-Congo
Native speakers: 370,000
Region: Kenya

Danish
Problem no: 9.9
Family: Indo-European
Native speakers: 6 mil.
Region: Denmark

Daza
Problem no: 5.8
Family: Nilo-Saharan
Native speakers: 380,000
Region: Chad, Niger, Sudan, Libya

Dinka
Problem no: 6.15
Family: Nilo-Saharan
Native speakers: 1.3 mil.
Region: Sudan

Dutch
Problem no: 4.11
Family: Indo-European
Native speakers: 25 mil.
Region: Netherlands

Dyirbal
Problem no: 7.2
Family: Pama-Nyungan
Native speakers: 8
Region: Northeast Queensland (Australia)

Embaloh
Problem no: 10.5
Family: Austronesian
Native speakers: 10,000
Region: Indonesia

Embera Chami
Problem no: 9.2
Family: Chocoan
Native speakers: 7,800
Region: Columbia

English
Problem no: 2.7
Family: Indo-European
Native speakers: 1.2 bil.
Region: UK, USA, Australia, New Zealand

Estonian
Problem no: 4.12, 9.10
Family: Uralic
Native speakers: 1.1 mil.
Region: Estonia

Evenki
Problem no: 4.5
Family: Tungusic
Native speakers: 26,580
Region: China, Russia

Fijian
Problem no: 3.6, 5.6
Family: Austronesian
Native speakers: 340,000
Region: Fiji

Finnish
Problem no: 4.12
Family: Uralic
Native speakers: 5.8 mil.
Region: Finland

Fitzroy River
Problem no: 5.8
Family: Pama-Nyungan
Native speakers: Unknown
Region: Queensland (Australia)

Gee
Problem no: 6.9
Family: Niger-Congo
Native speakers: 330,000
Region: Togo, Benin

Ge'ez
Problem no: 6.3
Family: Afroasiatic
Native speakers: Extinct
Region: Eritrea, Ethiopia

Greek
Problem no: 3.5, 5.7
Family: Indo-European
Native speakers: 13.5 mil.
Region: Greece

Guaraní

Problem no: 8.2
Family: Tupian
Native speakers: 6.5 mil.
Region: Paraguay, Bolivia, Argentina, Brazil

Gyarung

Problem no: 6.10
Family: Sino-Tibetan
Native speakers: 83,000
Region: Sichuan Province (China)

Hakhun

Problem no: 6.11
Family: Sino-Tibetan
Native speakers: 108,000
Region: Burma, India

Hanunó'o

Problem no: 5.8
Family: Austronesian
Native speakers: 13,000
Region: Philippines

Hausa

Problem no: 5.8, 8.6
Family: Afroasiatic
Native speakers: 60 mil.
Region: Nigeria, Niger, Cameroon, Benin, Chad

Hebrew

Problem no: 2.9
Family: Afroasiatic
Native speakers: 9 mil.
Region: Israel

Hmong

Problem no: 2.1
Family: Hmong-Men
Native speakers: 3.7 mil.
Region: China, Vietnam, Laos

Huli

Problem no: 9.5
Family: Engan
Native speakers: 150,000
Region: Papua New Guinea

Hungarian

Problem no: 10.6
Family: Uralic
Native speakers: 13 mil.
Region: Hungary

Iaai

Problem no: 5.17
Family: Austronesian
Native speakers: 4,100
Region: Ouvéa Island (New Caledonia)

Ibibio

Problem no: 5.8 *Native speakers:* 1.5-2 mil.
Family: Niger-Congo *Region:* South Nigeria

Ibo

Problem no: 5.8 *Native speakers:* 80 mil.
Family: Niger-Congo *Region:* Nigeria

Ilocano

Problem no: 5.15 *Native speakers:* 8.1 mil.
Family: Austronesian *Region:* Philippines

Indonesian

Problem no: 4.3 *Native speakers:* 43 mil.
Family: Austronesian *Region:* Indonesia

Irish

Problem no: 2.6, 4.1, 5.16 *Native speakers:* 72,000
Family: Indo-European *Region:* Ireland

Itelmen

Problem no: 6.4 *Native speakers:* 82
Family: Chukotko-Kamchatkan *Region:* Kamchatkan Peninsula (Russia)

Jalé

Problem no: 5.8 *Native speakers:* Unknown
Family: Trans New Guinea *Region:* New Guinea

Japanese

Problem no: 2.4, 5.5 *Native speakers:* 128 mil.
Family: Japonic *Region:* Japan

Javanese

Problem no: 2.10, 3.3 *Native speakers:* 82 mil.
Family: Austronesian *Region:* Java (Indonesia)

Kharia

Problem no: 10.4 *Native speakers:* 298,000
Family: Austroasiatic *Region:* India

La-Mi

Problem no: 4.6 *Native speakers:* 0
Family: Constructed *Region:* Taiwan

Lango
Problem no: 8.1 *Native speakers:* 2.1 mil.
Family: Nilo-Saharan *Region:* Uganda

Latvian
Problem no: 5.14 *Native speakers:* 1.75 mil.
Family: Indo-European *Region:* Latvia

Lepcha
Problem no: 2.8 *Native speakers:* 66,500
Family: Sino-Tibetan *Region:* Sikkim (India)

Ligurian
Problem no: 3.9 *Native speakers:* 600,000
Family: Indo-European *Region:* Liguria (Italy)

Luiseño
Problem no: 7.3 *Native speakers:* Extinct
Family: Uto-Aztecan *Region:* Southern California (USA)

Lunyole
Problem no: 4.3 *Native speakers:* 340,000
Family: Niger-Congo *Region:* Uganda

Luwian
Problem no: 2.2 *Native speakers:* Extinct
Family: Indo-European *Region:* Hittite Empire

Malagasy
Problem no: 8.8 *Native speakers:* 25 mil.
Family: Austronesian *Region:* Madagascar

Maltese
Problem no: 5.13 *Native speakers:* 520,000
Family: Afroasiatic *Region:* Malta

Manam
Problem no: 10.1 *Native speakers:* 8,000
Family: Austronesian *Region:* North of New Guinea

Mandar
Problem no: 4.3 *Native speakers:* 480,000
Family: Austronesian *Region:* Sulawesi (Indonesia)

Manobo
Problem no: 3.2
Family: Austronesian
Native speakers: 58,000
Region: Mindanao Region (Philippines)

Mansi
Problem no: 9.17
Family: Uralic
Native speakers: 12,300
Region: Russia

Māori
Problem no: 5.3
Family: Austronesian
Native speakers: 50,000
Region: New Zealand

Minangkabau
Problem no: 4.8
Family: Austronesian
Native speakers: 5.5 mil.
Region: West Sumatra (Indonesia)

Mundari
Problem no: 7.5
Family: Austroasiatic
Native speakers: 1.7 mil.
Region: India, Bangladesh, Nepal

Nasioi
Problem no: 5.8
Family: South Bougainville
Native speakers: 20,000
Region: Papua New Guinea

Nez-Perce
Problem no: 7.2
Family: Plateau Penutian
Native speakers: 20
Region: Idaho (USA)

N'gombe
Problem no: 5.8
Family: Niger-Congo
Native speakers: 150,000
Region: Democratic Republic of the Congo

Norwegian
Problem no: 5.11
Family: Indo-European
Native speakers: 5.32 mil.
Region: Norway

Nung
Problem no: 7.1
Family: Kra-Dai
Native speakers: 1 mil.
Region: Vietnam

Nupe
Problem no: 5.8
Family: Niger-Congo
Native speakers: 800,000
Region: Nigeria

Old Norse
Problem no: 3.7 *Native speakers:* Extinct
Family: Indo-European *Region:* Scandinavia

Palauan
Problem no: 5.10 *Native speakers:* 5,500
Family: Austronesian *Region:* Borneo

Proto-Algonquian
Problem no: 6.5 *Native speakers:* Extinct
Family: Algic *Region:* West of the USA

Quechua
Problem no: 4.3 *Native speakers:* 10 mil.
Family: Quechuan *Region:* South America

Quenya
Problem no: 9.1 *Native speakers:* 0
Family: Constructed *Region:* –

Roro
Problem no: 4.2 *Native speakers:* 15,000
Family: Austronesian *Region:* Eastern New Guinea

Rotokas
Problem no: 6.14 *Native speakers:* 4,320
Family: North Bougainville *Region:* Bougainville Island

Sandawe
Problem no: 7.10 *Native speakers:* 60,000
Family: Isolated *Region:* Tanzania

Selkup
Problem no: 9.12 *Native speakers:* 1,000
Family: Uralic *Region:* Russia

Sesotho
Problem no: 4.10 *Native speakers:* 5.6 mil.
Family: Niger-Congo *Region:* Lesotho, South Africa, Zimbabwe

Somali
Problem no: 3.1 *Native speakers:* 21.8 mil.
Family: Afroasiatic *Region:* Horn of Africa

Swahili
Problem no: 6.1, 6.7, 7.6, 9.8 *Native speakers:* 18 mil.
Family: Niger-Congo *Region:* Africa

Swedish
Problem no: 5.2 *Native speakers:* 10 mil.
Family: Indo-European *Region:* Sweden

Tabaq
Problem no: 10.7 *Native speakers:* 63,000
Family: Nilo-Saharan *Region:* Sudan

Tabasaran
Problem no: 6.6 *Native speakers:* 126,900
Family: Northeast Caucasian *Region:* North Caucasus

Tadaksahak
Problem no: 7.9 *Native speakers:* 100,000
Family: Nilo-Saharan *Region:* Mali, Niger

Tagbanwa
Problem no: 2.3 *Native speakers:* 2,000
Family: Austronesian *Region:* Palawan (Philippines)

Tahitian
Problem no: 7.2 *Native speakers:* 185,000
Family: Austronesian *Region:* French Polynesia

Tanna Island
Problem no: 5.8 *Native speakers:* Unknown
Family: Austronesian *Region:* Tanna Island (Vanuatu)

Tariana
Problem no: 6.8 *Native speakers:* 100
Family: Arawakan *Region:* Brazil

Tetum
Problem no: 8.7 *Native speakers:* 390,000
Family: Austronesian *Region:* East Timor

Thai
Problem no: 2.11 *Native speakers:* 36 mil.
Family: Kra-Dai *Region:* Thailand

Tifal
Problem no: 9.16 *Native speakers:* 4,000
Family: Trans New Guinea *Region:* Papua New Guinea

Tolaki
Problem no: 4.7 *Native speakers:* 330,000
Family: Austronesian *Region:* Sulawesi (Indonesia)

Turkish
Problem no: 8.4 *Native speakers:* 75.7 mil.
Family: Turkic *Region:* Türkiye

Tzeltal
Problem no: 5.8 *Native speakers:* 590,000
Family: Mayan *Region:* Mexico

Ulwa
Problem no: 5.9 *Native speakers:* 9,000
Family: Misumalpan *Region:* Nicaragua, Honduras

Umbu-Ungu
Problem no: 9.4 *Native speakers:* 77,000
Family: Trans New Guinea *Region:* Papua New Guinea

Upper Pyramid
Problem no: 5.8 *Native speakers:* Unknown
Family: Trans New Guinea *Region:* Papua Province (Indonesia)

Urhobo
Problem no: 5.8 *Native speakers:* 2 mil.
Family: Niger-Congo *Region:* Nigeria

Valley Yokuts
Problem no: 4.4 *Native speakers:* 20
Family: Yokuts *Region:* California (USA)

Vambon
Problem no: 9.13 *Native speakers:* 3,900
Family: Trans New Guinea *Region:* Papua Province (Indonesia)

Waorani
Problem no: 9.11 *Native speakers:* 1,800
Family: Isolated *Region:* Ecuador, Peru

Wappo

Problem no: 7.2
Family: Yuki-Wappo
Native speakers: Extinct
Region: Alexander Valley (California, USA)

Welsh

Problem no: 7.8
Family: Indo-European
Native speakers: 1 mil.
Region: Wales

Yoruba

Problem no: 9.6
Family: Niger-Congo
Native speakers: 41 mil.
Region: Nigeria

Yup'ik

Problem no: 9.3
Family: Eskimo-Aleut
Native speakers: 19,750
Region: West and southwest of Alaska (USA)

Zoque

Problem no: 4.3
Family: Mixe-Zoquean
Native speakers: 74,000
Region: Mexico

Zulu

Problem no: 5.1
Family: Niger-Congo
Native speakers: 12 mil.
Region: South Africa, Lesotho

Appendix B: Genealogical classification of the languages featured in the problems

Languages marked with "†" are extinct.

1. **AFROASIATIC FAMILY**
    a) **Chadic branch**: Hausa (5.8, 8.6);
    b) **Cushitic branch**
        i. North: Beja (7.4);
        ii. Lowland East: Somali (3.1);
    c) **Semitic branch > West Semitic**
        i. Central: Arabic (2.9, 4.9, 7.7), Hebrew (2.9), Maltese (5.13);
        ii. South: Ge'ez (6.3);
2. **ALGIC FAMILY**
    a) **Algonquian branch**: Proto-Algonquian † (6.5);
        i. Central: Cree (6.12);
3. **ARAWAKAN FAMILY**
    a) **Northern**
        i. North Amazonian: Tariana (6.8);
        ii. Ta-Maipurean: Arawak (10.2);
4. **AUSTROASIATIC FAMILY**
    a) **Munda branch**
        i. North: Mundari (7.5);
        ii. South: Kharia (10.4);

5. **AUSTRONESIAN FAMILY > MALAYO-POLYNESIAN**: Javanese (2.10, 3.3)
    a) **Barito**: Malagasy (8.8);
    b) **Celebic**: Tolaki (4.7);
    c) **Central-Eastern**
        i. Central > Timoric: Tetum (8.7);
        ii. Eastern > Oceanic
            A. *Central Pacific*: Fijian (3.6, 5.6);
            B. *Southern Oceanic*: Iaai (5.17), Tanna Island (5.8);
            C. *Western Oceanic*: Manam (10.1), Roro (4.2);
            D. *Polynesian > Tahitic*: Māori (5.3), Tahitian (7.2);
    d) **Philippine**
        i. Northern Luzon: Ilocano (5.15);
        ii. Greater Central Philippine
            A. *South Mangyan*: Hanunó'o (5.8);
            B. Manobo (3.2);
            C. *Palawanic*: Tagbanwa (2.3);
    e) **Malayo-Chamic**
        i. Chamic: Acehnese (5.8);
        ii. Malayic: Indonesian (4.3), Minangkabau (4.8);
    f) **North Bornean**: Palauan (5.10);
    g) **South Sulawesi**
        i. Bugis-Tamanic: Embaloh (10.5);
        ii. Northern: Aralle-Tabulahan (10.8), Mandar (4.3);
6. **NORTH BOUGAINVILLE FAMILY**: Rotokas (6.14);
7. **SOUTH BOUGAINVILLE FAMILY**: Nasioi (5.8);
8. **CHOCOAN FAMILY**: Embera Chami (9.2);
9. **CHUKOTKO-KAMCHATKAN FAMILY**
    a) **Kamchatkan**: Itelmen (6.4);

10. **ENGAN FAMILY**: Huli (9.5);
11. **ESKIMO-ALEUT FAMILY**
    a) **Eskimo**: Yup'ik (9.3);
12. **HMONG-MEN FAMILY**
    a) **Hmongic**: Hmong (2.1);
13. **INDO-EUROPEAN FAMILY**
    a) **Anatolian** †: Luwian (2.2);
    b) Armenian (2.5);
    c) **Balto-Slavic**
        i. Baltic: Latvian (5.14);
        ii. Slavic
            A. *South Slavic*: Bulgarian (5.4);
            B. *West Slavic*: Czech (9.7);
    d) **Celtic**
        i. Brittonic: Welsh (7.8);
        ii. Goidelic: Irish (2.6, 4.1, 5.16);
    e) **Germanic**
        i. North Germanic: Old Norse † (3.7);
            A. *East Scandinavian*: Danish (9.9), Swedish (5.2);
            B. *West Scandinavian*: Norwegian (5.11);
        ii. West Germanic: Dutch (4.11), English (2.7);
    f) **Hellenic**: Greek (3.5, 5.7);
    g) **Italic**: Ligurian (3.9);
14. **JAPONIC FAMILY**: Japanese (2.4, 5.5);
15. **KRA-DAI FAMILY**
    a) **Tai**: Nung (7.1), Thai (2.11);
16. **MAYAN FAMILY**: Tzeltal (5.8);
17. **MISUMALPAN FAMILY**: Ulwa (5.9);

18. **MIXE-ZOQUEAN FAMILY**: Zoque (4.3);
19. **MUSKOGEAN FAMILY**
    a) **Eastern**: Alabama (5.8);
    b) **Western**: Chickasaw (3.8);
20. **NIGER-CONGO FAMILY > ATLANTIC-CONGO**
    a) **Benue-Congo**
        i. Bantoid > Bantu
            A. *Northeast Bantu*
                - *Chaga-Taita*: Dabida (6.2);
                - *Great Lakes*: Lunyole (4.3);
                - *Sabaki*: Swahili (6.1, 6.7, 7.6, 9.8);
            B. *Southern Bantu*: Sesotho (4.10), Zulu (5.1);
            C. *Buja-Ngombe*: N'gombe (5.8);
        ii. Cross River > Efik-Ibibio: Ibibio (5.8);
    b) **Kwa**
        i. Gbe: Gee (6.9);
        ii. Potou-Tano: Akan (10.9);
    c) **Volta-Niger**
        i. Edoid: Urhobo (5.8);
        ii. Igboid: Ibo (5.8);
        iii. Nupoid: Nupe (5.8);
        iv. Yoruboid: Yoruba (9.6);
21. **NILO-SAHARAN FAMILY**
    a) **Saharan**: Daza (5.8);
    b) **Songhay**: Tadaksahak (7.9);
    c) **Eastern Sudanic**
        i. Northern > Nubian: Tabaq (10.7);
        ii. Southern > Nilotic

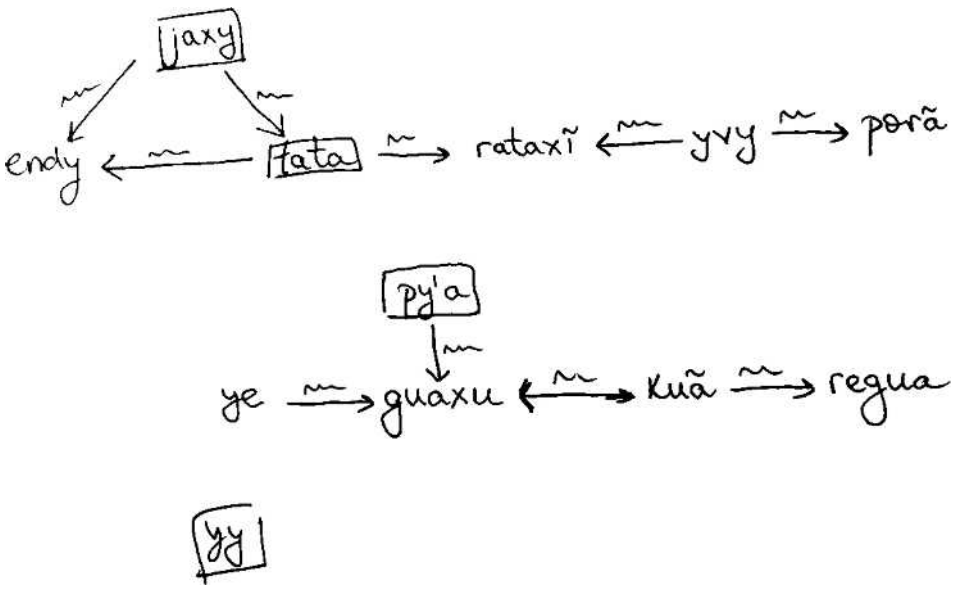
A. *Eastern*: Bari (4.13, 5.8), Lango (8.1);
B. *Western*: Dinka (6.15);

22. **NORTHEAST CAUCASIAN FAMILY**
    a) **Lezgic**: Tabasaran (6.6);
23. **NYULNYULAN FAMILY**: Bardi (10.3);
24. **PAMA-NYUNGAN FAMILY**
    a) **Dyirbalic**: Dyirbal (7.2);
    b) **Maric**: Fitzroy River (5.8);
25. **PLATEAU PENUTIAN FAMILY**: Nez-Perce (7.2);
26. **QUECHUAN FAMILY**: Quechua (4.3);
27. **SEPIK FAMILY**: Alamblak (9.14);
28. **SINO-TIBETAN FAMILY**
    a) **Bodo-Konyak-Jinghpaw**
        i. Konyak: Hakhun (6.11);
    b) Chinese (8.5);
    c) **Qiangic**: Gyarung (6.10);
    d) **Tibeto-Burman > Himalayish**: Lepcha (2.8);
29. **TRANS NEW GUINEA FAMILY**
    a) **Asmat-Awyu-Ok > Awyu-Ok**
        i. Awyu: Vambon (9.13);
        ii. Ok: Tifal (9.16);
    b) **Chimbu-Wahgi**
        i. Hagen > Aua-Gawil: Umbu-Ungu (9.4);
    c) **Dani**:
        i. Central Dani: Upper Pyramid (5.8);
        ii. Ngalik-Nduga: Jalé (5.8);
30. **TUNGUSIC FAMILY**: Evenki (4.5);

31. **TUPIAN FAMILY**: Guaraní (8.2);
32. **TURKIC FAMILY**
    a) **Oghuric**: Chuvash (3.4);
    b) **Common Turkic > Oghuz**: Turkish (8.4);
33. **URALIC FAMILY**
    a) **Finnic**: Estonian (4.12, 9.10), Finnish (4.12);
    b) **Samoyedic**: Selkup (9.12);
    c) **Ugric**: Hungarian (10.6), Mansi (9.17);
34. **UTO-AZTECAN FAMILY**
    a) **Takic**: Luiseño (7.3);
35. **YOKUTS FAMILY**: Valley Yokuts (4.4);
36. **YUKI-WAPPO FAMILY**: Wappo (7.2);
37. **ISOLATED LANGUAGES**
    a) **Africa**: Chabu (9.15), Sandawe (7.10);
    b) **South America**: Cushillococa Ticuna (4.14), Waorani (9.11);
    c) **Asia**: Ainu (6.13), Burushaski (7.11);
    d) **Europe**: Basque (8.3);
38. **ARTIFICIAL (CONSTRUCTED) LANGUAGES**
    a) **Artistic languages**
        i. Fictional languages: Quenya (9.1);
        ii. Secret languages: La-Mi (4.6);
    b) **International auxiliary languages**: Afrihili (5.12);

Appendix C: The handwritten graph corresponding to Problem 8.2

We preferred to circle the words that appear in the corpus rather than underline them in order to increase visibility.



References

Aikhenvald, Alexandra Y. & Robert M. W. Dixon (eds.). 2013. *Possession and ownership*. Oxford: Oxford University Press.

Berlin, Brent & Paul Kay. 1969. *Basic color terms: Their universality and evolution*. 1st edn. Berkeley: University of California Press.

Booij, Geert. 2012. *The grammar of words*. Oxford: Oxford University Press.

Cabredo Hofherr, Patricia & Anne Zribi-Hertz (eds.). 2013. *Crosslinguistic studies on noun phrase structure and reference*. Leiden: Brill Academic Publishers.

Carnie, Andrew. 2011. *Modern syntax: A coursebook*. Cambridge: Cambridge University Press.

Chrisomalis, Stephen. 2010. *Numerical notation: A comparative history*. Cambridge: Cambridge University Press.

Cinque, Guglielmo & Richard S. Kayne (eds.). 2008. *The Oxford handbook of comparative syntax*. Oxford: Oxford University Press.

Coon, Jessica. 2013. *Aspects of split ergativity*. Oxford: Oxford University Press.

Coon, Jessica, Diane Massam & Lisa Demena Travis (eds.). 2017. *The Oxford handbook of ergativity*. Oxford: Oxford University Press.

Coulmas, Florian. 1999. *The Blackwell encyclopedia of writing systems*. New York: Wiley-Blackwell.

Coulmas, Florian. 2003. *Writing systems: An introduction*. Cambridge: Cambridge University Press.

Cruttenden, Alan. 2021. *Writing systems and phonetics*. London: Routledge.

Daniels, Peter T. & William Bright. 1996. *The world's writing systems*. Oxford: Oxford University Press.

Dixon, Robert M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Donohue, Mark & Søren Wichmann (eds.). 2005. *The typology of semantic alignment*. Oxford: Oxford University Press.

Dousset, Laurent. 2012. *Australian aboriginal kinship: An introductory handbook with particular emphasis on the Western Desert*. Marseilles: Pacific-Credo Publications.

Fortescue, Michael. 2011. *Orientation systems of the North Pacific Rim*. Chicago: The University of Chicago Press.

Fortescue, Michael, Marianne Mithun & Nicholas Evans (eds.). 2017. *The Oxford handbook of polysynthesis*. Oxford: Oxford University Press.
Haarmann, Harald. 2004. *Geschichte der Schrift*. München: C. H. Beck.
Hengeveld, Kees, Heiko Narrog & Hella Olbertz (eds.). 2017. *The grammaticalization of tense, aspect, modality and evidentiality*. Berlin: De Gruyter Mouton.
Hurford, James R. 2011. *The linguistic theory of numerals*. Cambridge: Cambridge University Press.
Ifrah, George. 2000. *The universal history of numbers: From prehistory to the invention of the computer*. New York: John Wiley & Sons Inc.
Koptjevskaja-Tamm, Maria & Henrik Liljegren. 2017. Semantic patterns from an areal perspective. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*. Cambridge: Cambridge University Press.
Kroeger, Paul. 2022. *Analyzing meaning: An introduction to semantics and pragmatics*. Berlin: Language Science Press.
Ladefoged, Peter & Keith Johnson. 2014. *A course in phonetics*. Wadsworth: Cengage Learning.
Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.
Levinson, Stephen C. 2003. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.
Li, Paul Jen-kuei. 1985. A secret language in Taiwanese/台湾的秘密语言. *Journal of Chinese Linguistics* 13(1). 91–121. http://www.jstor.org/stable/23767095.
Parkin, Robert. 2021. *How kinship systems change*. New York: Berghahn Books.
Refsing, Kirsten. 1986. *The Ainu language: The morphology and syntax of the Shizunai dialect*. Aarhus University Press.
Rijkhoff, Jan. 2002. *The noun phrase*. Oxford: Oxford University Press.
Robinson, Andrew. 1999. *The story of writing*. New York: Thames & Hudson.
Rocca, Iggy & Wyn Johnson. 1999. *Course in phonology*. Oxford: Blackwell.
Setter, Jane. 2021. *Your voice speaks volumes*. Oxford: Oxford University Press.
Shibatani, Masayoshi. 1990. *The languages of Japan*. Reprint 1994. Cambridge University Press.
ten Hacken, Pius (ed.). 2016. *The semantics of compounding*. Cambridge: Cambridge University Press.
Vanhove, Martine (ed.). 2008. *From polysemy to semantic change*. Amsterdam: John Benjamins Publishing Company.
Zaslavsky, Claudia. 1999. *Africa counts: Number and pattern in African cultures*. Chicago: Chicago Review Press.
Zsiga, Elizabeth C. 2013. *The sounds of language: An introduction to phonetics and phonology*. Chichester: Wiley-Blackwell.

Linguistics Olympiad

Linguistics Olympiad: Training guide represents a unique and complex work aimed to help students and teachers alike prepare for the national and international Linguistics Olympiads. This guide identifies the most common types of problems and, for each of them, proposes a theoretical framework (basic linguistics concepts, as well as language typology data) together with a methodological approach, tailored for each type of problems, and, in the end, a selection of practice problems from past editions of national and international Linguistics Olympiads. This work is breaking new ground, being the first of its kind, featuring a large number of languages and problems, centred around the concept of problem-based learning.