Introduction to Linguistics

Michael Barrie

January 23, 2021
Acknowledgements

This textbook is a work in progress, and, I suspect, will likely remain so. I wish to thank previous students who have used this book and have made comments and requests for clarification. Some of the data used for discussion and problem sets originally came from my own fieldwork. I wish to thank several Blackfoot speakers including the late Beatrice Bullshields, Daniel Crazy Bull, Randa Weasel Head, the late Brent Prairiechicken, Shelly Prairiechicken, and Sandra Manyfeathers. Thanks also to Moonhyun Sung, Soowhan Lee, and Sihun Jung who assisted greatly in carrying out the fieldwork on Blackfoot. Thanks also to Martina Wiltschko, Heather Bliss, and Meagan Louie for discussions on Blackfoot grammar. Some Iroquoian data are thanks to Daisy Elijah (Oneida), Gloria Williams and the late Nora Carrier (Onondaga), and Barb Garlow and Ruby Williams (Cayuga). Thanks also to Marianne Mithun, Karin Michelson, Carrie Dyck, Amos Keye, and Kehte Deer for numerous helpful discussions on Iroquoian grammar. Thanks to Moonhyun Sung, Dongwoo Park, Jungu Kang and Inkie Chung for helping with the Korean data. Thanks especially to Inkie Chung for testing an earlier version of this textbook with his class and for his numerous suggestions for revisions. Some ideas for the structure of Chapter 6 (Semantics) are thanks to [Ai 2020].
# Contents

Acknowledgements ........................................................................................................... i

1 Foundations of the Study of Language ........................................................................ 1
   1.1 The Science of Language .......................................................................................... 1
   1.2 Language, Communication, and Animals ................................................................. 3
   1.3 Prescriptivism versus Descriptivism ........................................................................... 9
   1.4 Language and Thought ............................................................................................. 12
      1.4.1 Linguistic Determinism ...................................................................................... 12
      1.4.2 Linguistic Nativism ............................................................................................ 14
   1.5 Core Areas of Linguistics ......................................................................................... 16
   1.6 Properties of Universal Grammar ............................................................................ 19
   1.7 Languages and Dialects ............................................................................................ 20
   Exercises ....................................................................................................................... 22
   Further Reading ........................................................................................................... 23

2 Morphology: The pieces of words .............................................................................. 24
   2.1 Words and their Pieces ........................................................................................... 24
      2.1.1 Morphemes and Words ...................................................................................... 28
      2.1.2 Describing Morphemes ...................................................................................... 30
      2.1.3 The Positions of Morphemes ............................................................................. 33
   2.2 The Grammar of Derivational Morphology ............................................................. 35
      2.2.1 English Derivational Morphology ..................................................................... 36
      2.2.2 Korean Derivational Morphology ..................................................................... 40
   2.3 The Functions of Inflectional Morphology ............................................................... 40
      2.3.1 Noun Class ........................................................................................................ 41
      2.3.2 Number ............................................................................................................. 44
      2.3.3 Person ................................................................................................................ 45
      2.3.4 Tense ............................................................................................................... 46
      2.3.5 Case .................................................................................................................. 47
   2.4 Analyzing Morphology ............................................................................................ 48
5 Syntax

5.1 Introduction .................................................................................................................. 127
5.1.1 Basic Properties of Phrase Structure .................................................................... 128
5.2 Constituency Tests ........................................................................................................ 134
5.2.1 Proforms .................................................................................................................. 135
5.2.2 Stand Alone ............................................................................................................ 137
5.2.3 Coordination Tests .................................................................................................. 138
5.2.4 Movement Tests (advanced) .................................................................................... 139
5.3 The Content of Lexical Categories ............................................................................... 145
5.3.1 Noun Phrases ......................................................................................................... 145
5.3.2 Verb Phrases .......................................................................................................... 147
5.3.3 Adpositional Phrases ............................................................................................. 148
5.3.4 Adjective Phrases ................................................................................................... 150
5.3.5 Adverbial Phrases .................................................................................................. 151
5.3.6 One-replacement and nested VPs (advanced) ......................................................... 152
5.4 Clauses .......................................................................................................................... 154
5.5 Word Order Typology ................................................................................................. 161
5.6 Transformations ........................................................................................................... 166
5.6.1 Polarity Questions .................................................................................................. 166
5.6.2 Passivization ........................................................................................................... 167
Answers to Practice Questions .......................................................................................... 172
Exercises ............................................................................................................................ 175
Further Reading .................................................................................................................. 177

6 Semantics .......................................................................................................................... 178

6.1 Introduction .................................................................................................................... 178
6.2 Word and Sentence Meaning ....................................................................................... 179
6.2.1 Semantic Properties of words .............................................................................. 179
6.2.2 Semantic Relationships among Sentences .............................................................. 185
6.3 A Brief History of Semantic Thought ......................................................................... 189
6.4 Denotational and Compositional Semantics ............................................................... 190
Exercises ............................................................................................................................ 193
Further Reading .................................................................................................................. 195

7 Pragmatics ........................................................................................................................ 196

7.1 Introduction .................................................................................................................... 196
7.2 Implicature ..................................................................................................................... 197
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2.1 Conversational Implicature</td>
<td>199</td>
</tr>
<tr>
<td>7.2.2 Conventional Implicature</td>
<td>199</td>
</tr>
<tr>
<td>7.2.3 Scalar Implicature</td>
<td>200</td>
</tr>
<tr>
<td>7.3 Speech Acts</td>
<td>201</td>
</tr>
<tr>
<td>7.4 Gricean Maxims</td>
<td>202</td>
</tr>
<tr>
<td>Exercises</td>
<td>205</td>
</tr>
<tr>
<td>Further Reading</td>
<td>208</td>
</tr>
<tr>
<td>8 Sociolinguistics</td>
<td>209</td>
</tr>
<tr>
<td>8.1 Introduction</td>
<td>209</td>
</tr>
<tr>
<td>8.2 Dialectology</td>
<td>210</td>
</tr>
<tr>
<td>8.3 Social and Linguistic Variation</td>
<td>218</td>
</tr>
<tr>
<td>8.4 Language and Gender</td>
<td>220</td>
</tr>
<tr>
<td>8.5 Contact Languages</td>
<td>221</td>
</tr>
<tr>
<td>Exercises</td>
<td>223</td>
</tr>
<tr>
<td>Further Reading</td>
<td>225</td>
</tr>
<tr>
<td>9 Historical Linguistics</td>
<td>226</td>
</tr>
<tr>
<td>9.1 Introduction</td>
<td>226</td>
</tr>
<tr>
<td>9.2 Words over Time</td>
<td>231</td>
</tr>
<tr>
<td>9.3 Historical Sound Changes</td>
<td>235</td>
</tr>
<tr>
<td>9.4 Comparative Historical Reconstruction</td>
<td>242</td>
</tr>
<tr>
<td>9.5 Historical Syntax</td>
<td>243</td>
</tr>
<tr>
<td>9.6 Genetic Classification</td>
<td>243</td>
</tr>
<tr>
<td>9.6.1 The Altaic Superfamily</td>
<td>244</td>
</tr>
<tr>
<td>9.6.2 Indo-European Languages</td>
<td>246</td>
</tr>
<tr>
<td>9.6.3 Sino-Tibetan Family</td>
<td>248</td>
</tr>
<tr>
<td>9.6.4 Austroasiatic Family</td>
<td>248</td>
</tr>
<tr>
<td>9.6.5 Kra-Dai Family</td>
<td>249</td>
</tr>
<tr>
<td>9.6.6 Austronesian Family</td>
<td>249</td>
</tr>
<tr>
<td>Answers to Practice Questions</td>
<td>250</td>
</tr>
<tr>
<td>Exercises</td>
<td>251</td>
</tr>
<tr>
<td>Further Reading</td>
<td>252</td>
</tr>
<tr>
<td>10 Language Acquisition</td>
<td>253</td>
</tr>
<tr>
<td>10.1 Introduction</td>
<td>253</td>
</tr>
<tr>
<td>10.2 L1 Acquisition</td>
<td>254</td>
</tr>
<tr>
<td>10.2.1 Language in Infancy</td>
<td>255</td>
</tr>
<tr>
<td>10.2.2 First Sounds and Words</td>
<td>256</td>
</tr>
<tr>
<td>10.2.3 The Development of Grammar</td>
<td>257</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>10.3 Acquisition of More than One Language</td>
<td>259</td>
</tr>
<tr>
<td>Exercises</td>
<td>261</td>
</tr>
<tr>
<td>Further Reading</td>
<td>262</td>
</tr>
<tr>
<td>11 Language Documentation, Endangerment, and Revitalization</td>
<td>263</td>
</tr>
<tr>
<td>11.1 Introduction</td>
<td>263</td>
</tr>
<tr>
<td>11.2 Language Endangerment</td>
<td>264</td>
</tr>
<tr>
<td>11.2.1 Causes of Endangerment</td>
<td>266</td>
</tr>
<tr>
<td>11.3 Language Documentation</td>
<td>266</td>
</tr>
<tr>
<td>11.4 Language Revitalization</td>
<td>267</td>
</tr>
<tr>
<td>11.4.1 Case Study: Cornish</td>
<td>267</td>
</tr>
<tr>
<td>11.4.2 Case Study: Mohawk</td>
<td>268</td>
</tr>
<tr>
<td>Exercises</td>
<td>270</td>
</tr>
<tr>
<td>Further Reading</td>
<td>271</td>
</tr>
<tr>
<td>12 Psycholinguistics and Neurolinguistics</td>
<td>272</td>
</tr>
<tr>
<td>12.1 Introduction</td>
<td>272</td>
</tr>
<tr>
<td>12.2 Speech Errors</td>
<td>274</td>
</tr>
<tr>
<td>12.3 Processing</td>
<td>275</td>
</tr>
<tr>
<td>12.4 Psycholinguistic Experimental Methodology</td>
<td>276</td>
</tr>
<tr>
<td>12.5 Brain Structure and Language</td>
<td>280</td>
</tr>
<tr>
<td>12.5.1 Aphasia</td>
<td>282</td>
</tr>
<tr>
<td>12.5.2 Specific Language Impairment and Other Disorders</td>
<td>283</td>
</tr>
<tr>
<td>Exercises</td>
<td>284</td>
</tr>
<tr>
<td>Further Reading</td>
<td>285</td>
</tr>
<tr>
<td>13 Biolinguistics</td>
<td>286</td>
</tr>
<tr>
<td>13.1 Introduction</td>
<td>286</td>
</tr>
<tr>
<td>13.2 A Biological Foundation for Language?</td>
<td>286</td>
</tr>
<tr>
<td>13.3 A Genetic Basis for Language?</td>
<td>287</td>
</tr>
<tr>
<td>Exercises</td>
<td>290</td>
</tr>
<tr>
<td>Further Reading</td>
<td>291</td>
</tr>
<tr>
<td>Glossary</td>
<td>291</td>
</tr>
<tr>
<td>Answers to Exercises</td>
<td>292</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>301</td>
</tr>
<tr>
<td>Index</td>
<td>302</td>
</tr>
</tbody>
</table>
List of Tables

1.1 Yale Romanization of Korean ......................................................... 3

2.1 Lexical Entry for /-ly/ ................................................................. 35

2.2 Derivational Affixes of English ................................................... 37

2.3 Derivational Affixes of Korean ...................................................... 40

2.4 Chinese Classifiers ................................................................. 43

2.5 Pronouns of Tolomako .............................................................. 45

2.6 Pronouns in Chinese Languages .................................................... 45

3.1 Types of Sounds by Airstream Mechanism ......................................... 69

3.2 IPA and General Descriptions of Vowel Height ................................... 74

3.3 Vowels of English ........................................................................ 82

5.1 Idioms from different languages ................................................ 133

13.1 Blood types in humans .............................................................. 288
List of Figures

1.1 Arbitrary relationship between a cat and the words meaning cat ........................................... 4
1.2 Universal Hierarchy of Colour Terms .............................................................. 13
1.3 Map of Inuit-Dit Dialects .................................................................................. 15
2.1 Distribution of allomorphs .................................................................................. 49
2.2 ASL signs for to sit and chair .......................................................................... 57
3.1 Types of phonetics .......................................................................................... 66
3.2 Vocal Apparatus .............................................................................................. 67
3.3 Pulmonic Airflow .............................................................................................. 68
3.4 Consonants of the IPA chart ........................................................................... 70
3.5 Position of fingers to detect voicing .................................................................. 71
3.6 Vowels of the IPA chart .................................................................................. 73
3.7 ASL sign for ‘head’ .......................................................................................... 77
3.8 Korean Consonants of the IPA chart ................................................................. 79
3.9 Korean Vowels of the IPA chart ....................................................................... 80
3.10 English Consonants of the IPA chart ............................................................... 81
3.11 English Vowels of the IPA chart ..................................................................... 82
4.1 Minimali Pairs in ASL ...................................................................................... 89
4.2 Minimal and Near Minimal Pairs ..................................................................... 90
4.3 Environments in which [s] and [ʃ] are found .................................................... 91
4.4 Distribution of allophones ............................................................................... 91
4.5 Phonological Rule ............................................................................................ 91
4.6 Environments in which [t], [tf], and [ts] are found ........................................... 93
4.7 Environments for Korean oral stop consonants ................................................. 94
4.8 Amplitude timeline of "The bird flew away." .................................................... 102
6.4 Different uses of mug ....................................................................................... 181
6.1 Entry for bright ................................................................................................. 182
6.2 Entry for bank .................................................................................................. 183
ACCR: Accusative case
AGR ASP: Aspect
CL: Classifier
COMPL: Completive aspect
DECL: Declarative
DEF: Definite
DET: Determiner
DIV: Division
DUR: Durative
FUT: Future
GEN: Genitive case
GN: General Number
HAB: Habitual
IND: Indirect
INTER: Interrogative
LOC: Locative case
N: Noun
NEG: Negation
NOM: Nominative case
NSP: Non-specific
NUM/#: Number
PART: Participle
PL: Plural
PRF: Perfect
PRFV: Perfective
PROG: Progressive
Q: Quantifier
SG: Singular
SP: Specific
Chapter 1

Foundations of the Study of Language

By the end of this chapter you should be able to:

- explain the terms descriptivism and prescriptivism
- know the five core areas of theoretical linguistics
- briefly explain the notion of generative grammar
- explain the difference between competence and performance
- present arguments for universal grammar (UG).

1.1 The Science of Language

When asked what sets humans apart from the other animals, a common response is language. Language is different from mere communication, for which many species of animals have evolved elaborate forms. The much-studied waggle dance of bees is well-known example of animal communication. As the panel below indicates, bees are able to communicate the distance and direction of a food source over 35 m away. Their “conversations”, however, are limited to the topic of food, and can only make reference to the present. Bees cannot discuss the location of a food source from yesterday, last week or “when I was a young worker bee….” Later, we will discuss some properties of human language, and what sets it apart from animal communication.

For now, let us consider the following contrast and see what it has to do with the study of language. The superscript numbers show coreference. Thus, in sentence (1), John and him both refer to the same person, namely John. The asterisk, *, in front the last two sentences means that they are ungrammatical. We will discuss this concept in more detail in a bit, but for now, it just means that these sentences are unacceptable to a native speaker of English.

(1)  a. John¹ thinks that Mary likes him¹.
    b. John thinks that Mary² likes herself².
    c. *John¹ thinks that Mary likes himself¹.
    d. *John thinks that Mary² likes her².
Linguists have adopted a universal way of presenting linguistics data from any language in the world. A sentence is presented with words lined up with their glosses followed by an idiomatic translation into the language of use. We are using English, so we give an idiomatic translation into English. Consider the following examples closely.

(2) Presentation of Linguistic Data (da Cunha and Cintra 1985)

a. Este aluno obteve ontem uma nota boa.
   This student received yesterday a score good
   ‘This student received a good score yesterday.

b. Estes livro-s são barato-s.
   these book-PL are cheap-PL
   ‘These books are cheap.’

Notice the words are aligned for easy analysis. Portuguese, like English, uses -s to mark plural on nouns. Once a word bit (called a morpheme) has been identified, it is set off from the root by a hyphen and is indicated in the gloss. A standard set of abbreviations is used to identify the morphemes. This is a plural morpheme and is indicated with PL. Note that small capitals are used to label the morpheme. Observe from this example that, like English, the noun is marked for plural number; however, unlike English, the adjective is also marked for plural number.

In linguistic discussions Korean uses the Yale romanization system. This system may be unfamiliar to you, but it is the de facto standard among Korean linguists. Here is the representation for 민수가 사과를 먹었다. The romanization system is shown in Table 1.1. One deviation from the Yale system (other than the footnote below) that we make is the following. Potentially ambiguous romanizations are disambiguated with an apostrophe. For example 반가 and 방아 are both romanized as "panga". Here, however, they are romanized as "pan’ga" and "pang’a" respectively. Finally, note that this romanization system is used mostly in discussions of morphology and syntax. In phonetics and phonology we will use the International Phonetic Alphabet (IPA), which is introduced in Chapter 3.

(3) Minswu-ka sakwa-lul mek-ess-ta
    Minsoo-NOM apple-ACC eat-PST-DECL
    ‘Minsoo ate an apple.

Don’t worry if you don’t understand the abbreviations. We will discuss them in up coming chapters.

1Note that in the traditional Yale system 부 is transcribed as pu rather than as pwu since 브 is not historically found in Korean. However, because of recent borrowings such as 브랜드 and 블로그 I Romanize 부 as pwu in this manuscript. 블로그 then, is Romanized as pulloku.
In the early 20th century Saussure (1916) distinguished between language (French langue = 언어 or 랑그), speech (French parole = 발화 or 파돌). Often, the French terms langue and parole are used in English academic writing. The term language can refer to ‘a language’, such as French, Korean, Swahili, Cree or English. This sense of the word ‘language’ requires an article. We will discuss what ‘a language’ is below. The other sense of the word language, which concerns us here, is an abstract, stable system that enables speakers to communicate. In this sense, language corresponds to Chomsky’s notion of i-language, short for internal language (Chomsky 1986). Speech or parole refers to the actual execution of language in a given situation. In other words, it refers to the speaker’s use of language. Saussure gave the following analogy. Language is like the rules of chess, whereas speech is like the actual chess moves used by two players in a particular game of chess. In this section we will discuss several properties of Human language (in the sense of langue or i-language).

Hockett (1966) devised a set of properties that characterize human language. As we will see, some of these build on earlier work. Also, this list has since grown and changed over time. Some forms of animal communication have some of these properties, but, so far as is known, not all of them. Let’s go through a partial list of the properties of human language.

The first property is duality of patterning (이원적 대립) and was first proposed by Martinet (1960). This is sometimes called ‘double articulation’ based on a literal translation from the French double articulation. Language is composed of units of meaning and units of sound. The units of sound themselves do not generally have meaning. For example, the English word cat has three sounds: [kæt]. The individual sounds do not have meaning. Furthermore, the individual sounds can be re-arranged to form act or tack.
Arbitrariness refers to the lack of any intrinsic connection between units of meaning (the signified 기의) and the set of units of sound to refer to it (the signifier 기표). We say that the signifier and the signified together constitute the sign (기호). This concept started with Saussure (1916), where he argued that the relationship between the signifier and the signified is arbitrary.

![Figure 1.1: Arbitrary relationship between a cat and the words meaning cat](image)

Although arbitrariness generally holds in natural language, there are instances of iconicity, in which there is some connection between sound and meaning. Onomatopoeic words such as ‘splash’ and ‘meow’ are thought by English speakers to sound like the concept they denote. Likewise Korean words such as 짝짝짝 (clapping), 야옹 (meow), and 칡칙폭폭 (choo-choo, sound of a train) are onomatopoeic.

Creativity is the ability to create a virtually infinite number of utterances based on a finite lexicon and a grammar. It also refers to the ability of the hearer/perceiver to understand any novel utterance (Chomsky, 1966). This was originally called productivity by Hockett, but the term ‘productivity’ has a more specific meaning in morphology. You’ve probably never heard either of the following sentences before, but you understand exactly what they mean.

(4)  
   a. Several bright purple gorillas flew by eating avocados in the shape of sea urchins.
   b. 많은 큰 노란색 두루미가 맛고 주스를 마시면서 오늘 아침에 뛰어 지나갔어요.

Creativity exists at the word level, too. Imagine the nonce word flark. Let’s say it refers to the feeling of happiness combined with being full. Once we know what this word means, then we can say that we feel a bit flarkish or really flarkiful. We can say that we are walking along whistling flarkifully. We will have more to say about this in the next chapter on morphology (word structure).

Displacement refers to the ability to talk about things that are not present, events in the past or future, possible event, untrue events, or objects that don’t exist. While animal communication generally lacks displacement, recall that honey bees are able to communicate sources of nectar that aren’t present, thus demonstrating displacement. Sentences from natural human language easily demonstrate this point.

(5)  
   a. If it hadn’t rained last weekend we could have gone to the beach.
b. 더 열심히 공부했으면 시험을 안 떨어졌겠어요.

Blackfoot (블랙풋), an Algonquian language spoken in Canada (Alberta) and the US (Montana), has an interesting demonstrative system for identifying whether its referent is visible or not. Indoors, it typically means whether the entity is in the room or not. Here are some examples for interest. Note that in Blackfoot demonstratives must appear with proper nouns (unlike English and Korean).

(6) Annáhka  [Blackfoot]
      that.invisible John  áako’towa
      ‘John will arrive.’

In this example the demonstrative encodes the fact that John is not visible (not present). In the next example John must be in the room and visible to the speaker and addressee.

(7) Anná [Blackfoot]
      that  John ihpomma
      ‘John bought something.’ (John must be present)

Learnability refers to the fact that all non-pathogenic humans can acquire language. Upon exposure to a speech community a child will naturally and effortlessly acquire language. This is unlike other skills such as reading or mathematics. Simply being exposed to people reading to doing math is not enough for a child to acquire these skills. It is unclear how much direct speech is required to language acquisition. In some cultures adults barely speak to young children; nevertheless, children still acquire language in these societies (Ochs and Schieffelin, 1984; Cristia et al., 2017). The Critical Period Hypothesis holds that exposure to language must take place before a certain age (usually around puberty), otherwise language acquisition cannot take place or is at least severely hindered. This hypothesis is hard to test and there is at the moment only anecdotal evidence to support it. This evidence comes primarily from feral children who were raised in the absence of any kind of exposure to language.

Genie: The Feral Child

Genie’s case provides some insight into language acquisition. She was rescued at the age of 13 from a highly abusive household. Her father forced her to live in a solitary room with no contact with anyone except him. He never spoke to her, but made only barking noises. When she was found she could not speak. Genie never acquired language but did learn a few words and phrases (Curtiss et al., 1975).

Human languages are organized hierarchically (Chomsky, 1957) rather than linearly. There is no human language with rule such as "move the auxiliary three words to the left". Consider the following English data.
John is happy.

Is John happy?

The child over there is happy.

Is the child over there happy?

To form a yes/no question we see that the auxiliary moves to the front of the sentence. We can see that trying to formulate a rule such as "move the auxiliary one word to the left" or "move the auxiliary four words to the left" does not work. We can illustrate the same point with Korean data.

(8)  a. 철수가 내 사과를 먹었다.
    b. 내 사과를 철수가 먹었다.

A naive interpretation of these sentences is that Korean optionally moves a phrase one word to the left. Consider the sentence in (9-a). If we move the direct object one word to the left the result is unintelligible, (9-b). The correct result is in (9-c).

(9)  a. 이 학생이 내 사과를 먹었다.
    b. 이 내 사과를 학생이 먹었다.
    c. 내 사과를 이 학생이 먹었다.

Instead, we treat the individual strings 이 학생이 and 내 사과를 as hierarchical units. A much more telling example of hierarchicality in human language is shown in the following example.

(10) The girl who is playing is named Rosie.

To form a yes/no question we do not move the first auxiliary, rather we move the auxiliary in the main clause.

(11)  a. *Is the girl who playing is named Rosie?
    b. Is the girl who is playing named Rosie?

Interestingly, when children are acquiring language they never make mistakes like the one in (11-a). It seems children never entertain a linear hypothesis in working out the grammar of the language they are acquiring. We will see more examples of hierarchical structure later when we look at syllable structure.

Human languages exhibit meaningful silence. Although the concept of meaningful silence has been around for a long time, it was suggested as a universal for human language by Kayne (2003). Earlier we discussed the concepts of arbitrariness and dual patterning. The relationship between the units of sound and the units of meaning is arbitrary. One aspect of human language is also meaningful silence, in which a unit of meaning is not paired up with sound. First, consider the following Italian examples.
Here, there are two bits of meaning that concern us. There’s the /-o/, which means singular (SG), and the /-i/, which means plural (PL). The situation in English is slightly different. The bit of meaning for plural in English corresponds to /-s/. However, there is no corresponding bit for singular in English. Consider the following example.

(13) Mary gave the book to John.

This sentence can only mean that Mary gave John one book. That is, the direct object can only be understood as singular, even though there is no overt bit of meaning that means ‘singular’. We say that ‘singular’ in English has a phonologically null correspondence. The following Korean example illustrates the same property.

(14) a. 내가 사과를 먹었어요.
   b. 내가 사과를 먹어요.

Korean has an overt bit of meaning that refers to past tense (-었어요). There is no overt bit of meaning for present tense, however. Nevertheless, example (14-b) can only be understood as present tense.

Recursion in language refers to the ability to embed smaller structures in larger structures repeatedly with the same mechanism (Hauser et al., 2002). Here are some examples.

(15) a. John’s hat
    b. John’s brother’s hat
    c. John’ brother’s friend’s hat

(16) a. This is the house that Jack built.
    b. This is the malt that lay in the house that Jack built.
    c. This is the rat that ate the malt that lay in the house that Jack built.
    d. This is the cat that killed the rat that ate the malt that lay in the house that Jack built.
    e. This is the dog that worried the cat that killed the rat that ate the malt that lay in the house that Jack built.

(17) a. 민수가 사과를 먹었다.
b. 영희는 민수가 사과를 먹었다고 말했다.
c. 철수는 영희가 민수가 사과를 먹었다고 말했다고 생각한다.

Let’s look at a simple demonstration. A sentence typically consists of a subject and a predicate. The subject is a noun phrase (NP) and the predicate is a verb phrase (VP). Here is a simple depiction. Do not worry about the formalism, yet. We will discuss it in more detail in Chapter 5.

(18)
```
\[ S \rightarrow NP \text{ VP} \]
```

We can make a rule for the formation of the S node as follows.

(19)
```
S \rightarrow NP \text{ VP}
```

Notice that the VP is internally complex. We can write a rule for the VP as follows.

(20)
```
VP \rightarrow V \text{ NP}
```

This gives us the following structure.

(21)
```
\[ S \rightarrow NP \text{ VP} \]
\[ NP \rightarrow \text{ Susan} \]
\[ VP \rightarrow V \text{ NP} \]
\[ V \rightarrow \text{ ate} \]
\[ NP \rightarrow \text{ an apple} \]
```

Now, some verbs do not take an NP as an object. Some verbs take a clause as an object, like in the Korean examples in [17]. Let’s look at an example.

(22)
```
a. John thinks Mary ate an apple.
b. VP \rightarrow V \{NP, S\}
```
We can now see how recursion is produced by this small set of rules. VP maps to V and S. S maps to NP and VP. VP maps to V and S, and so forth.

This ends the discussion on properties of human language. The next section discusses the difference between how linguists study language and how people attempt to control language.

1.3 Prescriptivism versus Descriptivism

Prescriptivism is the practice of dictating, by virtue of some authority, acceptable standards of language use. Prescriptivist attitudes are most often applied to written language, but are often applied to spoken language. Examples of prescriptive authorities include the 국립국어원 and the Académie Française. A prescriptive grammar is a collection of the prescribed rules of a given language, and usually includes some common proscribed structures (i.e., sentences). Almost always are the proscribed examples in prescriptive grammars those that are found in common use. A prescribed structure is one which is deemed acceptable by some authority and a proscribed structure is one which is deemed unacceptable. A prescribed sentence is given in example (23), followed by a proscribed sentence.

(23) a. To whom do you wish to speak?
   b. Who do you wish to speak to?

Many “rules” of English grammar are actually not strongly proscribed by reputable grammars. For instance, you may have heard that split infinitives are to be avoided; however, since 1983 the Chicago Manual of Style has ceased frowning on constructions like to boldly go where no one has gone before. Many grammar mavens also decry the practice of stranding prepositions, even though these to are tolerated by contemporary style guides, especially in informal settings. Despite the fact that speakers of English have been stranding prepositions for over 700 years, the myth still persists that stranding prepositions is to be avoided in any situation, no matter how informal.

Let’s consider a specific example or a prescriptive rule in more detail. We will undertake the deconstruction
of a common prescriptive rule starting with an examination of the following sentences.

(24)  a. Everyone forgot their homework yesterday.
     b. Someone left their dirty dishes on the table.
     c. Who remembered to bring their sleeping bag?

All of these sentences are proscribed in English by some prescriptive authorities because of the putatively faulty agreement between the subject and the possessive pronoun modifying the object. *Everyone, someone, and who* are singular, and thus require singular pronouns.

a. Everyone forgot his homework yesterday.
b. Someone left his dirty dishes on the table.
c. Who remembered to bring his sleeping bag?

**Generic They in Literature**

But to expose the former faults of any person, without knowing what their present feelings were, seemed unjustifiable. [Austen, Pride and Prejudice]

There’s not a man I meet but doth salute me, As if I were their well acquainted friend. [Shakespeare, Comedy of Errors, Act IV Scene 3]

The following website lists several examples of generic they from English literature.

http://www.crossmyt.com/hc/linghebr/austheir.html

Since about the 1960s the use of *his* in these sentences has been considered by some writers to be sexist. Various alternatives have been suggested such as *his or her, s/he*. Only *he or she* has ever really caught on, and when it does appear, it is found only in written English. People generally revert to generic *they* in speech. Nevertheless, generic *they* is proscribed by some authorities. One purported justification for this is that it is more logical if elements in a sentence agree for number, person, and gender. There are several situations where this simplistic view of agreement does not hold, however. English ‘royal we’ referring to a 1st person singular referent is a well-known example. French *on* is third person singular, however, it can agree with first person singular elements.

(25)  On a oublié notre musique.
      one has forgotten our  music
      ‘We forgot our music.’

[French]

Also, German *Sie* is third person plural, but can be used to refer to second person singular referents in formal situations. In French, the 2nd person plural pronoun is used to politely refer to an individual addressee. It is used when addressing strangers or someone older than the speaker. Consider the following example
Observe that *vous* triggers plural agreement on the verb, but singular agreement on the adjective. All of these forms are considered standard use of the language and are sanctioned by the Académie Française (the governing authority on the prescribed use of the French language). Thus, the argument based on the logic of agreement is flawed at best, since there exist numerous examples that violate this principle. Let’s look at the history of the word *they* (and its derivatives *them*, *their*, etc.) briefly. *They* was borrowed from Scandanavian into English in the 14th century. Before that, there were separate plural forms for *he* and *she*. Almost immediately upon its introduction into English, *they* was used as a generic pronoun as in the sentences at the beginning of this section. In fact, generic *they* with a singular antecedent can be found in the works of Austen, Shakespeare and many other writers. It was not until the end of the 18th century until a prescriptive rule against generic *they* was formulated. The following timeline of arguments shows the development towards the proscription of generic *they*:

Some will set the Carte before the horse, as thus ... yet in speaking at the least, let us kepe a natural order, and set the man before the woman for maners Sake. (Wilson [1553])

This rule dictates that the order *Mary and John* is incorrect. The correct order is *John and Mary*.

The Relative shall agree in gender with the Antecedent of the more worthy gender... The Masculine gender is more worthy than the Feminine. (Poole [1646])

The masculine Person answers to the general Name, which comprehends both Male and Female; as Any Person, who knows what he says. (Kirby [1746])

This rule advocated the use of gender-neutral *he* and was adopted into British law by an act of Parliament (Weatherall [2005] 15).

Pronouns must always agree with their antecedents, and the nouns for which they stand, in gender, number, and person; ... Of this rule there are many violations ... (Murray [1795])

Similar attitudes were also expressed in the United States.

... the fact remains that his is the representative pronoun, as mankind includes both men and women... (White [1880])

It is clear, then, that overtly sexist attitudes were the driving force behind the proscription against generic *they* and the prescription of gender-neutral *he*. Furthermore, many proscribed sentences are difficult to make prescriptively correct while still maintaining the given structure of the sentence:

(27) a. Everyone forgot their homework, but they remembered their lunch.

b. ?*Everyone forgot his homework, but he remembered his lunch.
Descriptivism is the approach that linguists take in the study of language. Linguists do not prescribe how language is to be used (although we are often asked to). Rather, we describe, record and analyze the way language is actually used. This includes language in all its forms, including its dialects (the form of a language used in a geographically determined area), sociolects (the form of a language used by a given socio-economic class) and even idiolects (the particular form of a language used by a specific person). A common problem in descriptive linguistics is eliminating the observer effect. If people are conscious of the fact that their language is being observed they may modify their speech. Perhaps they may try to adhere to perceived prescriptive norms, or they may adjust their speech by speaking more slowly or clearly to help the researcher. There are techniques, however, for eliciting naturalistic data, both in the field and in experimental settings.

1.4 Language and Thought

Language is intimately entwined with other cognitive aspects of the human mind/brain. The earliest studies in linguistics were conducted by Pāṇini and Aristotle. Modern studies in linguistics can be traced back to the 18th century. This long history has culminated into a collective knowledge of the science of language, which we present in this reading. Despite the general agreement on many aspects of human language, there remain, as in many fields, significant points of disagreement. These points of disagreement most notably relate to the relationship between language and thought. We review some of these points of disagreement here.

1.4.1 Linguistic Determinism

The Sapir-Whorf Hypothesis is one of the better known concepts among the general public; however, it has also generated a significant amount of controversy. This hypothesis is a bit of a misnomer as Edward Sapir and Benjamin Lee Whorf never co-authored a paper, and they never put forth such a hypothesis as described here. Nevertheless, their writing have strongly influenced our understanding of the concepts described here, so the name has become standard. In simple terms the Sapir-Whorf Hypothesis holds that language shapes our range of thoughts. There is a strong version and a weak version.

(28) Sapir-Whorf Hypothesis
   a. Strong Version: Linguistic Determinism holds that language entirely determines our cognitive processes.
   b. Weak Version: Linguistic Influence holds that language influences our range of cognitive processes, but is not deterministic.

It is now generally accepted that Linguistic Determinism is false. The unfortunate case of Genie (see page 5) shows that Linguistic Determinism cannot be correct. Genie was raised completely sheltered from all forms of language; however, she is capable of thought. Early conceptions of Linguistic Influence suggested that speakers
The Sapir-Whorf Hypothesis has generated discussion in several areas recently, two of which we discuss here. The first concerns colour terms. Berlin and Kay (1969) propose a universal hierarchy of colour terms as follows. If a language has only two colour terms, the two will be ‘black’ and ‘white’. If a language has three colour terms, the third will be ‘red’. If a language has four colour terms, the fourth will be either ‘green’ or ‘yellow’. Figure 1.2 shows the full range of possibilities. In languages with only two words for colour, ‘white’ and ‘black’, the word for ‘white’ generally refers to any light colour and the word for ‘black’ generally refers to any dark colour. In effect, all visible colours will generally be associated with some name in a given language. Since the perception of colour is a biological phenomenon, the naming of colours is thought to be universal.

Levinson, however, has suggested that Yélî Dnye is an exception to the universal theory of colour naming proposed by Berlin and Kay. Levinson shows that Yélî Dnye fails to have names for certain portions of the visible spectrum, a property not predicted by Berlin and Kay’s theory.

Yélî Dnye is a language isolate spoken in Papual-New Guinea. It has an unusually large inventory of sounds and postpositions.

Another property of language that features prominently in discussions of linguistic determinism is direction terminology. Languages such as Korean and English use egocentric directions such as ‘left’ and ‘right’, which change according to the speaker. To give someone directions you might say, "Walk 200m then turn left." Some languages, such as Guugu Yimithirr, use geographic directions. Only compass directions such as ‘south’ and ‘northeast’ are used. A Guugu Yimithirr speaker doesn’t talk about a "left" hand or "right" hand, but a "south" hand or "north east" hand, depending on the orientation of the speaker. As such, speakers of Guugu Yimithirr are argued to have a much better sense of direction than speakers of languages that use egocentric directions. See Haviland (1998) for a discussion on this topic.
There are two persistent myths that are related to the Sapir-Whorf Hypothesis. Although both of these myths have been debunked by the linguistics community at large, they often surface in popular literature from time to time. The first is less well known, but stems directly from Whorf’s work. Whorf (1964) had originally claimed that Hopi has "no words, grammatical forms, construction or expressions that refer directly to what we call 'time' [and that Hopi had] no general notion or intuition of time as a smooth flowing continuum in which everything in the universe proceeds at equal rate, out of a future, through the present, into a past.” From this quote the myth persists that the Hopi people have no concept of time. Malotki (1983) published an in depth study of tense and time in Hopi showing that Whorf’s original claims do not hold.

The other well known myth concerns the number of words for snow in Inuktitut. From time to time one sees headlines claiming an unusually large number of words for ‘snow’ in "Eskimo". First note that the term Eskimo is outdated and considered offensive. The aboriginal people living in the far north of North America prefer to be known as Inuit. The language they speak is a set of dialects known as Inuktitut (see Figure 1.3). In actuality, Inuktitut has approximately the same number of words for ‘snow’ as English does. The history of this myth is chronicled in Pullum (1991).

### 1.4.2 Linguistic Nativism

Nativism holds that certain cognitive aspects are innate or in-born. Such properties are expected to be uniform across the species. Linguistic Nativism is also known as **Universal Grammar** holds that the child is endowed with certain grammatical principles (Chomsky, 1965). Such a theory of grammar is typically expressed within the **Principles and Parameters Framework** (Chomsky, 1981). This framework works as follows.

(29) Principles and Parameters

a. Principles: Those properties of grammar that are invariant across the species and are potentially found in all languages.

b. Parameters: Those properties of grammar that vary across the species and must be determined during the process of language acquisition.

Universal Grammar is described in more detail in other parts of this chapter, so we eschew an in depth discussion here.

---

Figure 1.3: Map of Inuktitut Dialects
Pirahã has figured prominently in discussions of Universal Grammar. Everett (2005) has argued that Pirahã is an exception to Universal Grammar as it does not allow for recursion (see page 7).

Pirahã is a Mura language spoken in Amazon in Brazil. Despite the small number of native speakers (approximately 400), the language is healthy as most speakers are monolingual. Despite the claims that the syntax of Pirahã challenges Universal Grammar, the phonology and morphology of Pirahã do conform to our understanding of human language.

1.5 Core Areas of Linguistics

Linguistics, simply put, is the study of language. There are six core theoretical areas to the study of language:

- phonetics – the study of sounds
- phonology – they study of sound patterns
- morphology – the study of how words are put together
- syntax – the study of how sentences are put together
- semantics – the study of meaning
- pragmatics – the study of how meaning interacts with real-world knowledge

Let’s look at each of these a bit more specifically. Phonetics is the study of the sounds and signs of language. We include signs here because there exist many signed languages around the world, which have the same kinds of structure as spoken languages (Klima and Bellugi, 1979; Emmorey et al., 2003). Phonetics asks questions such as how sounds are produced and perceived. It also asks what the physical properties of the sounds are. What kinds of sounds exits in natural languages? How do the physical properties of the sounds affect the sounds in a particular language?

Phonology asks how sounds are organized in human language. How does the presence of one sound affect nearby sounds? What kinds of sounds can co-occur in a syllable? For instance, why is plink a possible word in English, but not tlink? What are the properties of sounds that are important in a given language?

Morphology ask how words are put together. What are the parts of the words that have specific meanings? Do they change at all? For instance, baker and driver both have an -er in them. What does that mean? What kinds of words are possible and what kinds are impossible? Why does cats end with an s-sound while dogs ends with a z-sound (even though they are both spelled with an ‘s’)?

Syntax asks how words are organized in a sentence. Why is English subject-verb-object while Korean is subject-object-verb? How can words move around in a sentence? Why are questions words always at the begin-

ning of a sentence in English, but not in Chinese? What kinds of restrictions are there in moving words around. How come we can say *wash up the dishes* but not *wash up them? Why does it have to be *wash them up?*

Semantics asks how meaning is computed from a sentence. How do we understand that the sentence *The King of France is bald* when France no longer has a king? How do we get the different meanings in a sentence such as *Mary saw the boy with binoculars*? How are synonyms handled in a language?

Finally, pragmatics asks how real world knowledge is incorporated into language use. If you ask a friend to go for dinner and she replies, *"I have to work late"*. How do you understand this to mean she cannot go? She implied that she cannot go for dinner because of her work obligations. How are implicatures such as this handled in natural language?

In addition to the core areas discussed above, linguists also work in the following domains:

- sociolinguistics – the study of language in a social context
- historical linguistics – the study of how language changes over time
- psycholinguistics – the study of how the mind processes language
- neurolinguistics – the study of how the brain processes language
- first language acquisition – the study of how language is acquired as a child from birth
- second language acquisition – the study of how adults and children acquire a second language
- documentation and field linguistics – the study of how to record and document endangered languages

Sociolinguistics is concerned with how language varies across speakers within a society. It is obvious that not everyone who speaks the same language speaks exactly the same way. There are differences based on gender, age, region, socio-economic status, sexual orientation and race. The divisions are not always clear cut either. In fact, there is typically a great deal of variation. What kinds of variation are possible?

Historical linguistics is concerned with how languages change over time. Old English is nearly incomprehensible to a modern speaker of English. Why is this? Historical linguistics is also concerned with the relationships among languages. For instance English, German and Hindi are all related; however, English and German are much more closely related in terms of when the languages diverged from a common ancestor. Likewise, Mandarin, Cantonese and Burmese are also all related as members of the Sino-Tibetan family, but again, Mandarin and Cantonese are much more closely related.

Psycholinguistics is the study of how the mind processes language. How are words stored in the mind? What do slips of the tongue tell us about language? Do bilinguals process language differently from monolinguals? Do they have two separate grammars in their minds or one big grammar?

Neurolinguistics is the study of how the brain processes language. With the narrowing gap between traditional psychology and cognitive science, the differentiation between psycholinguistics and neurolinguistics is becoming quite blurred. Nevertheless neurolinguistics is concerned with the brain structures that are implicated...
in language use. What can go wrong with language if those parts of the brain are injured? What do speech defects tell us about the organization of language and of the human brain?

First language acquisition is concerned with how children acquire their first language so effortlessly in the absence of explicit instruction. This question goes back to Plato, who observed that children acquire language without any instruction. What kinds of universal properties are there to first language acquisition? It has been observed that children around the world acquire language pretty much the same way. Why is this?

Second language acquisition is concerned with how adults and older children acquire a second (or third or fourth) language later in life. What kind of effect does the speakers first language have on acquiring a second language? Why is it so much more difficult to acquire a language as an adult rather than as a child?

There are over 7000 languages in the world, many with very little or no documentation. Furthermore, many of these are in danger of becoming extinct. Field workers work directly with speakers of these languages to document them. Much of this work was started in the early 20th century in North America by linguists such as Leonard Bloomfield and Mary Haas. Often this is done with the purpose of revitalization in mind. Communities are becoming aware of the precarious states of their language and sometimes enlist the aid of linguists to help with revitalization efforts.

There are two major camps in the theoretical analysis of language: functional linguistics and formal linguistics. Functional linguistics is primarily concerned with the relationship between language and function (gathering information – question formation; introducing a new topic into conversation; highlighting topics of conversation, etc.) and between language and the real world (politeness, social relations between speakers, etc.). Functionalism grew from the Prague School in the 1920’s and was later developed by Michael Halliday. Role and Reference Grammar is one of the main theories of functional syntax, despite its formal appearance. We will not be dealing with functional syntax in this textbook. Historically, formal linguistics is concerned with the structural properties of language and was concerned only with the automata for piecing words together to form sentences (in the case of formal syntax). Formal linguistics is often seen as starting with Noam Chomsky’s dissertation, summarized in his published version [Chomsky] [1957], although it has its roots in Pāṇini. Today, formal linguistics does take meaning into account. Here’s an example of where the two approaches diverge.

(30) Can you close the window?

This sentence is usually taken as a polite request to close the window, although this is not part of the literal meaning of the sentence. Functional linguistics are interested in this communicative feature of the sentence, whereas a formal linguist is traditionally only interested in the literal meaning and the structure (although a branch of linguistics known as formal pragmatics does address the communicative features as well). Formal linguistics is concerned with the hierarchical properties of language. This textbook will deal exclusively with formal linguistics, although we will address those issues of communication that are of interest to functional linguists. Specifically, we will deal mostly with Generative Grammar. What do we mean by ‘generative grammar’? To answer this, we must know what we mean by ‘grammar’. Grammar is the tacit knowledge that each
Each one of us has an innate ability to acquire whatever language we are surrounded with as we grow up. If you grew up in an English-speaking environment, you will know that (31) is an ill-formed sentence of English without ever having had any explicit instruction indicating that (31) is ungrammatical. (Recall that *=ungrammatical.) Indeed, most English speakers would be hard-pressed to explain why (31) is bad... they just know that it is. Likewise, a speaker of French knows that (32-a) is ungrammatical. The grammatical version is in (32-b).

(31) *Who do you wonder whether Peter likes?
(32) French
   a. *Qui crois-tu que viendra?
      who think-you that will.come
      ‘Who do you think will come?’
   b. Qui crois-tu qui viendra?
      who think-you who will.come
      ‘Who do you think will come?’

1.6 Properties of Universal Grammar

Unlike other learned behaviours, all humans acquire language. All non-pathogenic humans and no animals acquire language. In rare, tragic cases where people are sheltered from all forms of human language until after the onset of puberty, language acquisition ceases to be an effortless activity, and proceeds slowly. In these situations, the individual never acquires an adult level of fluency in any language. A well-known example is Genie (see page 5).

There have also been many attempts to teach language to higher primates (Gardner and Gardner 1969). None of these attempts has ever been successful (Seidenberg and Petitto 1979). Non-human primates such as Washoe and Nim Chimsky were able to acquire a set of signs; however, it was never clear whether these animals employed a set of recursive rules productively to produce language. Rather, given the small number of signs the primates knew, they would stumble upon somewhat consistent patterns from time to time.

Before ending this chapter we will discuss one other theory of language acquisition and some of the problems that have been noted for it. The proposal we discuss is learning by analogy. It has been suggested that children learn language by analogy. This seems like a reasonable possibility given the large variety of generalizations that can be made about grammar. For instance, the first analogy the child could draw is that direct objects are always placed after verbs. This would account for the two-word stage during language acquisition, which consists usually of sequences of a verb and a direct object. Under this approach, there is no underlying
universal grammar. The child simply learns language the same way he or she learns any other aspect of life. Consider, however, the following examples (Gleitman [1994]).

(33) a. Susan painted the red barn.
   b. Susan painted the barn red.
   c. Susan saw the red barn.
   d. *Susan saw the barn red.

Under the analogy approach it is expected that children would produce speech errors during acquisition along the lines of the last example – however, children never seem to produce errors of this kind. This is unexpected assuming that children make use of analogies during language acquisition, but is compatible with an approach that assumes UG. In short, UG doesn’t make structures available that could produce the last sentence, which is why children never make errors of this kind.

1.7 Languages and Dialects

Linguists have precise definitions for the terms ‘language’ and ‘dialect’, which differ from what the general public sometimes uses. The definitions used by linguists are as follows.

If two distinct speech forms are mutually intelligible they are two dialects of the same language.

If two distinct speech forms are not mutually intelligible they are two separate languages.

In casual speech people often use the word language to refer to the standard form of a language and the word dialect to refer to a non-standard regional variety. When linguists use the term language they are referring to a set of dialects, not the standard form of the language. The Standard form of a language is an artificial form that is used for pedagogical and official purposes. We’ll come back to this point below.

The picture painted above has some complications. Imagine three hypothetical towns ten kilometers apart each, A, B and C. Let’s say they all speak the same language, but the dialects are slightly different. Let’s call them dialects A, B and C. Now, let’s say in our hypothetical example that speakers of dialect A and B can understand each other and that speakers of dialects B and C can understand each other. However, speakers of dialects A and C cannot understand each other.

This scenario exemplifies a dialect continuum (방언 연속체). Dialect continua are extremely common around the world. One well-known dialect continuum is the Dutch-German continuum. Someone from Amsterdam and someone from Munich cannot understand each other – but people from the smaller boarder towns on the Dutch/German border can understand one another. Another dialect continuum is Cree-Ojibway. Cree and
Ojibwe are spoken from Alberta to Quebec, but not all dialects are mutually intelligible. Neighbouring dialects tend to be mutually intelligible, but dialects that are spoken in communities far apart tend not to be.

We often talk of Chinese as a language. Linguistically, Chinese is a subfamily of languages (it is a branch of the Sino-Tibetan family). The most widely spoken Chinese language is Mandarin (northern China and Taiwan). Another familiar Chinese language is Cantonese (Hong Kong and Guangdong province), also known as Yue. A 10-year-old child from Beijing and a 10-year-old child from Hong Kong will not understand a word the other is saying – thus, Mandarin and Cantonese are two different languages. More importantly, there is no dialect of Mandarin that is mutually intelligible with any dialect of Cantonese.

Mandarin, however, is the official language of China. Sometimes, Mandarin is simply referred to as “Chinese”. Because of the official status of Mandarin, all the other Chinese languages are sometimes referred to as “dialects” – but this is a different use of the word “dialect” than described above. In this course, we will use the terms we introduced above.

A sprachbund (언어동조대) is a concept related to the current discussion but crucially differs from a dialect continuum. A sprachbund is a set of unrelated languages or dialects that share properties because of their geographical proximity.

One well-known sprachbund is the Balkan Sprachbund and includes Albanian, Bulgarian and Romanian (among others). These languages, though distantly related, share certain properties as a result of contact with one another over a long period of time. One such property is the lack of infinitives. When languages share properties as a result of contact we call these similarities a sprachbund effect. Sprachbund effects can sometimes give rise to the mistaken notion that a group of geographically related languages are historically related.
Exercises

**Question 1.** Consider the properties of communication and human language. What sets language apart from communication? Do you think you can make a case for higher primates possessing language, or just a more complex form of communication?

**Question 2.** Consider the following sentences. Do they break any prescriptive rules of English grammar? Are they acceptable to you as a native speaker? If not, try to determine the source of the unacceptability. If you are not a native speaker of English, try to find someone who is.

1. Who did you give the book to?
2. Who do you wonder whether Peter gave a book to?
3. John persuaded Mary to go to the opera.
4. John promised Mary to go to the opera.
5. Me and John went to the movies.

**Question 3.** How do the following two short dialogues argue against the hypothesis that language is learned by positive reinforcement (first dialog from Uriagereka, 1998:7)?

Child (4 yrs): Eat shit!
Father: (shouting angrily) Go to your room!

***

Child (2 yrs): Me eated it.
Mother: You certainly did, cutie-pie. (hands child another cookie)
Further Reading

- **Ghosh (2010)** - wonderful short book on language and prescriptivism
- **Roberts (2017)** - excellent lay person introduction to language and linguistics
- **McCulloch (2019)** - a lively discussion on how the internet has shaped language
- **Adger (2019)** - an excellent lay person introduction to the science of language

**online resources:**

- All Things Linguistic: [https://allthingslinguistic.com/](https://allthingslinguistic.com/)
- Babel: The language magazine [https://babelzine.co.uk/](https://babelzine.co.uk/)
- Language Log: [https://languagelog.ldc.upenn.edu/nll/](https://languagelog.ldc.upenn.edu/nll/)
- Superlinguo: [https://www.superlinguo.com/](https://www.superlinguo.com/)
Chapter 2

Morphology: The pieces of words

By the end of this chapter you should be able to:

- be able to identify pieces of words in your native language,
- know the major categories of morphemes found in language,
- be able to describe the properties of morphemes found in language
- be able to explain the difference between derivational and inflectional morphology,
- know and be able to explain the basic properties of compounds,
- understand how new words are introduced into language.

2.1 Words and their Pieces

Anyone who has thought about words even for a few moments will undoubtedly notice that certain pieces of words tend to recur. Consider the following two sets of English and Korean words.

(1) a. friend, friendly, friendless, friendlessness, friendliness, friendship
    b. telephone, phonology, phonetics, phonograph, xylophone, saxophone

(2) a. 학생 haksayng ‘student’ 대학교 tayhakkyo ‘university’ 학원 hakwen ‘hagwon’ (a kind of private school)
    b. 경제력 kyengceylyek ‘economic power’ 영도력 yengtolyek ‘leadership’ 중력 cwunglyek ‘gravity’

The first set contains the form friend in each word. Each word in turn has something to do with the concept of ‘friend’. The second set contains the form phone. Here, the link between the words is more tenuous, but all the words have to do with the concept of sound. Likewise, in the Korean examples the first set contains 학 (hak), which has to do with studying, and the second set contains 력 (lyek), which has to do with power or force. Virtually all languages contain words with internal complexity of the kind found above. Some languages contain words with a great deal of internal complexity, and some languages contain much less complexity. Let’s
look at some sets of words in some other languages. See if you can pick out the common meaning associated with the recurring word-bit in red.

(3) a. Mandarin: péngyǒu ‘friend’ yǒuyì ‘friendship’ yǒuhǎo ‘friendly’ zhànyǒu ‘ally’
   b. Swahili: mwanaamia ‘club member/party member’ mwanaanches ‘citizen’ mwanaakijiji ‘villager’
   c. Hungarian: ír ‘to write’ írás ‘act of writing’ irat ‘document’ irodalom ‘literature’
   d. Kubeo: põe ‘person’ põe#kil-be ‘to be alone’ põe#õa ‘to deliver a baby’

**Kubeo** is a Tucanoan language spoken in Columbia and Brazil. It has a complex morphology and is tonal. The Kubeo are a small, ethnically diverse society in the northwestern Amazon. The map below shows the location of living and extinct Tucanoan languages.

**Oneida** is an Iroquoian language spoken in Canada (Ontario) and the United States (New York and Wisconsin). It has an extremely complex morphological structure and agreement system. The Oneida, along with the other Northern Iroquoians constitute the league of the Six Nations, which played an important role in the history of Canada and the US. The stop sign to the right is an example of Oneida.

Let’s look at the English word ‘friendliness’. Observe that it can be broken down into /friend/, /-ly/, and /-ness/. It is merely a peculiarity of English orthography that /-ly/ changes to “li” in the written form. If you listen to the word, there is no change in the pronunciation. The piece /-ly/ changes nouns into adjectives (friendly, fatherly, lovely, etc.). The piece /-ness/ changes adjectives into nouns, and indicates a state or property (friendliness, calmness, awkwardness, etc.). These are the smallest bits of meaning that the word ‘friendliness’ can be broken down into. We cannot split ‘friend’ apart into “fr” and “iend” and ascribe meaning to them. These smallest bits of meaning are called morphemes. When we give linguistic examples morphemes are separated by hyphens as in examples (4), (5), and (6) below, and are given a “gloss” – a label indicating the meaning or function of each morpheme. There are usually standard abbreviations for the glosses (such as PL for plural), which can be found in the appendix on page 302.

**Morpheme:** The smallest segmental unit of meaning within a word

(4) English

   a. friendliness
      friend-ly-ness

Be careful not to confuse morphemes with syllables. A morpheme can contain one or more syllables as in the following words:

- **dog** 개
- **candle** 소금
- **banana** 개구리

Also, a single syllable can contain more than one morpheme

- **cat-s** 잡 (자-로)
- **play-ed** 갔다 (가-다-다)
b. sleeplessness
  sleep-less-ness

(5) Korean
  a. 가셨습니다
     ka-si-ess-sup-ni-ta
     가-시-있-습-니다
     go-HON-PST-FORMAL-INDIC-DECL
     ‘went’ (polite, formal style, see [210])

(6) Mandarin
  a. 言语学者们
     yú-yüe-xüe-zhˇe-men
     linguistics-study-person-PL
     ‘linguists’

Morphology is the study of how these pieces of words—the morphemes—are put together to form complete and complex words. Thus, we say friendliness but not *friendlessly. In Korean we say ka-si-ess-sup-ni-ta (가셨습니다), but not *ka-ta-si-sup-ess-ni (*가다시었습니다). In Mandarin we say t¯a-men (‘we’), but not *men-t¯a.

Every language has rules for putting morphemes together to form words. From the discussion below, we will see that there are various ways to classify morphemes. One basic distinction is between lexical morphemes (more often simply called roots, a convention we will follow) and grammatical morphemes. The root of the word typically gives the core meaning of the word. It identifies the referent or concept in the real world. Thus, the knowledge of roots is entwined with real-world knowledge. Grammatical morphemes, on the other hand, provide grammatical meaning such as tense, plural, categorial distinctions such as noun and verb, negation, and a wide range of other meanings. Consider the following examples.

(7) a. friend-li-ness dis-heart-en-ing

b. 먹-었-다  짓-누르-다

Roots identify an object or concept in the real world, and they involve real world knowledge. For instance, knowing the difference between zebras and gazelles involves specialized knowledge about animals, but not about English grammar. On the other hand, grammatical morphemes encode grammatical concepts such as tense and plurality. Thus, knowing the difference between ‘zebra’ and ‘zebras’ requires grammatical knowledge. We say that the use of grammatical morphemes forms part of the grammatical competence of the speaker.

The repository of this learned information is the mental lexicon - a model of all stored linguistic information of morphemes. There are certainly numerous words in the mental lexicon, with entries we can notate as follows. Note that the definitions are approximate.
The word in [8] is **monomorphemic** – that is, it consists of one morpheme. What about **polymorphemic** words (words with more than one morpheme)? Is the word *cats* stored in the mental lexicon of the English speaker or the word 가셨습니다 stored in the mind of the Korean speaker? We have the following questions. Are individual morphemes also stored in the mental lexicon? or only words? First, let’s consider how many words there are in a given language. It is actually quite difficult to estimate the number of words in any language. Do we count *cat* and *cats* as two words or one? Likewise, do we count the Korean words 가셨습니다 (went, formal, honorific), 가요 (goes, informal polite), 가다 (to go) as three words? Think for a moment about a plausible model for the mental lexicon. If we store whole words, then we store not only *cat*, *dog*, and *zebra*, but also *cats*, *dogs*, and *zebras*. Another possibility is that English speakers store *cat*, *dog*, and *zebra*, on the one hand, and just the plural morpheme itself, *-s*, on the other, with the knowledge that this morpheme attaches to count nouns. In the remainder of this section we will consider these two possibilities.

An important experiment, the Wug Test, [Berko (1958)](#), showed that even very young children have mental access to morphemes, the pieces of words. Upon being introduced to the nonce word *wug*, children had no problem forming the plural form *wugs* on their own. This was the first line of experimental evidence that the mental lexicon does not store only fully formed words, but must also store individual morphemes. These morphemes are then put together as needed to form words, much the same way words are put together to form sentences as needed. The goal of this chapter is to introduce the principles of word formation.

We must be clear to distinguish between synchronically active morphology and historical remnants. To take two clear examples, consider the following pairs of English words.

(9) a. cats
   b. lady

The morphological analysis of the first example is clear, but not so for the second example.

(10) a. cat-s (CAT-PL)
   b. la-dy (> hlæf-dige = loaf-kneader, i.e., someone who makes bread)

As discussed with the *wug*-experiment above, the plural morpheme is accessible to contemporary English speakers. The word *lady* in Old English was a compound, *hlæf-dige*, where *hlæf* means ‘bread’ and is cognate with Modern English ‘loaf’, and *dige* means ‘knead’. Nevertheless, no one would suggest that a speaker of Modern English stores the English word *lady* as a compound with two analyzable units. To be clear, the word *cats* has two morphemes in Modern English, whereas the word *lady* is a single morpheme in Modern English. In this chapter, we are concerned only with morphology that is active in the speaker’s linguistic competence.
and not with historical facts. Nevertheless, there are cases where the distinction is not so clear-cut (as with deceive, perceive, etc. discussed below). Note that some recent research shows that some compounds, such as tombstone, and suppletive forms, such as give, are processed compositionally as "tomb" + "stone" and "give" + PST, respectively (Flick et al., 2018; Lewis et al., 2011; Stockall and Marantz, 2006).

2.1.1 Morphemes and Words

Morphemes are the smallest bits of meaning within words. To identify morphemes, several words or sentences must be compared. Ideally, once we have a large set of data, we compare words that differ in one aspect of meaning and see how the words differ in form. We surmise that the difference in form corresponds to the difference in meaning. This process is called morphemic analysis. As a simple illustration, consider the following Indonesian data.

(11) Indonesian (Sneddon et al., 2012)

<table>
<thead>
<tr>
<th>pekerja</th>
<th>'worker'</th>
</tr>
</thead>
<tbody>
<tr>
<td>perenang</td>
<td>'swimmer'</td>
</tr>
<tr>
<td>pedagang</td>
<td>'trader'</td>
</tr>
<tr>
<td>petenis</td>
<td>'tennis player'</td>
</tr>
</tbody>
</table>

Looking at these data, we see they all have a common morpheme, pe-, which we surmise indicates an agentive, 'do-er' meaning. It seems to correspond to the English -er morpheme also present in the data. Thus, the first word contains two morphemes: pe-, an agentive morpheme and kerja, the root meaning 'work'. Of course, a real analysis of Indonesian morphology would consist of dozens of words to make sure we haven't simply stumbled across four words that happen to begin with 'pe'.

To understand the next data set we need to understand the notion of agreement. In many languages the form of a verb (and sometimes a noun or adjective) varies with the subject of the verb. In some languages, it varies with the subject and the object of the verb. Consider the following Spanish data.

(12) Spanish Agreement

<table>
<thead>
<tr>
<th>com-o</th>
<th>'I eat'</th>
<th>com-emos 'we eat'</th>
</tr>
</thead>
<tbody>
<tr>
<td>com-es</td>
<td>'you eat'</td>
<td>com-eis 'you.(PL) eat'</td>
</tr>
<tr>
<td>com-e</td>
<td>'s/he eats'</td>
<td>com-en 'they eat'</td>
</tr>
</tbody>
</table>
The form of the verb *comer* ‘to eat’ changes according to the subject of the sentence. Note, also, that this is true in English, too, but to a much lesser degree. For the third person singular form an /-s/ appears on the present tense form of the verb in English. Consider the following Spanish examples.

(13) a. Yo y María com-emos las manzanas.
    I and Maria eat-1PL the apples
    ‘Maria and I eat apples.’

b. José com-e las manzanas.
    Jose eat-3.SG the apples
    ‘Jose eats apples.’

In the first example the verb agrees with the first person plural subject. In the second example the verb agrees with the third person singular subject.

Consider now the following Czech data. Here, we see a small problem if we execute the above procedure blindly.

(14) Czech

<table>
<thead>
<tr>
<th></th>
<th>‘I plant’</th>
<th>‘we plant’</th>
</tr>
</thead>
<tbody>
<tr>
<td>sázím</td>
<td>‘I plant’</td>
<td>sázíme</td>
</tr>
<tr>
<td>sázíš</td>
<td>‘you (SG) plant’</td>
<td>sázíte</td>
</tr>
<tr>
<td>sází</td>
<td>‘he/she/it plants’</td>
<td>sázejí</td>
</tr>
<tr>
<td>rozumím</td>
<td>‘I understand’</td>
<td>rozumíme</td>
</tr>
<tr>
<td>rozumíš</td>
<td>‘you (SG) understand’</td>
<td>rozumíte</td>
</tr>
<tr>
<td>rozumí</td>
<td>‘he/she/it understands’</td>
<td>rozumejí</td>
</tr>
</tbody>
</table>

Look first at the forms for ‘I plant’ and ‘I understand’. Both of these forms have a first person singular subject and both are present tense. They differ only in the lexical meaning (*plant* versus *understand*). What they have in common is the form *-ím*. So, we surmise that this morpheme means first person singular. (Note that the actual method for determining morphemes this way requires us to look at dozens and dozens of verbs in the language. We are just looking at two verbs to illustrate the principle of morphemic analysis.) We can also assume that the verb roots are that part of the word less the first person singular agreement marker. Keeping these forms in mind, we can also figure out other subject morphemes, such as the second person singular subject morpheme. The word for *you (SG) plant* is *sázíš*. Since *sáz* is the root, the second person singular subject morpheme must be *-ís*. So, we posit the following morphemes in the Czech lexicon.

(15) *-ím 1.SG; -íš 2.SG; sáz ‘plant’; rozum ‘understand’

However, if we look at two forms for ‘I plant’ and ‘you (SG) plant’ we get a different picture. Here, the lexical meaning is the same (‘plant’) but the subject differs. Now, the part of the two words that is the same is *sází,*
which leads us to posit that the first person singular subject marker is just -m. Following the same process as above, we expect the following morphemes.

(16) -m 1.SG; -š 2.SG; sází ‘plant’; rozumí ‘understand’

How do we solve this quandary? This very problem underscores the importance of using a large data set for doing morphemic analysis on unfamiliar languages. Nevertheless, there is enough data here to solve this problem. If we compare the third person plural forms to the other forms, we see that the third person plural morpheme must be -ejí, meaning that the roots must be sáz and rozum. Here, then, are all the morphemes in the Czech data above.

(17) Czech Morphemes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-ím</td>
<td>sáz</td>
<td>‘plant’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ís</td>
<td>rozum</td>
<td>‘understand’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ím</td>
<td>-ejí</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Practice 2.1: How many morphemes are there in the following Korean and English words? Identify them:

사과, 먹는다, 많이, 갑어요, 일반적으로, 받으셨습니다, 갑시다, 왔습니다: houses, cat, alienates, dripped, undeniable, devilishly, withdrawal, decriminalization

2.1.2 Describing Morphemes

Morphemes can be described as either bound or free. As their names imply, a free morpheme can appear on its own, while a bound morpheme must appear with at least one other morpheme. In English most roots are free and most grammatical morphemes are bound. Let’s look at roots first. By way of illustration, here are some free roots in a number of languages.

(18) Free Roots in Various Languages

a. English: run, dog, eat, green
b. Mandarin: jiā家 (‘home’), rì 日 (‘sun’), kàn看 (‘see/look’)
c. Korean: kay 개 (‘dog’), kwi 귀 (‘ear’), chayk책 (‘book’)
d. Thai: dek (‘child’), phaea (‘goat’), ngen (‘silver’)

In some languages the majority of roots are bound. This is common in many languages of North America and Australia. Consider the following Onondaga example of a sentence with two bound roots.

(19) waʔkhninú: neʔ ganakdaʔ
In this example, the roots *hninu* and *nakt* are bound. They can never be uttered on their own, rather, they must appear with affixes. This is the case for all verbs and nearly all nouns in Onondaga.

In English, bound roots are found in a number of words borrowed from Latin. Consider the following examples.

(20) receive, deceive, conceive, perceive, transpire, perspire, conspire

The form *ceive* has a predictable alternant, *cept*, in words like *reception*, *deception*, *conception*, and *perception*. New words are not freely formed with these bound roots. It is an open question whether words like *receive* are stored as wholes in the speaker’s mental lexicon or whether words like *receive* are put together actively in the mind of the speaker the same way a word like *dogs* is.

Consider now the following Korean data.

(21) Korean Bound Root *cwu* 주

maykcwu 백주 potocwu 보도주 socwu 소주 *cwu 주 swul 술
‘beer’ ‘wine’ ‘soju’ (‘alcohol’) ‘alcohol’

Notice that there is a recurring form, *cwu* 주, which appears in the names of various alcoholic beverages. As with *ceive* and *spire*, it is a bound form. It cannot appear on its own. There is a separate free lexical item, *swul*, 술, that can appear alone to mean alcohol.
Practice 2.2: Consider the following Portuguese data. What morphemes are there in the following words? What is the approximate meaning of these morphemes?

- lente ‘slow’
- lentemente ‘slowly’
- triste ‘sad’
- tristemente ‘sadly’
- tristeza ‘sadness’
- limpeza ‘cleanliness’

Now consider the following additional data. Do you need to revise the morphemes you identified above? Are the morphemes bound or free? [Note there are masculine and feminine forms of some adjectives. Do not worry about this distinction, yet. Just note what morphemes you think are involved.]

- limpo ‘clean’ (for masculine nouns)
- limpamente ‘cleanly’
- limp (‘clean’)
- limpa ‘clean’ (for feminine nouns)
- rápido ‘quick’ (masculine)
- rápidamente ‘quickly’
- rápida ‘quick’ (feminine)
- bela ‘beautiful’ (feminine)
- belo ‘beautiful’ (masculine)
- beleza ‘beauty’

Grammatical morphemes can be either free or bound. Bound grammatical morphemes are either affixes (접사) which attach to a word, or clitics (접어), which attach to a phrase. We will come back to clitics near the end of the chapter. Free grammatical morphemes are usually called particles (소사).

Morphemes are described in terms of productivity. Productivity refers to the degree to which the morpheme can attach to relevant bases. For example, the Korean affix -ㅁ is highly productive because it can attach to virtually any verb (감, 어려움, 가르침, 먹음, etc.). The affix -개 has a relatively low productivity as it can attach to some verbs (집개, 지우개, 닦개). In English, the affix -ness has relatively high productivity. It can attach to most adjectives (cuteness, silliness, blueness, randomness, softness, furriness, etc.). The affix -th, however, has extremely low productivity. It attaches to a very small number of adjectives (warmth, length, width, breadth).

Finally, morphemes are also described in terms of compositionality. Compositionality refers to the degree to which the meaning of the whole can be deduced from its parts. The English affix -ness described above is compositional since the meaning of a word bearing this affix can be determined from the base. For example, cuteness refers to the state of being cute. If an English speaker learns a new adjective, they can easily deduce the meaning of the noun resulting from adding the affix -ness. The affix -ment, however, is much less compositional. The meanings of words with the affix -ment cannot easily be deduced from the base in some case, although it is clearer in other cases. Consider the following examples.

<table>
<thead>
<tr>
<th>base</th>
<th>meaning</th>
<th>base+ness</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>depart</td>
<td>to leave</td>
<td>department</td>
<td>a section of a large office or of a large store</td>
</tr>
<tr>
<td>apart</td>
<td>separated</td>
<td>apartment</td>
<td>a single dwelling in a large building</td>
</tr>
<tr>
<td>govern</td>
<td>to oversee</td>
<td>government</td>
<td>a political body that oversees a country, province or city</td>
</tr>
<tr>
<td>resent</td>
<td>to despise</td>
<td>resentment</td>
<td>feeling of ill will</td>
</tr>
<tr>
<td>install</td>
<td>to build in</td>
<td>installment</td>
<td>a periodic payment</td>
</tr>
<tr>
<td>improve</td>
<td>to make better</td>
<td>improvement</td>
<td>the result of making something better</td>
</tr>
</tbody>
</table>
2.1.3 The Positions of Morphemes

Affixes are typically described in terms of where they appear in relation to the base. You may already be familiar with the terms *prefix* (접두사) and *suffix* (접미사). A prefix appears before its base and a suffix appears after its base. Cross-linguistically suffixes are more common than prefixes. Let’s start by looking at the following Mandarin data. The morpheme *-men* is a kind of plural marker in Mandarin. It is sometimes called a *collectivizer* because it refers to a group of non-specific size and is not necessary to indicate plurality the way the English plural suffix is. It attaches after (or to the right of) the root, so it is a suffix.

(22) Mandarin plural suffix *-men*
   a. wo - I; women - we
   b. ni - you; nimen - you.PL
   c. ta - s/he; tamen - they
   d. xuesheng - student; xueshengmen - students
   e. laoshi - teacher; laoshimen - teachers

It is possible for more than one suffix to attach to a root. Korean can attach several suffixes to a root. The word 가셨습니다 is broken down as 가-시-었-습-니-다 (ka-si-ess-sup-ni-ta). The morphological gloss is shown in the following example.

(23) ka-si-ess-sup-ni-ta
   go-HON-PST-FORMAL-INDIC-DECL
   ‘went’ (subject is honorific)

Again, do not worry about the meanings of the individual morphemes too much at this stage. For now, it is important to recognize that it is possible for more than one suffix to appear in a word.

As mentioned, prefixes are rarer cross-linguistically. One fairly productive English prefix is *-re*: redo, re-edit, re-heat, re-shelve, re-apply, and so forth. As with suffixes, it is possible for there to be more than one prefix attached to a root. Consider the following Onondaga data (Woodbury 2018, 422/574).

(24) Onondaga Prefixes
   a. đe-ŋ-nyó:hga?
      t-ɛ-k-ɛ-ʔny-o:hk-a?
      DUC-FUT-1 SG.AG-SRFL-hand-cover-PUNC
      ‘I will cover my hands./I will put gloves on.’
   wa?-e-hwa?ekw-a? ne? če?nhétshæ?
   FACT-3SG.AG-hit-PUNC DET ball
   ‘She batted the ball.’

**Infixation** involves adding an affix inside a root (Yu, 2007). Consider the following Tagalog data (Schachter and Otanes, 1972, 370). Observe that the past tense morpheme -in- appears inside the root.

(25) Tagalog Infixes

<table>
<thead>
<tr>
<th>base</th>
<th>infixed form</th>
</tr>
</thead>
<tbody>
<tr>
<td>bigyan</td>
<td>‘give’</td>
</tr>
<tr>
<td>ibili</td>
<td>‘buy for’</td>
</tr>
<tr>
<td>basa</td>
<td>‘read’</td>
</tr>
</tbody>
</table>

Observe in the data above that the infix appears immediately to the right of the first consonant.

binigyan

ibinili

binasa

Here is another example from Hua (Haiman, 1980, 195), also known as Yagaria.

(26) Hua Negation

<table>
<thead>
<tr>
<th>base form</th>
<th>negated form</th>
</tr>
</thead>
<tbody>
<tr>
<td>zgavo</td>
<td>‘embrace’</td>
</tr>
<tr>
<td>harupo</td>
<td>‘slip’</td>
</tr>
<tr>
<td>rapai</td>
<td>‘be fat’</td>
</tr>
</tbody>
</table>

In Hua, the infix -?a- is placed before the final syllable.

**Hua or Yagaria** is a Trans-New Guinea language spoken by about 20,000 people in Papua New Guinea. It has dual number (*agaea* = she/he; *tagaea* = they two; *pagaea* = they three or more).

English and Cantonese have a process of lexical infixation (also called *tmesis*). Consider the following examples. The numbers in the Cantonese data refer to tones. See Chapter 3 for details on tone.

(27) abso-bloody-lutely, fan-fucking-tastic.
/-ly/  Adj
<N, friend, father, mother, scholar, . . . > _-
/li/

characteristic of or embodying the main qualities of

Table 2.1: Lexical Entry for /-ly/

(28) a. bin1'dou6 'where'
     b. bin1-gwai2-dou6 (<devil>where) 'where the hell'

2.2 The Grammar of Derivational Morphology

Recall from above that words are built up according to the morphological rules of the language. A current topic of investigation in morphology and in psycholinguistics is how words are stored and accessed in the mental lexicon. Consider the following. English speakers know the words ‘friend’ and ‘friendly’. Are both of these stored individually in the mental lexicon? Simply answering yes to this question does not capture the observation that /-ly/ is a common, recurring morpheme in English with consistent properties, namely that it turns nouns into adjectives. On the other hand, simply stating that ‘friend’ and ‘-ly’ are stored separately and put together as needed leads to over-generation. Why can’t /-ly/ be attached to any noun giving rise to such non-words as *dogly or *cactusly? Adding to the confusion in this matter is the existence of bound roots discussed above. Do speakers form words such as perceive, deceive, and so forth online as necessary? Or are they stored as atoms with the similarities among them merely coincidental to the modern English speaker? There are even more opaque cases. The English word infant historically has four distinct morphemes, which, put together literally mean ‘not speaking’. It is unlikely that we would want to say that infant is formed online.

Nevertheless, there are still many generalizations that can be made about the formation of words. Consider the list of words to which /-ly/ is added. (Note, this is different from the /-ly/ suffix that turns adjectives into adverbs.)

(29) friend -> friendly; mother -> motherly; father -> fatherly; scholar -> scholarly

We describe the properties of morpheme in lexical entries. The first line of the entry states the category of the lexical item. The second line indicates the morphological properties of the lexical item. In the ordered set <X, A>, X indicates the category that the lexical item attaches to and the set A, a, b, c, . . . lists those particular bases the affix attaches to when the lexical item in question is not fully productive. The third line gives details on the pronunciation; however, we will concern ourselves with phonological details in chapter 3. The fourth line contains the meaning of the lexical item. Thus, we can state the lexical entry for /-ly/ more formally as follows.

The first line indicates that /-ly/ is an adjective. The second line indicates that it attaches to nouns and indicates the set of particular nouns that it attaches to. Since the list is quite large, not every single noun is
The underscore (_) indicates the position of the morpheme /-ly/ with respect to its base. Here, it appears as a suffix to the set of nouns given. Another way to think about these to lines is as follows. /-ly/ is a suffix that attaches to any one of the nouns given in the list to form an adjective. The third line is the pronunciation of the morpheme. We will cover this in the phonetics and phonology chapters later, so will not consider it further here. We should note, of course, that the pronunciation of a lexical item is an obligatory component of its lexical entry. The last line is the definition of the morpheme. We can draw a morphological tree for *motherly* as follows.

Let’s look at another example. Consider the following Korean data.

2.2.1 English Derivational Morphology

Although English has very little inflectional morphology, it has a great deal of derivational morphology. The Table 2.2 lists several derivational affixes of English (Fabb, 1988). This is only a partial list of the derivational morphemes in English. Note that the meanings given are only approximate. Please consult a refutable dictionary for precise meanings of these affixes.

Derivational morphemes typically select bases of a certain category (as shown in Table 2.2). The reversative prefix *de-* selects a verb and gives rise to a verb with a new meaning. Derivational morphemes sometimes select specific derivational morphemes. For instance, *-ity* selects *-able*, and *-able* selects *-ify*. The following examples show these selectional restrictions.

We capture these structures in tree diagrams. The tree is built up from the root working out until the full structure...
<table>
<thead>
<tr>
<th>affix</th>
<th>base</th>
<th>resultant category</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>be-</td>
<td>adj or verb</td>
<td>verb</td>
<td>causation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>befriend, bestrange, bewet</td>
</tr>
<tr>
<td>de-</td>
<td>verb</td>
<td>verb</td>
<td>reversative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>de-ice, dehumidify</td>
</tr>
<tr>
<td>mis-</td>
<td>verb</td>
<td>verb</td>
<td>wrongly or in error</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>misrepresent, misbelieve, misclassify</td>
</tr>
<tr>
<td>un-¹</td>
<td>adj</td>
<td>adj</td>
<td>negative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>unattractive, unequal, unconstitutional</td>
</tr>
<tr>
<td>un-²</td>
<td>verb</td>
<td>verb</td>
<td>reversative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>undo, unwind, undress</td>
</tr>
<tr>
<td>in-</td>
<td>adj</td>
<td>adj</td>
<td>negative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ineligible, incredible, illegal, impossible (note sound changes)</td>
</tr>
<tr>
<td>-able</td>
<td>verb</td>
<td>adj</td>
<td>ability of action</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>readable, doable, excitable</td>
</tr>
<tr>
<td>-ar</td>
<td>noun</td>
<td>adjective</td>
<td>of or pertaining to</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linear, columnar</td>
</tr>
<tr>
<td>-ate</td>
<td>noun or root</td>
<td>verb</td>
<td>act or carry out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>formulate, abbreviate, rejuvenate</td>
</tr>
<tr>
<td>-ation</td>
<td>noun or verb</td>
<td>noun</td>
<td>the act of making or doing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>flirtation, affixation, ionization</td>
</tr>
<tr>
<td>-ify</td>
<td>noun or adj</td>
<td>verb</td>
<td>to make</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>acidify, beautify, falsify</td>
</tr>
<tr>
<td>-ion</td>
<td>verb or root</td>
<td>noun</td>
<td>the action or state</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>consumption, prosecution, regression</td>
</tr>
<tr>
<td>-less</td>
<td>noun</td>
<td>adj</td>
<td>lacking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>penniless, cordless, helpless</td>
</tr>
<tr>
<td>-ness</td>
<td>adj</td>
<td>noun</td>
<td>state or quality of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>happiness, darkness, sleekness</td>
</tr>
</tbody>
</table>

Table 2.2: Derivational Affixes of English
is built. Affixes must be added according to their subcategorization frames. Here are the trees for all the words in example (32).

(33) a. 

```
  V     Adj     N
 /     \     /
read   -able  read
    \     /
     V     Suf
     -ity
     \     /
      read  -able
```

b. 

```
  N     V     Adj     N
 /     \     \     /
class  -ify  V     Suf
      \     \     /
       class -able
```

The affix -ation requires some comment. It is clearly composed of -ate and -ion; however, in Modern English it has become a fixed form. Thus, formulation is composed of formula-ate-ion while taxation is composed of tax-ation. These are shown in the following morphological trees.

(34)

```
  N
 /     /
V     Suf
/     /
N     Suf
     -ion
     |
     tax
     |
     -ation
     |
     |
     |
     |
     formula
     |
     -ate
```

Recall that -ate forms verbs. Since to tax is already a verb it cannot appear with -ate independently of -ion. That is, there is no such word *taxate. There is, however, a word formulate. The following example shows how tax is used as a verb.

(35) a. The government taxed the people.

b. The government’s taxation of the people.

Determining the correct morphological structure for a complex word can be tricky if there are both prefixes and suffixes. Consider the word declassify. There is both a prefix and a suffix. How do we decide which one attaches first? The prefix de- only attaches to verbs, and the root class is a noun, not a verb. First the suffix -ify attaches to noun to form the verb classify. Then the prefix can be added. Here is the structure.
Sometimes we have to consider the meaning of the word. Consider the word *undeniable*. The prefix *-un* can attach to both verbs and adjectives. When it attaches to verbs it has the meaning of reversing an action. Thus *untie* means to undo the action of tying something. So, it can only attach to verbs that refer to a reversible action. There is no verb *undeny* because one cannot undo the action of denying something. Thus, the suffix must attach first. Here is the tree for *undeniable*.

Finally, consider the word *unlockable*. Again, the prefix *-un* can attach either to verbs or to adjectives (with different meanings). In this case, *

The first tree takes the adjective *lockable* as a base and attaches the prefix *-un* giving rise to the meaning ‘not lockable’.

**Practice 2.3:** Draw morphological trees for the following words: energize, unreliability, destabilization, mistrustful.
Table 2.3: Derivational Affixes of Korean

<table>
<thead>
<tr>
<th>affix</th>
<th>base</th>
<th>resultant category</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>고-</td>
<td>noun</td>
<td>noun</td>
<td>high</td>
</tr>
<tr>
<td>뇌-</td>
<td>verb</td>
<td>verb</td>
<td>repetitive or reversative</td>
</tr>
<tr>
<td>-답</td>
<td>noun</td>
<td>verb</td>
<td>characteristic of</td>
</tr>
<tr>
<td>-로</td>
<td>stative verb</td>
<td>adv</td>
<td>주관적으로, 충동적으로</td>
</tr>
<tr>
<td>-적</td>
<td>noun</td>
<td>verb</td>
<td>characteristic of</td>
</tr>
<tr>
<td>-지</td>
<td>noun</td>
<td>verb</td>
<td>characteristic of</td>
</tr>
<tr>
<td>-이۱</td>
<td>verb</td>
<td>noun</td>
<td>instance of</td>
</tr>
<tr>
<td>-이۲</td>
<td>stative verb</td>
<td>adv</td>
<td>adverbial function</td>
</tr>
</tbody>
</table>

2.2.2 Korean Derivational Morphology

As we will see in the next section, Korean has a great deal of inflectional morphology. Here, we discuss some derivational morphology in Korean. Table 2.3 contains a partial list of Korean derivational morphemes (Yeon and Brown, 2011). Again, the meanings are only approximate. Note that the postposition -로 productively selects the suffix -적.

Generally very few derivational morphemes appear in a single word in Korean (in contrast to English, as we saw above). The trees for Korean are relatively straightforward to draw. Consider the words 높이 and 정열적으로. The structures are shown in the following trees.

(39) N          PP
    V Suf      N P
    늘이 Suf   -으로
                정열적

2.3 The Functions of Inflectional Morphology

We have already seen above that morphemes can have specific functions. For instance, in English the morpheme /-ly/ changes a noun into an adjective. In this section we will look at the functions of inflectional morphology. Inflectional morphemes indicate grammatical properties, a number of which we will look at here. Sometimes, the dividing line between inflectional morphology and derivational morphology is grey, leading some people...
to suspect that the division between these two is not real. Nevertheless, many people still distinguish between
them and the distinction is sometimes useful.

2.3.1 Noun Class

The first type of inflection we will look at is noun class (명사부류). Many languages around the world divide
their nouns up into various classes. One well-known type of classification is based on gender. Consider the
following French data.

(40) French Gender
a. la pomme vert-e
   the.FEM apple green-FEM
   ‘the green apple’
b. la petit-e porte
   the.FEM small.FEM door
   ‘the small door’
c. le crayon vert-Ø
   the.MASC pencil green-MASC
   ‘the green crayon’
d. le petit-Ø four
   the.MASC small-MASC oven
   ‘the small oven’

The French nouns crayon (‘pencil’) and four (‘oven’) are masculine, while the nouns pomme (‘apple’) and
porte (‘door’) are feminine. For the most part, gender is an idiosyncratic property of the noun, and must simply
be encoded in the lexical entry of the noun. In French, words such as homme (‘man’) and femme (‘woman’) are
predictably masculine and feminine, respectively. Many words relating to people can change, depending on the
referent, such as avocat (‘(male) lawyer’) and avocate (‘female lawyer’), although the masculine form is the
default form. However, for the vast majority of words the gender must simply be memorized.

Some languages divide up the class of nouns based on animacy. As with gender above, it is largely unpredict-
dable, although words relating to people and animals are always animate. Consider the following Blackfoot
data (Frantz, 1997). Observe that there are different markers for animate singular and plural nouns versus inan-
imate singular and plural nouns.

(41) Blackfoot Animacy
a. ponoká-wa
   elk-ANIM.SG
   ‘elk’
b. í’ksisako-yi
   meat-INANIM.SG
   ‘meat’
Noun classes can also be identified by a free morpheme rather than by a bound morpheme as seen in French and Blackfoot above. These free morphemes are called classifiers (분류사). Classifiers are particularly common in East and Southeast Asia, but are found in other parts of the world, too. Here are some Korean classifiers.

(42) Korean Classifiers
   a. 개 kay - generic classifier, used when a more specific one is not available
   b. 줄 cwul - used for items in a line, kimbap, string, rope, etc.
   c. 권 kwen - used for books
   d. 마리 mali - used for animals (and disrespectfully for people)

Here are some examples. Note that Korean has two number systems: Native Korean numbers and Sino-Korean numbers (borrowed from Chinese). The majority of classifiers use Native Korean numbers for quantities up to 99; however, in practice most people use Native Korean number only as far as 20, except for to express the age of a person. For higher quantities, Sino-Korean numbers are used.

(43) a. 사과 두 개 sakwa twu kay (apple two CL) ‘two apples’
    b. 낙지 네 마리 nakci ney mali (small.octopus four CL) ‘four small octopuses’

Chinese languages have dozens of classifiers. Table 2.4 provides a small sample. Note that the character enclosed in brackets is the simplified form. Here are some examples.

(44) Mandarin classifiers
   a. 三 个 rén three CL person ‘three people’
   b. 五 张 zhī paper ‘five pieces of paper’

Bantu languages are also well known for having a complex system of noun classification [Meinhof 1906]. Noun class markers in Bantu languages are prefixes and have separate forms for singular and plural. Here are some examples. The following charts are just partial lists of the full set of noun class prefixes in Swahili and Shona.
Table 2.4: Chinese Classifiers

<table>
<thead>
<tr>
<th>Character</th>
<th>Mandarin</th>
<th>Cantonese</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>個(个)</td>
<td>ge</td>
<td>go³</td>
<td>generic classifier, also used for people</td>
</tr>
<tr>
<td>名</td>
<td>míng</td>
<td>ming⁴</td>
<td>honorific classifier for people</td>
</tr>
<tr>
<td>張(张)</td>
<td>zhāng</td>
<td>jeung¹</td>
<td>for flat objects such as paper and blankets</td>
</tr>
<tr>
<td>隻(只)</td>
<td>zhī</td>
<td>jek³</td>
<td>for animals and for a single entity of paired objects (a shoe, a chopstick, etc.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pet⁶</td>
<td>for shapeless clumps such as mud or feces (Cantonese only)</td>
</tr>
<tr>
<td>堅</td>
<td></td>
<td>bung⁶</td>
<td>walls (Cantonese only)</td>
</tr>
<tr>
<td>頂(顶)</td>
<td>dǐng</td>
<td>deng²</td>
<td>objects with protruding tops (hats)</td>
</tr>
<tr>
<td>把</td>
<td>bā</td>
<td>ba²</td>
<td>things with a handle or long, slender objects (scissors, keys, knives, etc.)</td>
</tr>
</tbody>
</table>

Table 2.4: Chinese Classifiers

(45) Swahili noun classes

<table>
<thead>
<tr>
<th>singular</th>
<th>plural</th>
<th>use</th>
</tr>
</thead>
<tbody>
<tr>
<td>m-</td>
<td>wa-</td>
<td>people</td>
</tr>
<tr>
<td>ki-</td>
<td>vi-</td>
<td>utensils and many inanimate objects</td>
</tr>
<tr>
<td>m-</td>
<td>mi-</td>
<td>trees, plants, natural forces</td>
</tr>
</tbody>
</table>

(46) Shona noun classes

<table>
<thead>
<tr>
<th>singular</th>
<th>plural</th>
<th>use</th>
</tr>
</thead>
<tbody>
<tr>
<td>mù-</td>
<td>và-</td>
<td>people</td>
</tr>
<tr>
<td>chi-</td>
<td>zvì-</td>
<td>utensils and many inanimate objects</td>
</tr>
<tr>
<td>mù-</td>
<td>mi-</td>
<td>trees, plants, natural forces</td>
</tr>
</tbody>
</table>

(47) Bantu noun class prefixes

a. Swahili ([Mohamad, 2001])
   (i) m-tu ‘person’; wa-tu ‘people’
   (ii) wa-limu ‘teachers’
   (iii) m-nazi ‘a coconut tree’; mi-nazi ‘coconut trees’
   (iv) ki-lima ‘hill’; vi-lima ‘hills’

b. Shona ([Fortune, 1980])
   (i) mù-kómáná ‘boy’; và-kómáná ‘boys’
   (ii) chi-nhù ‘thing’; zvì-nhù ‘things’

Finally, Australian languages have a unique system of noun classes. Dyirbal, for instance, divides nouns along the following classes ([Dixon et al., 1972]).

(48) Dyirbal Noun Classes
Dyirbal is a highly endangered Pama-Nyungan language spoken in northeastern Australia.

Recalling the Sapir-Whorf Hypothesis discussed in Chapter 1, it is tempting to jump to the wrong conclusion regarding the nature of the noun classes in Dyirbal. This is especially apparent in Lakoff’s book *Women, Fire, and Dangerous Things* (Lakoff, 2008). Plaster and Polinsky (2007) make clear, however, that comparative historical data on other related languages and acquisition data suggest that children do not acquire noun classes by grouping women, fire and dangerous things into a semantically uniform set. Rather, children rely on statistical and phonetic cues in acquiring noun classes. Adults often come up with after-the-fact explanations for the logic of noun classes, but these explanations often have nothing to do with how children acquire language and more to do with societal perceptions of gender.

### 2.3.2 Number

Many languages have some way of encoding the quantity of items. Thus, the English word book refers to one item while the word books refers to two or more items. This distinction is called number. Specifically, English distinguishes singular (one item) and plural (more than one item). Number distinctions are found both on nouns and on verbs. Consider the following Upper Sorbian data.

(49) Number in Upper Sorbian

<table>
<thead>
<tr>
<th>Singular</th>
<th>Dual</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>hród ‘castle’</td>
<td>hrodaj ‘(two) castles’</td>
<td>hrody ‘(three or more) castles’</td>
</tr>
<tr>
<td>dzelam ‘(I) work’</td>
<td>dzelamoy ‘(we two) work’</td>
<td>dzelamy ‘(we three or more) work’</td>
</tr>
</tbody>
</table>

Upper Sorbian distinguishes between singular, dual (고수/고양수), and plural. Dual indicates two items. In languages that have mark dual number, plural indicates three or more items.

Although rare, some languages include trial number in their pronominal system, but not on nouns. The Austronesian language Tolomako distinguishes singular, dual, trial, and plural on its pronouns. Table 2.5 lists the pronouns of Tolomako (adapted from Guy, 1972). Note that the pronouns exhibit an inclusive/exclusive distinction, discussed below.
2.3.3 Person

Person distinctions in natural language include the first person (인칭/the speaker, ‘I’ - 나), the second person (인칭/the addressee, ‘you’ - 너), and third person (인칭/anyone else, ‘he/she/it’ - 걐 (colloquial)). Person marking is found on pronouns in numerous languages around the world, including English as shown in the example sentences above. Consider the Chinese data in Table 2.6.

Additionally, some languages exhibit a *clusivity* distinction in the first person. Consider the following examples.

(50) a. We’re going out for dinner tonight...What do you want to eat?
    b. We’re going out for dinner tonight...Can you stay home and watch the dog?

Both sentences use a first person plural pronoun, but the interpretation is different. We in (50-a) includes the addressee, but we in (50-a) does not. The Korean pronoun *우리* has the same ambiguity (but not *저희*, which excludes the hearer, so is exclusive). Consider the following Tongan data.

(51) a. Te mau fakataha he Falaite kahaú.
    ‘We will meet next Friday.’

---

1The forms 그 and 그녀 are often given as third person pronouns in Korean; however, these are seldom used in actual speech. 걐 (< 걐어), however, is often used in casual speech.
b. Oku tau lau ia ko ha taki.
    PRS 1NCL.PL regard 3SG PRED ha leader
    ‘We regard him (as) a leader.’

2.3.4 Tense

An event can be marked as taking place before the moment of speech (past tense), during the moment of speech (present tense), or after the moment of speech (future tense). This information is encoded as a morpheme in some cases. In English, verbal morphology encodes only a past/non-past distinction. Future tense is encoded by a separate lexical item. Consider the following examples.

(52) a. I played the piano.
    b. I play the piano.
    c. I will play the piano.

The form play is used both in the present tense and the future tense. To express futurity, a separate modal will is used. Past tense is encoded by the morpheme spelled orthographically as ‘-ed’ on the verb. Languages differ as to whether tense is encoded morphologically or in some other way.

Consider the following Swahili data (Ojiambo, 2010). Observe that past, present and future are encoded as a prefix on the verb. Note further that person is also encoded as a prefix on the verb.

(53) a. ni-na-soma
    1.SG-PRS-study
    ‘I am studying.’

b. ni-li-soma
    1.SG-PST-study
    ‘I studied.’

c. ni-ta-soma
    1.SG-FUT-study
    ‘I will study.’

Korean has a variety of ways to express tense, including more than one way to express future. We will consider just a few here. Consider the following data. They show distinct suffixes for present, past, and future. Note there is a declarative suffix indicating that this form of the verb is used for statements. Don’t worry about this suffix now. We will discuss it in Chapter 5.
2.3.5 Case

Case (격) indicates the grammatical relationship of a noun phrase with the sentence. The grammatical subject in many languages appears with nominative case (주격) and the direct object appears with accusative case (대격). Here are some examples.

(55) Nominative and Accusative Case
   a. I saw him. (I - nominative; him - accusative)
   b. 영희가 책을 읽었다. (Yenghuy-ka - nominative; chayk-ul - accusative)
   c. mater alit pull-os
      mother.NOM nourishes children-ACC
      ‘Mother nourishes the young ones.’ [Latin]

The Korean example clearly shows the nominative case suffix (-ka/가) and the accusative case suffix (-ul/을). Not all languages have an overt suffix for both nominative and accusative. In the Latin example we see an accusative suffix (-os), but there is no nominative suffix. Finally, Modern English has no overt case suffixes; however, pronouns inflect for case. For example, the first person singular nominative pronoun is I, and the first person singular accusative pronoun is me.

Dative case (여격) is used to indicate the indirect object. Consider the following Lithuanian example (Mathiassen, 1996).

(56) Mergait-ė davė vaik-ui obuol-i
    girl-NOM gave boy-DAT apple-ACC
    ‘The girl gave the boy an apple.’ [Lithuanian]

Case is often marked by bound morphemes as shown in the examples above. Case can also be marked by free morphemes. This is common in Austronesian languages. Here is an example from Tahitian (Polinsky and Potsdam, 1997). Note that there is no marker for nominative case. Accusative case (ACC) is marked by the particle ‘i. (‘ represents a glottal stop.)
(57) a. te ma’ue nei te mau manu.  
ASP fly ASP DET PL bird  
‘The birds are flying.’  

b. ‘ua ‘ite te tamaiti ‘i te mau manu  
PERF see DET child ACC DET PL bird  
‘The child saw (the) birds.’

2.4 Analyzing Morphology

We can figure out the morphemes of a language by examining closely related words. We will look at Swahili again. Although some of the morphemes have been shown above, consider these data as though you are seeing them for the first time.

(58) a. ninasoma ‘I am studying’  
b. tulisoma ‘We studied’  
c. tutasoma ‘We will study’  
d. unasoma ‘You(sg.) are studying’  
e. mlisoma ‘You(pl.) studied’  
f. atasoma ‘S/he will study’  
g. walisoma ‘They studied.’  
h. nilisoma ‘I studied’

We start by comparing two words that differ by only one feature. The forms for *I am studying* and *You(sg.) are studying* differ only by the person (1st person singular versus 2nd person singular). Let’s compare these forms.

ninasoma  
unasoma

The parts of the words that differ are shown in red. We surmise that *ni-* means 1SG and that *u-* means 2SG. Now, let’s look at the forms for *I am studying* and *I studied*. These two forms differ only by tense (present versus past). The rest of the word is the same.

ninasoma  
nilisoma

Here, the only parts of the words that differ are *na-* and *li-.* Again, we surmise that *na-* means present tense and that *li-* means past tense. Furthermore, looking at the data, we see that *soma* is the only consistent part, so we can surmise that it means *study.* You should now be able to segment the remaining data on your own.
Practice 2.4: Segment the following Finnish data into as many morphemes as you can find.

<table>
<thead>
<tr>
<th>Finnish Word</th>
<th>English Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>puhun</td>
<td>‘I speak’</td>
</tr>
<tr>
<td>puhuin</td>
<td>‘I spoke’</td>
</tr>
<tr>
<td>puhuisin</td>
<td>‘I would speak’</td>
</tr>
<tr>
<td>puhut</td>
<td>‘you (sg.) speak’</td>
</tr>
<tr>
<td>puhuit</td>
<td>‘you (sg.) spoke’</td>
</tr>
<tr>
<td>puhuisit</td>
<td>‘you (sg.) would speak’</td>
</tr>
<tr>
<td>puhumme</td>
<td>‘we speak’</td>
</tr>
<tr>
<td>puhuimme</td>
<td>‘we spoke’</td>
</tr>
<tr>
<td>puhuisimme</td>
<td>‘we would speak’</td>
</tr>
<tr>
<td>puhutte</td>
<td>‘you (pl.) speak’</td>
</tr>
<tr>
<td>puhuitte</td>
<td>‘you (pl.) spoke’</td>
</tr>
<tr>
<td>puhuisitte</td>
<td>‘you (pl.) would speak’</td>
</tr>
</tbody>
</table>

Morphemes often appear in two or more alternants whose distribution is predictable. What we mean by this is that there can be more than one form of a morpheme, and there is an observable pattern in which these forms appear. When there is more than one such form, each form is called an **allomorph**. Let’s look at an example.

Consider the following Korean data.

(59) Korean 부 and 불

<table>
<thead>
<tr>
<th>Korean Word</th>
<th>English Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>자연</td>
<td>‘naturalness’</td>
</tr>
<tr>
<td>부자연</td>
<td>‘unnaturalness’</td>
</tr>
<tr>
<td>도덕</td>
<td>‘morality’</td>
</tr>
<tr>
<td>부도덕</td>
<td>‘immorality’</td>
</tr>
<tr>
<td>자유</td>
<td>‘freedom’</td>
</tr>
<tr>
<td>부자유</td>
<td>‘lack of freedom’</td>
</tr>
<tr>
<td>공정</td>
<td>‘fairness’</td>
</tr>
<tr>
<td>불공정</td>
<td>‘unfairness’</td>
</tr>
<tr>
<td>완전</td>
<td>‘completeness’</td>
</tr>
<tr>
<td>불완전</td>
<td>‘incompleteness’</td>
</tr>
<tr>
<td>합격</td>
<td>‘success’</td>
</tr>
<tr>
<td>불합격</td>
<td>‘failure’</td>
</tr>
<tr>
<td>안정</td>
<td>‘stability’</td>
</tr>
<tr>
<td>불안정</td>
<td>‘instability’</td>
</tr>
</tbody>
</table>

The two allomorphs are 부 (pwu) and 불 (pwul). What is their distribution? Looking at the data we see that pwu appears before ‘c’ and ‘t’ while pwul appears in all other environments, or elsewhere.

We notate this as follows.

```
<table>
<thead>
<tr>
<th>pwul / ___ {t,c}</th>
</tr>
</thead>
<tbody>
<tr>
<td>pwu / __</td>
</tr>
</tbody>
</table>
```

Figure 2.1: Distribution of allomorphs

This notation captures the following facts. The morpheme pwul has two allomorphs: pwu and pwul. The allomorph pwu appears before ‘t’ or ‘c’. The allomorph pwul appears elsewhere. Note for now the formulation above just refers to the orthographic forms of the romanization of the Korean words. Once we cover phonetics and phonology in more detail, we will revisit this notation.

We will come back to morphological analysis in more detail in chapter 4.
2.5 Typological Properties of Morphology

Here we discuss some well known typological properties of morphology (Humboldt, 1825). We discuss typology more in chapter 5. Briefly, linguistic typology is the study of how properties cluster in language. For instance, OV languages such as Korean tend to have postpositions (서울에) while VO languages such as English tend to have prepositions (to Seoul). We discuss two ways of describing the morphology of languages. It is important to remember that these two properties are spectra rather than absolutes. For instance, people often say that Mandarin is an analytic or isolating language, while Mohawk is a polysynthetic language. Polysynthesis and analyticity are two ends of a spectrum. Languages can fall in the middle of this spectrum.

2.5.1 Polysynthesis and Analyticity

This spectrum refers to the level of morphological complexity of words. A polysynthetic language has morphologically complex words, while analytic languages, also known as isolating languages (고립어), have very little morphological complexity.

Polysynthetic languages contain words with a great deal of morphological complexity. Many of the languages of the Americas and of Australia are polysynthetic, although polysynthetic languages are found all over the world. Consider the following Inuktitut example.

\[(60) \text{annulaksi-kkami-nginna-jualu-gasu-lauqsima-guma-nngit-tsaq-galuaq-tunga} \]
\[\quad \text{imprison-again-really-a.lot-try-ever-want-NEG-EMPH-EMPH-DECL.1.S.S} \]
\[\quad \text{‘I would never ever even want to try to end up in jail ever again even for a bit.’ (Johns, 2007)} \]

Analytic (or isolating) languages have much less morphological complexity in their words. Consider the following Mandarin example.

\[(61) \text{wǒ-men bù huì chī nà ge píngguǒ} \]
\[\quad \text{1-PL NEG will eat DEM CL apple} \]
\[\quad \text{‘We will not eat that apple.’} \]

The terms analytic and synthetic are also used to describe how a concept is encoded. For example, in English future tense is encoded analytically. That is, the morpheme that encodes future tense is a separate word. Past tense, however, is encoded synthetically. That is, the morpheme that encodes past tense is an affix that is attached to the verb. Consider the following examples.

\[(62) \text{a. John will play the piano. [‘will’ is a separate word]} \]
\[\quad \text{b. John played the piano. [‘-ed’ is a suffix attached to the verb]} \]

In Korean, future tense can be either synthetic or analytic.
Finally, in English the comparative can be either synthetic or analytic.

(64) a. taller, tastier [synthetic comparative]
    b. more tasty, more intelligent [analytic comparative]

2.5.2 Fusion and Agglutination

The second spectrum refers to how much information is encoded in a single morpheme. Agglutinative languages tend to encode one grammatical concept per morpheme. Fusional languages tend to encode many grammatical concepts per morpheme.

Turkish is a typical agglutinative language. Consider the following Turkish data.

(65) Turkish Agglutination
   a. yap-ı-la¸ s-t–ma-mı¸ s
      build-NLZR-VLZR-CAUS-PASS-NEG-EV.PRF
      ‘(It) has not been built up.’
   b. Gör-ü¸ s-tür-ül-e-me-ye
      see-RECIP-CAUS-PASS-PSB-NEG-PSB also PSB-IMPF INTER-PST.COP-2PL
      de bil-iyor mu-ydu-nuz?
      also bil-iyor PSB IMPF mu-ydu-nuz?
      ‘Did it also happen that you were not allowed to see each other?’

Observe that there are many morphemes on each word, and that each morpheme tends to correspond to a single feature. Note that there are some portmanteau morphemes in these data. A portmanteau morpheme expresses more than one concept at once. For example, the morpheme /ydu/ expresses both past tense (PST) and the copula (COP = 계사 / BE동사). It is doubtful that any language exists that is purely agglutinative. Such a language would spell out every morpho-syntactic feature on a separate affix. Like polysynthesis and analyticity, agglutination and fusionality are ends of a spectrum. Other agglutinative languages include Korean, Indonesian, Finnish, Japanese, and Telugu.

Fusional languages express multiple concepts on a single morpheme. Again, no language can be purely fusional. Spanish encodes tense and agreement fusionally. Consider the following Spanish examples.

(66) a. com-o
    eat-1SG.PRS
    ‘I eat.’
   b. com-e
    eat-2SG.PRS
    ‘You eat.’
c. com-í
   eat-1SG.PST
   ‘I ate.’

d. com-iste
   eat-2SG.PST
   ‘You ate.’

In the Latin word *bon-us* ‘good’, the suffix *-s* indicates masculine gender, singular number and nominative case. Other fusional languages include Hebrew, Punjabi and Russian.

### 2.6 Compounding

A wide-spread word-formation strategy in many languages around the world is **compounding**. A compound (합성어) is defined as a word containing two or more lexical roots. English examples include *textbook, snowman, white board, and smart phone*. Consider the following Cantonese data.

(67) Cantonese Compounds

<table>
<thead>
<tr>
<th>yu.yin</th>
<th>yu.yin-hok</th>
<th>yu.yin-hok-ga</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>language-study</td>
<td>language-study-person</td>
</tr>
<tr>
<td>‘language’</td>
<td>‘linguistics’</td>
<td>‘linguist’</td>
</tr>
<tr>
<td>fa</td>
<td>fa-hok</td>
<td>fa-hok-ga</td>
</tr>
<tr>
<td>change</td>
<td>change-study</td>
<td>change-study-person</td>
</tr>
<tr>
<td>‘change’</td>
<td>‘chemistry’</td>
<td>‘chemist’</td>
</tr>
</tbody>
</table>

Compounds can be classified in terms of their headedness as follows. An **endocentric compound** contains a single head. An **exocentric compound** contains no head inside the compound. A **dvandva** contains two heads. Endocentric compounds are perhaps the most familiar, so we’ll start with those. English *greenhouse* and Korean *치약* are both endocentric compounds. In particular, they are both right-headed compounds since the right portion is the head. The head determines the category of the compound and gives a clue (sometimes vague) as to the meaning. For *greenhouse*, the two members are *green* (an adjective) and *house* (a noun). The compound *greenhouse* is a noun. It is roughly speaking a kind of house—a house for plants to live in rather than for people to live in. *치약* is a kind of medicine (roughly), but certainly not a kind of tooth. Both of these compounds are represented as follows.
Compounds can be recursive. That is a compound can consist of smaller compounds. Consider the compound *university administration office*. This refers to an office for the university administration. Thus *university administration* must form a constituent first. Here is the structure.

```
(69) N
    /   N
   /    office
university administration
```

Consider now the compound *head green house*. This compound refers to a green house that is larger or more important than the others. Here is the structure.

```
(70) N
    /   N
   /    N
  head Adj N
   /    green house
```

Finally, consider the complex compound *head green house inspector*. This compound is ambiguous. It could either refer to the inspector of the head green house, or it could refer to the green house inspector who is in charge of all the other green house inspectors. Here are the two trees, respectively.

```
(71) N
    /   N
   /    inspector
head Adj N
   /    green house
```

```
N
/   N
/    N
/     inspector
/      Adj N
/       green house
```

```
N
/   N
/    N
/     inspector
/      Adj N
/       green house
```
Practice 2.5: Draw trees for the following English compounds: dog food, blackbird, hotdog, cottage cheese, scrap paper, dog food bowl, toilet bowl cleaner

In an exocentric compound, neither member is the head of the compound. Consider the Korean compound 돌대가리. 돌대가리 is neither a kind of stone nor a kind of head. It is a kind of person. Neither 돌 nor 대가리 is the head of the compound. English exocentric compounds include flatfoot and old money. A flatfoot is not a type of foot or a type of flatness. It is a detective. Old money is not a kind of money or oldness. It refers to a wealthy person who has inherited money from their family.

Finally, a dvandva is a compound in which both members are the head of the compound. For example, a singer-songwriter is someone who is both a singer and a songwriter. In Korean, 천지 refers to both the sky and the earth.

When the head of an endocentric compound has an irregular plural, the whole compound takes the same irregular plural. When the head of an exocentric compound has an irregular plural, however, the whole compound typically takes a regular -s for the plural. There is some speaker variation, however. Compare the following examples.

(72) English endocentric compounds
   a. wisdom tooth - wisdom teeth
   b. field mouse - field mice
   c. Canada goose - Canada geese

(73) English exocentric compounds
   a. flatfoot - flatfoots
   b. Maple leaf - Maple leafs (name of a hockey team)
   c. bigfoot - bigfoots

The compound saber tooth is interesting in this regard. When it refers to a type of tooth, it is an endocentric compound, and the plural is saber teeth. When it refers to a type of extinct tiger it is an exocentric compound, and the plural is saber tooths.

2.7 Non-concatenative Morphology

In many languages the morphological concepts discussed in this chapter are not added by prefixes and suffixes as discussed here. The word concatenate means to place in a line by a series of links—like a chain. Looking back at the examples above, the order of the morphemes can be described by a linear chain. Recall the Swahili data in (58) on page 48.
The following grammatical and derivational changes cannot be expressed as a simple lining-up of morphemes. Thus, we call it non-concatenative morphology. We will consider apophony, suppletion, tri-consonantal roots, and tone change.

2.7.1 Apophony

Apophony (모음 교체), also called internal change, involves a sound change inside a word to signal a grammatical concept. Usually it is inflectional. Modern English exhibits a great deal of apophony in its Germanic vocabulary.

(74) Apophony in English
   a. plural: foot/feet; tooth/teeth; goose/geese
   b. tense: sing/sang/sung; come/came

Dinka (a Nilotic language spoken in South Sudan) has a similar phenomenon. Number is shown by vowel change.

(75) Dinka Apophony
   a. dom ‘field’, dum ‘fields’
   b. kat ‘frame’, ket ‘frames’

Finally, Bemba (a Bantu language spoken primarily in Zambia) has an elaborate form of apophony to indicate causation.

(76) Bemba Apophony

<table>
<thead>
<tr>
<th>Intransitive Verb</th>
<th>Causative Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>luba ‘to be lost’</td>
<td>lufya ‘to cause to be lost’</td>
</tr>
<tr>
<td>koma ‘to be deaf’</td>
<td>komya ‘to cause to be deaf’</td>
</tr>
<tr>
<td>pona ‘to fall’</td>
<td>ponya ‘to cause to fall’</td>
</tr>
</tbody>
</table>

2.7.2 Suppletion

Consider the following English verbs and their past tense forms.

(77) English past tense forms

<table>
<thead>
<tr>
<th>play</th>
<th>played</th>
<th>show</th>
<th>showed</th>
</tr>
</thead>
<tbody>
<tr>
<td>go</td>
<td>went</td>
<td>bring</td>
<td>brought</td>
</tr>
</tbody>
</table>
On regular verbs the past tense suffix -ed is added to show past tense. The past tense of go, however, changes completely to went. Suppletion (보충법) refers to the situation in which a completely different form of the root is used in a particular environment. In this example we say that the past tense form of go is the suppletive form went. The past tense of bring is also irregular; however, notice that the form is not completely different. We say that it exhibits partial suppletion.

There are two verbs in Korean that employ suppletion in the negative. To make a verb negative in the short form 안 appears to the left of the verb. Thus, the negative of 먹어요 is 안 먹어요. The verb 알다 in the present is 알아요. In the negative, however, it is 몰라요. Likewise, the negative of 있어요 is 없어요.

2.7.3 Reduplication

Reduplication (첩어) involves copying all or part of a base to signal a grammatical contrast. Copying the entire base is called full reduplication, and copying part of the base is called partial reduplication. The following Indonesian example shows full reduplication to express plurality [Sneddon et al., 2012].

<table>
<thead>
<tr>
<th>Indonesian Reduplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>piring ‘plate’</td>
</tr>
<tr>
<td>piring-piring ‘plates’</td>
</tr>
<tr>
<td>buku ‘book’</td>
</tr>
<tr>
<td>buku-buku ‘books’</td>
</tr>
<tr>
<td>rumah ‘house’</td>
</tr>
<tr>
<td>rumah-rumah ‘houses’</td>
</tr>
<tr>
<td>orang ‘person’</td>
</tr>
<tr>
<td>orang-orang ‘people’</td>
</tr>
</tbody>
</table>

Swahili uses full reduplication on verbs to indicate frequentative aspect. Note that reduplication indicates other additional properties, but we show frequentative aspect here [Ojambo 2010].

<table>
<thead>
<tr>
<th>Swahili Reduplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>piga ‘to strike’</td>
</tr>
<tr>
<td>pigapiga ‘to strike repeatedly’</td>
</tr>
<tr>
<td>tupa ‘to throw’</td>
</tr>
<tr>
<td>tupatupa ‘to throw repeatedly’</td>
</tr>
<tr>
<td>sema ‘to speak’</td>
</tr>
<tr>
<td>semasema ‘to speak repeatedly’</td>
</tr>
<tr>
<td>cheza ‘to play’</td>
</tr>
<tr>
<td>chezacheza ‘to play repeatedly’</td>
</tr>
</tbody>
</table>

American Sign Language (ASL) uses full reduplication to derive nouns from verbs. Later we will see more details in the structure of ASL. Individual morphemes are typically written in CAPS. Consider the following examples [Supalla and Newport, 1978]. An example of ‘to sit’ and ‘chair’ are shown in Figure 2.2. The verb ‘to sit’ consists of a motion in which the index finger and the middle finger of the right hand are brought down to the index finger and middle finger of the right hand. This sign is reduplicated to derive the word for ‘chair’.

<table>
<thead>
<tr>
<th>ASL Reduplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>To see videos of these and other signs visit American Sign Language University</td>
</tr>
</tbody>
</table>
Marshallese is an Austronesian language spoken on the Marshall Islands. Marshallese employs partial reduction to express the concept of wearing an article of clothing (Moravcsik, 1978).

(81) Marshallese Partial Reduplication

<table>
<thead>
<tr>
<th>kagir</th>
<th>‘belt’</th>
<th>kagirgir</th>
<th>‘to wear a belt’</th>
</tr>
</thead>
<tbody>
<tr>
<td>takin</td>
<td>‘sock’</td>
<td>takinkin</td>
<td>‘to wear socks’</td>
</tr>
</tbody>
</table>

English possesses contrastive focus reduplication (Ghomeshi et al., 2004). This kind of reduplication indicates a prototypical token of the type indicated by the base. Here is an example.

A: I want a salad.

B: We have some fruit salad.

A: No, I mean a salad-salad.
There are several kinds of salad; however, a prototypical salad contains lettuce, tomatoes, carrots and other vegetables. When speaker A says that they want a ‘salad-salad’, they are indicating that they want a prototypical salad and not a fruit salad or potato salad or any other kind of salad.

2.7.4 Triconsonantal Roots

Semitic languages contain triconsonantal roots, sometimes called triliteral roots. Consider the following Arabic words. Note that the apostrophe ’ indicates a glottal stop. It functions as any other letter. The symbol d represents the "th" sound as in English this. Observe that each root consists of three consonants with various vowels (and sometimes other consonants) intervening. Try to find the set of three consonants for each root.

(82) Arabic Triconsonantal Roots

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>kataba</td>
<td>‘he wrote’</td>
</tr>
<tr>
<td>qara’a</td>
<td>‘he recited’</td>
</tr>
<tr>
<td>kadaba</td>
<td>‘he lied’</td>
</tr>
<tr>
<td>katabtu</td>
<td>‘I wrote’</td>
</tr>
<tr>
<td>qarab’u</td>
<td>‘I recited’</td>
</tr>
<tr>
<td>kaďabtu</td>
<td>‘I lied’</td>
</tr>
<tr>
<td>yaktubu</td>
<td>‘he writes’</td>
</tr>
<tr>
<td>yaqra’u</td>
<td>‘he recited’</td>
</tr>
<tr>
<td>yakdibu</td>
<td>‘he lies’</td>
</tr>
<tr>
<td>aktubu</td>
<td>‘I write’</td>
</tr>
<tr>
<td>aqra’a</td>
<td>‘I recite’</td>
</tr>
<tr>
<td>akdiбу</td>
<td>‘I lie’</td>
</tr>
</tbody>
</table>

The root for ‘write’ is k-t-b. The root for ‘recite’ is q-r-. Finally, the root for ‘lie’ is k-d-b. Note that the word Qur’an (usually spelled without ’ in English) is derived from the root q-r-.

2.7.5 Prosodic Morphemes

Sometimes grammatical information is manifested prosodically. We will look at stress and tone here. We discuss tone in more detail in the next chapter, but we illustrate the concept here.

First, lexical category in English is often indicated by stress change. Consider the following pairs of words, where the acute accent indicates primary stress. In bisyllabic words, nouns are usually stressed on the first syllable and verbs are usually stressed on the second syllable. There is some speaker variation in these forms, however. Some speakers can use décrease and incréase with initial stress as either a noun or a verb.

(83) Lexical Categories in English and Stress Change

<table>
<thead>
<tr>
<th>noun</th>
<th>verb</th>
<th>noun</th>
<th>verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>récord</td>
<td>recórd</td>
<td>cónflict</td>
<td>conflíct</td>
</tr>
<tr>
<td>présent</td>
<td>presént</td>
<td>próduse</td>
<td>producé</td>
</tr>
<tr>
<td>décrease</td>
<td>decréase</td>
<td>incréase</td>
<td>incréase</td>
</tr>
<tr>
<td>réject</td>
<td>réjéct</td>
<td>súspect</td>
<td>suspéct</td>
</tr>
</tbody>
</table>

As mentioned, tone is discussed in more detail in the next chapter. Tone refers to the pitch of the vowel. Many Bantu languages use pitch grammatically. In the following example tone is used to indicate plural number.
on nouns. The grave accent (â) indicates low tone and the acute accent (á) indicates high tone. High tone on the final two syllables of a word indicates plural number in Ngiti.

(84) Ngiti number marked by tone ([Lojenga, 1994] p. 135)

<table>
<thead>
<tr>
<th></th>
<th>singular</th>
<th>plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>teacher</td>
<td>màlimò</td>
<td>malímó</td>
</tr>
<tr>
<td>chief</td>
<td>kamà</td>
<td>kámá</td>
</tr>
<tr>
<td>angel</td>
<td>màlàyikà</td>
<td>màlàyiká</td>
</tr>
</tbody>
</table>
Answers to Practice Questions

Practice Question 1.
Korean: 1, 3 먹는 다, 2 하 면, 4 가르 치 어 요, 3 일반 적 으로, 6 반 으 시 었 십 니 다, 4 가르 치 다, 5 오 었 슬 니 까

English: 2 house-s, 1, 3 alien ate-s, 2 drip ed, 3 un deny able, 3 devil ish ly, 2 withdraw al, 5 de crim in al ize ation

Practice Question 2.
lente - slow; triste - sad; limp - clean; -mente - adverb marker; -eza - state or quality of
-o - masculine noun; -a - feminine noun; rapid - fast; bel - beautiful; note that 'triste' becomes 'trist' when it appears with the suffix -eza.

Practice Question 3.

Practice Question 4.

Practice Question 5.
Exercises

Question 1. Draw morpheme trees for the following English words: unreliable, defroster, inapplicability, treatments, de-ionization, deactivation, impossibility, re-enactment

Question 2. Draw morpheme trees for the following English compounds. Note that some are ambiguous: frying-pan cleaner, steel knife sharpener, biology textbook illustrator, ferry boat restaurant, physics laboratory apparatus clerk, Korean teacher application form

Question 3. Draw morpheme trees for the following Korean words: 먹었어요, 거머잡았다.

Question 4. Consider the following from Snoop Doggy Dogg’s *Tha Shiznit* (1993):

"The surgeon is Dr. Dr[IZ]e (Dre) / so I[IZ]ay (lay) and pl[IZ]ay (play) / with D O double G[IZ]ee (G) / the fly human being"

What kind of affix is /iz/? How does it attach? Consider the following data as well.

<table>
<thead>
<tr>
<th>bottle</th>
<th>b[IZ]ottle</th>
<th>ahead</th>
<th>ah[IZ]ead</th>
</tr>
</thead>
<tbody>
<tr>
<td>dollars</td>
<td>d[IZ]ollars</td>
<td>behave</td>
<td>beh[IZ]ave</td>
</tr>
<tr>
<td>Google</td>
<td>G[IZ]oogle</td>
<td>effect</td>
<td>eff[IZ]ect</td>
</tr>
<tr>
<td>soldiers</td>
<td>s[IZ]oldiers</td>
<td>exchange</td>
<td>exch[IZ]ange</td>
</tr>
<tr>
<td>weddings</td>
<td>w[IZ]eddings</td>
<td>surprise</td>
<td>surpr[IZ]ise</td>
</tr>
</tbody>
</table>

What kind of affix is /iz/? How does it attach? Consider the following data as well.

<table>
<thead>
<tr>
<th>bottle</th>
<th>b[IZ]ottle</th>
<th>ahead</th>
<th>ah[IZ]ead</th>
</tr>
</thead>
<tbody>
<tr>
<td>dollars</td>
<td>d[IZ]ollars</td>
<td>behave</td>
<td>beh[IZ]ave</td>
</tr>
<tr>
<td>Google</td>
<td>G[IZ]oogle</td>
<td>effect</td>
<td>eff[IZ]ect</td>
</tr>
<tr>
<td>soldiers</td>
<td>s[IZ]oldiers</td>
<td>exchange</td>
<td>exch[IZ]ange</td>
</tr>
<tr>
<td>weddings</td>
<td>w[IZ]eddings</td>
<td>surprise</td>
<td>surpr[IZ]ise</td>
</tr>
</tbody>
</table>

Question 5. Consider the following Korean data. What kind of morphological process is taking place here? Describe how the forms in list B are formed from the base in list A. Note that the forms in list B indicate a temporal extension of the sound indicated in list A. Thus, in the first example /sik/ refers to crispy noise and /sisik/ refers to crispy noise that lasts longer. For this first part consider only the data in the first table. Data from Jun (1994). Note that in the Romanization ‘wu’ represents the [u] sound as in moo and who. See the table for the Romanization of Korean on page 3 for more details.

<table>
<thead>
<tr>
<th>list A</th>
<th>list B</th>
</tr>
</thead>
<tbody>
<tr>
<td>/sik/</td>
<td>식 'crisp'</td>
</tr>
<tr>
<td>/cik/</td>
<td>직 (sound of tearing)</td>
</tr>
<tr>
<td>/twung/</td>
<td>동 ‘boom’</td>
</tr>
<tr>
<td>/pwung/</td>
<td>봉 ‘vroom’ (sound of car engine)</td>
</tr>
</tbody>
</table>

Consider now the following additional data. Do you have to revise your answer? Explain what morphological process is taking place here and how it works taking into account all the data in the question.
Question 6. Identify the morphemes in the following Pévé data (adapted from [Shay and Wambadang, 2019]). Note that the tones have been omitted. The symbol \( \odot \) represents a reduced vowel.

<table>
<thead>
<tr>
<th>list A</th>
<th>list B</th>
</tr>
</thead>
<tbody>
<tr>
<td>/asak/</td>
<td>아삭 ‘crunch’</td>
</tr>
<tr>
<td>/wacak/</td>
<td>와삭 ‘munch’</td>
</tr>
<tr>
<td>/ppituk/</td>
<td>빙득 ‘creak’</td>
</tr>
<tr>
<td>/cwulwuk/</td>
<td>주룩 ‘vroom’ (sound of car engine)</td>
</tr>
</tbody>
</table>

Question 7. Identify all the morphemes in the following Cree data. (Note, there is no morpheme that corresponds to it in these data. The colon indicates that the vowel is long; however, it does not play a role in the answer to this question.)

| niwa:pahte:n | I see it. | nitapin | I sit. |
| kiwa:pahte:n | You (SG) see it. | kitapin | You (SG) sit. |
| niwa:pahte:na:n | We see it. | nitapina:n | We sit. |
| kiwa:pahte:na:wa:w | You (PL) see it. | kitapina:wa:w | You (PL) sit |
| nima:chishe:n | I cut it. | |
| kima:chishe:n | You (SG) cut it. | |
| nima:chishe:na:n | We cut it. | |
| kima:chishe:na:wa:w | You (PL) cut it. | |

Question 8. Consider the following Uzbek data. List all the morphemes present and provide all allomorphs and their conditioning environments. Note that ‘y’ is a consonant, as in English yes. State the order of the morphemes in the Uzbek noun. Note there is no morpheme that corresponds to English the.

| kitobingdan | ‘from your book’ | opamga | ‘to my sister’ | uyimda | ‘in my house’ |
| opalarim | ‘my sisters’ | olmangda | ‘in your apple’ | kitobimga | ‘to my book’ |
| opangdan | ‘from your sister’ | uyingga | ‘to your house’ | uyimdan | ‘from my house’ |
| kitoblaringdan | ‘from your books’ | uylarga | ‘to the houses’ | olmang | ‘your apple’ |
| kitoblarda | ‘in the books’ | olmam | ‘my apple’ | omalarim | ‘my apples’ |
Further Reading

- **Harley (2017)** - excellent introduction to English word structure

- **Katamba and Stonham (2006)** - introductory textbook on morphology, commonly used in undergraduate classes

- **Shi (2015)** - overview chapter on Korean word formation

- **Aronoff (1994)** - This is an advanced theoretical discussion advocating for treating morphology as a module distinct from phonology and syntax. Although the technical discussion is advanced the description of the morphological data is clear and accessible.
Chapter 3

Phonetics

By the end of this chapter you should be able to:
- understand the basic principles of articulatory phonetics,
- be able to read the IPA chart,
- be aware of acoustic and perceptual phonetics
- know and understand the features of consonants, vowels, and tones,
- be familiar with phonetic variation among the world’s languages,
- be able to transcribe words in your native language and/or other languages.

3.1 Introduction

Phonetics is the study of the physical gestures used to produce language. Traditionally, phonetics involved the study of the sounds of language; however, the last fifty to sixty years of research has revealed that signed languages are processed in the mind/brain the same way spoken languages are. Phonetics has consequently evolved to consider the gestures used in signed languages. Spoken languages, then, are composed of discrete units of sound, and signed languages are composed of discrete units of manual and facial gestures. These units often blend together into larger components that cannot be separated. For instance a vowel and its tone cannot be pulled apart into two units of speech. Nevertheless, the vowel and its tone can be identified as distinct units. Likewise, in signed languages manual gestures can be broken down into components such as palm orientation and finger aperture. Phonetics is the study of how these bits of sound or bits of movement in the case of signed languages are put together. It is also the study of the properties of these sounds. We identify three major fields of study within phonetics.

- **articulatory phonetics** the study of how sounds are produced
- **perceptual phonetics** the study of how sounds are perceived
- **acoustic phonetics** the study of the physical properties of sounds

We will deal only with articulatory phonetics here. Historically, the study of phonetics involved training in
the recognition of the sounds of the world’s languages. Although the phonetician’s laboratory now includes a vast array of apparatus to study, basic training on qualitative recognition of speech sounds is still a valuable tool. You can consult some of the websites in the further reading section at the end of the discussion for some useful websites to help you train your ear.

3.2 The Vocal Tract

The vocal tract is roughly similar across many chordates. Nevertheless, various evolutionary adaptations in humans have made speech possible (Fitch 2000). Note that such adaptations have come at a cost, however. A lowered pharynx provides us with the opportunity to articulate a much wider array of speech sounds, but it put our esophagus and trachea so close together that we often choke, sometimes fatally, on the food we eat (Fitch, 2000).

Let us begin this discussion by introducing the articulators of the vocal tract. The earliest modern descriptions of the articulators of the oral tract can be traced back to Sweet (1877). The articulators are shown in figure 3.2. We distinguish between the tip of the tongue, the blade of the tongue (just behind the tip), and the body of the tongue (the remainder). In some languages the distinction between the tip and the blade is important, but in both English and in Korean they can be used interchangeably and vary from one speaker to the next. We will not worry about the distinction here.

We move on to articulators that are less commonly known. The alveolar ridge (치조제) is just behind the upper teeth. If you feel with the tip of your tongue you should feel a distinct bump or ridge leading up to the roof of the mouth. Not everyone has a distinct alveolar ridge. Curling your tongue further back you should reach the highest point in your mouth. This is a hard, bony surface called the palate or hard palate (경구개). If your tongue is flexible enough you should be able to curve it a bit further back to feel a soft, fleshy surface
Figure 3.2: Vocal Apparatus
near the back of your mouth called the *velum or soft palate* (연구개). A very small number of people can curl their tongues far back enough to touch the tip of the *uvula* (목젖) with their tongues. This is a small flap of tissue which looks a bit like a punching bag. You should be able to see it in the mirror. The glottis is the space between the two vocal folds (the flaps of tissue that cover the trachea). It is located inside the larynx.

### 3.3 The Sounds of the World’s Languages

Sounds in spoken languages are produced with airflow. There are three *airstream mechanisms* in human language (Pike [1943]). Airflow from the lungs is called *pulmonic* and is found in all spoken languages. In some languages, the velum is closed and the speaker raises the larynx. This pushes air out through the mouth. This is called *glottalic* airflow. Finally, the tongue can be used to create negative air pressure pockets in the mouth. This is called *velaric* airflow. Additionally, airflow can be directed outwards, *egressive* or can be directed inwards, *ingressive*. Let’s go through each of these in turn.

Pulmonic egressive sounds are found in virtually all of the world’s spoken languages. Both English and Korean make use of only pulmonic egressive sounds, with some very minor exceptions noted below. If you make an extended [s] sound you will notice your lungs gradually getting emptier and emptier. Spoken human languages typically do not use pulmonic ingressive airflow. A gasp to show surprise is usually articulated with pulmonic ingressive airflow, but that is a paralinguistic act. We discuss clicks below, but we mention here that the Khoisan language !Xôô has nasalized clicks in which the nasal articulation is produced with pulmonic ingressive airflow.

Glottalic egressive sounds are more commonly known as *ejectives*. They are produced by closing the vocal folds and raising the larynx, thus forcing air out through vocal passage. Ejectives are found in about 20% of the world’s languages. The Siouan language Lakhota exhibits ejectives. You can listen to these sounds at the [UCLA Phonetics Lab Data website](https://phoneticslab.ucla.edu).

(1) a. [p’o] ‘foggy’  
   b. [pəYo] ‘mallard’  
   c. [k’u] ‘to give’  
   d. [kah] ‘that’

Figure 3.3: Pulmonic Airflow

Glottalic ingressive sounds are more commonly known as *implosives*. They are produced by closing the vocal folds and lowering the larynx. This creates a slight vacuum in the oral cavity. When the glottis is released air is drawn into the lungs across the glottis creating an implosive consonant.

Velaric egressives do not exist in human language. Velaric ingressives are known as *clicks*. Consider the
following Xhosa examples. Again, to hear these examples go the the [UCLA Phonetics Lab Data website](http://www.cah.ucsd.edu/labs/phonetics/data/).

(2) a. [ukúkəla] ‘to gine fine’
   b. [ukúkəlo] ‘to break stone’
   c. [úkolo] ‘peace’

Table 3.1 shows the classification of sounds based on airstream mechanisms.

Sounds are typically divided into consonants and vowels. Consonants are articulated with complete or partial closure of the air tract. Vowels are made with no turbulent airflow. We describe each below in turn.

### Practice 3.1: How many sounds are there in the following English and Korean words? Pay close attention to the discrepancy between English orthography and English sounds. For example, the word *through* has seven letters but only three sounds.

| 돈, 좋아, 양, 밥, 곤, 닭이, 없다, 사과, 별로, 선생님, 슈퍼마트, 동동주; dog, cats, back, shell, knife, cringe, after, water, thought, doughnut, dripped, laughed, funny, courtesy, flinched, elephants, giraffe

### 3.3.1 Consonants

Consonants are described in three ways: (i) laryngeal features, (ii) place, and (iii) manner. We run through each of these in turn. Observe the consonant chart in Figure 3.4. The places of articulation are shown across the top row. The manners of articulation are shown down the left column. Differing laryngeal features are shown within the same cell. As noted in the chart, when there are two symbols in the same cell, the symbol on the left is voiceless and the symbol on the right is voiced. This happens to correspond to distinctive laryngeal features in English; however, as we will see for Korean, the set of laryngeal features differs from one language to the next.

**Laryngeal Features:** Laryngeal features refer to how the vocal folds and the larynx are used to produce speech sounds. English sounds are either *voiceless* (무성음) or *voiced* (유성음). Voiceless sounds include [s] and [t]. Voiced sounds include [z] and [d]. If you make an extended [z]-sound you should feel your vocal folds vibrating if you put your fingers over your larynx, see Figure 3.5 on page 71. Practice saying the following words. The initial sounds all contrast for voicing. The first word begins with a voiceless sound, and the second word begins with a voiced sound.
**THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)**

<table>
<thead>
<tr>
<th>CONSONANTS (PULMONIC)</th>
<th>© 2005 IPA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plosive</strong></td>
<td></td>
</tr>
<tr>
<td>Bilabial</td>
<td>p b</td>
</tr>
<tr>
<td>Labiodental</td>
<td>t d</td>
</tr>
<tr>
<td>Dental</td>
<td>t d j</td>
</tr>
<tr>
<td>Alveolar</td>
<td>k g</td>
</tr>
<tr>
<td>Postalveolar</td>
<td>q g</td>
</tr>
<tr>
<td>Retroflex</td>
<td>?</td>
</tr>
<tr>
<td><strong>Nasal</strong></td>
<td>m n</td>
</tr>
<tr>
<td>Trill</td>
<td>B r</td>
</tr>
<tr>
<td><strong>Tap or Flap</strong></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td></td>
</tr>
<tr>
<td>Lateral Fricative</td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>u j</td>
</tr>
<tr>
<td>Lateral approximant</td>
<td>l j</td>
</tr>
</tbody>
</table>

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

**Figure 3.4: Consonants of the IPA chart**

(3) **Voicelessness versus voicing on initial sound.**

<table>
<thead>
<tr>
<th>feel</th>
<th>veal</th>
</tr>
</thead>
<tbody>
<tr>
<td>thigh</td>
<td>thy</td>
</tr>
<tr>
<td>sink</td>
<td>zinc</td>
</tr>
<tr>
<td>chin</td>
<td>gin</td>
</tr>
</tbody>
</table>

Korean has three laryngeal features: **voiceless** (usually called *plain, lenis, or lax* (무기음) in descriptions of Korean phonetics and phonology), **aspirated** (유기음), and **fortis** (also called *tense*. We will follow tradition here and call the voiceless consonants plain. Plain consonants include [p] as in 반, [t] as in 단, and [k] as in 간.

Aspirated consonants include [pʰ] as in 판, [tʰ] as in 탄, and [kʰ] as in 간. Fortis consonants include [pʰ] as in 빵, [tʰ] as in 팻, and [kʰ] as in 까지. Observe that the superscript ‘ʰ’ is used to indicate aspirated sounds. There is no official diacritic to indicate Korean fortis sounds. Cross-linguistically, they are rare. Peter Ladefoged has used [ʰ] to indicate fortis sounds, so we will follow this convention.

**Place**: Looking across the top of the consonant section of the IPA chart we see the places of articulation from the front of the vocal tract to the back. We run through them in turn.

**Bilabial** (양순음) sounds are articulated with both lips. Bilabial sounds include [p], [b], [m], [n], [φ], and [β]. English words that begin with bilabial sounds include *pit, bit, and mitt*. Korean words that begin with bilabial sounds include 반, 판, 밥, and 만. Notice how both lips are involved in making the initial sounds in these words. We’ll cover the other sounds later.

**Labiodental** (순치음) sounds are articulated with the lower lip and the upper teeth. English has the labiodental sounds [f] and [v]. Korean does not have labiodental sounds.

**Dental** (치음) sounds are articulated with the tip or the blade of the tongue and the teeth. English examples
include [θ] as in *thin* and [ð] as in *this*. Korean examples include [ɑ] as in 단, [tʰ] as in 탄, and [tʰ] as in 팔. Observe that the symbol [t] is used to indicate a dental sound. A dental stop is also found in some English words such as *width* [wɪdθ]

*Alveolar* (치경음) sounds are articulated with the tip or the blade of the tongue and the alveolar ridge. English has numerous alveolar consonants, including [t], [d], [n], [s], [z], [l], and [s] as in *run*. Furthermore, [r], called a *flap* is a common variant of [t] and [d] in many dialects. You can hear it in the word *butter* in most North American dialects of English.

Note, the Korean coronal sounds, [t], [n], and so forth, have been reported as dental by some authors, alveolar by some authors and as intermediate by some authors. For convenience we will notate these as alveolar [t], although the dental pronunciation is also found in certain positions. Compare the ㄴ-sound /n/ in 안 with 나타.

*Postalveolar* (후치경음) sounds are made with the tip or blade of the tongue and the area of the roof of the mouth just behind the alveolar ridge. Postalveolar sounds in English include [ʃ] as in *ship* and [ʒ] as in *beige*.

*Retroflex* (권설음) sounds involve the the tip of the tongue curled up to the highest point of the roof of the mouth. Some English speakers produce an *r*-sound as a retroflex, as in her [hɜ]. Korean does not possess retroflex sounds. If you have studied Mandarin you will have come across many retroflex sounds, such as the initial [ʂ] as in *是*.

*Palatal* (경구개음) sounds are produced with the body of the tongue and the hard palate. Both English and Korean have one palatal sound, [j] as in *yes* and 요.

*Velar* (연구개음) sounds involve the body of the tongue and the velum or soft palate. Examples include [k] as in *ski* and 간, [g] as in *gate*, and [ŋ] as the final sound in *king* and 강.

*Uvular* (구개수음) and *pharyngeal* (인두음) sounds are not found in either English or Korean. A uvular nasal is found in Japanese, in particular in word-final position before a pause (such as at the end of a sentence). An example is *Nihon* [ɲihoŋ] (*Japan*). Standard French has a uvular trill, *Paris* [paʁi], but the vast majority of speakers use a voiced or voiceless uvular fricative, [paʁi].

*Glottal* (성문음) sounds involve the vocal folds. Both English and Korean have the sound [h] as in *high* and 한.

**Manner**: The first column of the IPA chart lists the various manners of articulation.

*Plosives* (파열음) are produced by completely blocking the flow of air in the vocal tract and then releasing the pressure built up. The sound produced by suddenly releasing this pressure is called a plosive or a stop consonant. English plosives include [p] as in *spin*, [b] as in *bin*, [t] as in *stun*, [d] as in *done*, [k] as in *school*, and [g] as in *ghoul*. Korean plosives include [p] as in 반, [pʰ] as in 판, [pʰ] as in 반히, [t] as in 단, [tʰ] as in 단,
[t*] as in 람.

**Nasals** (비강음) are produced by stopping the flow of air through the oral tract, lowering the velum, and allowing air to escape through the nose. Nasals include [m] as in *mat* and 만, [n] as in 나, [ŋ] as in *gnat*, and [ŋ] as in *sing* and 상.

**Trills** (전동음) are produced rapidly vibrating an articulator. Trills are not used in Korean and are found only in some dialects of English. In some Scottish dialects ’r’ is pronounced as a trill, [r].

**Taps or flaps** (타음을) are produced rapidly flapping an articulator against a surface. Both English and Korean have an alveolar flap, [r] as in *autumn* (in some dialects of English) and 얼음.

**Fricatives** (마찰음) are made by forming a tight constriction along the vocal tract creating turbulent airflow. English has a number of fricatives, including [f] as in *fine*, [v] as in *vine*, [θ] as in *thin*, [ð] as in *they*, [s] as in *sing*, [z] as in *zing*, [ʃ] as in *assure*, [ʒ] as in *azure*, and [h] as in *hat*. Korean has three fricatives: [s] as in 산, [s*] as in 산값, and [h] as in 한. ㅅ is typically articulated as [sʰ] by most speakers, especially in word-initial position (Chang, 2013; Cho et al., 2002).

**Approximants** (접근음) are consonants with no turbulent airflow. One type of approximant is the glide, such as [w] in *water* and 원주민, [j] in *yes* and 예, and [ŋ] in 꺾 and French *oui* (‘yes’). The other type of approximant is the liquid such as [l] as in *rabbit* and [l] as in *mill* and 밥.

**Affricates** (파찰음) are made with a complete stoppage of airflow through the vocal tract followed by a release with turbulent airflow. Phonetically, they are described as a plosive followed by a fricative. As such they are considered a sequence of two independent sounds and thus do not appear on the IPA chart. English has two affricates: [ʧ] as in *cheap* and [ʤ] as in *jeep*. Korean has three affricates: [?] as in 잔, [t?] as in 차, and [t*] as in 잔.

**Practice 3.2:** Each set of sounds below has one member that does not belong. What is the consonant that does not belong, and what phonetic property do the other sounds have in common?

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>t</td>
<td>d</td>
<td>n</td>
<td>r</td>
<td>v</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>p</td>
<td>b</td>
<td>n</td>
<td>c</td>
<td>k</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>b</td>
<td>m</td>
<td>b</td>
<td>p</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>b</td>
<td>t</td>
<td>m</td>
<td>z</td>
<td>g</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>j</td>
<td>m</td>
<td>nj</td>
<td>n</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>s</td>
<td>n</td>
<td>β</td>
<td>ʃ</td>
<td>ɣ</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>q</td>
<td>g</td>
<td>ɣ</td>
<td>ɦ</td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>l</td>
<td>[ə]</td>
<td>j</td>
<td>l</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The descriptions above provide the way to give the phonetic descriptions of phonetic symbols. Here are some examples. Pay close attention to the order of the descriptors.
Figure 3.6: Vowels of the IPA chart

**Order of phonetic descriptors:**
1) laryngeal feature
2) place
3) manner

[m] voiced bilabial nasal

[f] voiceless labiodental fricative

[k] voiceless velar plosive

[pʰ] aspirated bilabial plosive

[s*] fortis alveolar fricative

**Practice 3.3:** Identify the consonants in the following words.

돈, 좋아, 알, 밤, 곤, 닭이, 않다, 사과, 별로, 선생님, 슈퍼마트, 동동주; dog, cats, back, shell, knife, cringe, after, water, thought, doughnut, dripped, laughed, funny, courtesy, flinched, elephants, giraffe

### 3.3.2 Vowels

Vowels are classified in terms of height, backness, and roundedness. In some languages, including English, vowels are also classified in terms of tenseness/laxness.
**Height:** The IPA chart identifies four height categories. Very few languages actually require four height distinctions, however, and most pedagogical descriptions of vowels (including both English and Korean) make-do with three. We list the IPA terms and the general terms here. Hereafter we will use the general terms.

<table>
<thead>
<tr>
<th>IPA</th>
<th>general</th>
</tr>
</thead>
<tbody>
<tr>
<td>close</td>
<td>high</td>
</tr>
<tr>
<td>close-mid</td>
<td>mid</td>
</tr>
<tr>
<td>open-mid</td>
<td>low</td>
</tr>
<tr>
<td>open</td>
<td>low</td>
</tr>
</tbody>
</table>

Table 3.2: IPA and General Descriptions of Vowel Height

*High* vowels (고모음) are pronounced with the tongue in a high position (‘close’ to the roof of the mouth). Both Korean and English have the high vowel [i] as in *ski* and 스키. For classification purposes, the vowels i, y, and u are classified as high vowels. The IPA term for these vowels is ‘near close’.

*Mid* vowels (중모음) are pronounced with the tongue in a mid position. Both English and Korean have the mid vowel [o] as in *sew, know* and 조교.

*Low* vowels (저모음) are pronounced with the tongue in a low position (with the mouth ‘open’). For classification purposes, the vowels u and a are classified as low vowels.

**Backness:** Backness refers to where the highest point of the tongue is in the mouth with respect to the lips. Either the front, centre, or back of the tongue is the highest. Front and back vowels are very common.

In *front* vowels (전설 모음) The tip and blade of the tongue are raised higher than the rest of the tongue. You might feel as though you are pushing your tongue forward to accomplish this. Front vowels include [i] as in *ski* and 스키, [e] as in *bed* and 개, and [æ] as in *cat*.

In *central* vowels (중설 모음) the body of the tongue is raised higher than the rest of the tongue.

In *back* vowels (후설 모음) the back of the tongue is raised. You should feel the tip of your tongue being retracted into your mouth when you articulate an [u] sound.

**Roundedness:** Vowels can be rounded or unrounded. This feature refers to the rounding of the lips.

In *rounded* vowels (원순 모음) the vowels are in a rounded state. If you say the word *coo* (the sound of a pigeon) or 후후 you should notice that your lips are rounded.

In *unrounded* vowels (비원순 모음) the lips are spread or in a relaxed state. If you make a clear [i] sound as in *ski* or 스키, you should notice that your lips are unrounded or spread.

**Tense/Lax:** *Tense* vowels (긴장 모음) and *lax* vowels (이완 모음) in English do not reliably correlate to any phonetic property. The English vowels [i], [u], [e], [o], and [ə]/[u] are tense. The vowels [ɪ], [ʊ], [ɛ], [æ], and [æ] are lax. Dutch and German make a similar distinction for tense and lax vowels. It was once thought that the tense/lax distinction correlates to tongue root position (advanced tongue root versus retracted tongue root), but it has been shown that correlation does not hold up ([Ladefoged and Maddieson](1996)). The phonetic property of tongue root placement is an important property in many languages, such as Bantu languages, but...
it turns out not to be the factor that distinguishes tense and lax vowels in English.) The tense vowels are also
sometimes referred to as long vowels, and the lax vowels as short vowels; however, phonetic vowel length does
not reliably correlate to the tense/lax distinction. To add to the confusion, in languages that make use of tense
and lax vowels, their distribution is not the same from one language to the next. In general, the close vowels
(high vowels), /i y u u/ are tense, and the row just below (called ‘near close’), /i y a/ are lax. The close-mid
vowels are usually tense, and the open-mid vowels are generally lax, though /a/ is tense in English.

Finally, vowels can differ in length. Long vowels are indicated with a diacritic that looks similar to a colon.
Here is a long open front unrounded [aː]. Japanese has a clear distinction in vowel length. Consider the following
example.

(4) Japanese Vowel Length
  a. [itsuki] ‘moon’; [itsuːki] ‘airflow’
  b. [oːdʒiːsaŋ] ‘uncle’; [oːdʒiːsaŋ] ‘grandfather’

Seoul Korean used to have a distinction between long and short vowels; however, this distinction has been nearly
lost (Kang, Yoonjung et al., 2015). For example, [nun] ‘eye’ and [nuːn] ‘snow’ are pronounced identically as
[nun] by virtually speakers of Seoul Korean. The Jeolla dialect of Korean maintains a length contrast, however
(Jun, 1998).

Practice 3.4: Each set of sounds below has one member that does not belong. What is the consonant that does
not belong, and what phonetic property do the other sounds have in common? Refer to the phonetic properties
in Figure 3.6 on page 73

- i y u u o
- e i e y
- y u u e
- a e u e
- i o e y

3.3.3 Diphthongs

Diphthongs (어중모음) formed by a sequence of a vowel and a glide, in either order. There are three diphthongs
in English as follows.

(5) English Diphthongs

[au] high [hau], ride [raɪd], live [laɪv]
[au] how [hau], loud [laʊd], boughs [baʊz]
[oi] boy [boɪ], noise [noɪz], voice [voɪs]
Diphthongs are sometimes written with glides instead of vowels. Thus, [aɪ] is sometimes written as [aj]; [aʊ] is sometimes written as [aw]; and [oɪ] is sometimes written as [oj]. Furthermore, the quality of these diphthongs (like vowels in general) has been changing over times. The English word *by* is often pronounced as [bae] in contemporary speech.

Diphthongs in Korean consist of a glide followed by a vowel. Examples include 귀 [kqi], 뭐 [mwâ], 왕 [wan], and 의 [tti].

**Practice 3.5:** Identify the vowels in the following words.

돈, 좋아, 알, 밤, 끝, 달이, 없다, 사과, 별로, 선생님, 슈퍼마트, 동동주; dog, cats, back, shell, knife, cringe, after, water, thought, doughnut, dripped, laughed, funny, courtesy, flinched, elephants, giraffe

### 3.4 Phonetic Features of Signed Languages

The gestures used in signed languages are described in much the same way as in spoken languages (Brentari, 2019; Stokoe et al., 1965; Sandler and Lillo-Martin, 2006). The following parameters have been accepted for the phonetic and phonological description of signed languages. We distinguish between the dominant hand (the hand the signer uses for most gestures, often the right hand) and the non-dominant hand (the other hand).

1. **hand shape** - includes properties of which fingers are used and how they are shaped (concave, spread, or compact)
2. **movement** - includes direction (to or from signer) and repetition (single or multiple)
3. **place of articulation** - head, trunk, non-dominant hand, arm. These features can be modified by a setting such as [hi], [lo], [proximal], etc.
4. **orientation** - includes features such as weather the palm of the dominant hand is facing up or down

Consider the ASL sign for ‘head’ in Figure 3.7. The hand shape is [spread], movement is [single], place of articulation is [head] and a contour feature [hi lo] (meaning the movement is from the hi part of the head to the low part of the head), and the orientation is downward.

### 3.5 Tones

Many languages around the world use pitch to create contrastive sounds. Tones are either **level tones** or **contour tones**. The pitch of a level tone does not change, while the pitch of a contour tone does. Tones are traditionally divided into five possible level tones, where 5 indicates the highest pitch (Chao, 1930). The following IPA symbols are used.

1. Image from American Sign Language University

76
(6) Level Tones
   a. 5
   b. 4
   c. 3
   d. 2
   e. 1

It is very rare for a language to have more than three level tones. Thus, any given language makes use of only a subset of the tones in (6). Contour tones are described as a sequence of two or more level tones. Most contour tones contain only two level tones; however, as we will see for Mandarin below, it is possible to have a contour tone that contains a sequence of three level tones. Mandarin and Cantonese have several contour tones, so we will turn to examples from these two languages.

Mandarin is probably the most well known language that makes use of pitch to create tones. Mandarin has four tones as follows. The first tone is a level tone and the other three tones are contour tones. They are indicated by the beginning and end points. Consider the second tone. It starts at 3 (around the middle of the tonal register) and ends at 5 (the highest point of the tonal register). Thus, it is a rising tone. The third tone is a complex contour tone that starts at 2, then goes down to 1, then rises to 4. As mentioned, complex contour tones are quite rare cross-linguistically.

(7) Mandarin Tones
a. 妈 'mother' [ma:]  
b. 麻 'hemp' [ma:]  
c. 马 'horse' [ma:]  
d. 骂 'scold' [ma:]  

Peter Ladefoged’s website has an example of the four tones of Mandarin:  
[http://www.phonetics.ucla.edu/vowels/chapter2/chinese/recording2.1.html](http://www.phonetics.ucla.edu/vowels/chapter2/chinese/recording2.1.html)

Cantonese has six tones (although in the traditional Chinese linguistic tradition it has nine based on the historical derivation of the tones). Cantonese has three level tones and three contour tones.

(8) Cantonese Tones

a. 詩 'poem' [si:]  
b. 屎 'excrement' (vulgar) [si:]  
c. 試 'try' [si:]  
d. 時 'time' [si:]  
e. 市 'city' [si:]  
f. 事 'matter' [si:]  

To hear the tones of Cantonese, visit Peter Ladefoged’s website:  
[http://www.phonetics.ucla.edu/vowels/chapter2/cantonese/recording2.2.html](http://www.phonetics.ucla.edu/vowels/chapter2/cantonese/recording2.2.html)

Finally, we will look at an example of tones in Ibibio ([Urua, 2004](#)). Ibibio has two tones, a high level tone and a low level tone. The IPA notations are [a] and [ã], respectively; however, since there are only two tones, the high and low tones are notated as [á] and [à], respectively, for convenience. Consider the following examples.

(9) Ibibio Tones

dá ‘stand’  
dà ‘friend’  
kóp ‘scoop food’  
kàp ‘lock’  
jómmó ‘be pregnant’  
jómmó ‘boo at’

Related to tone is the notion of pitch accent. Pitch accent is the use of pitch to make one syllable more prominent in a word. First, consider the English words dessert and desert. The only difference between these two words is the placement of stress. Stress is on the second syllable in dessert and on the first syllable in desert. In a pitch accent language a sequence of tones is spread across a word. The difference between tonal languages and pitch accent languages is that tone languages associate a tone or sequence of tones to a syllable, and pitch accent languages associate a tone or sequence of tones to a word. Pitch accent languages include Norwegian, Swedish, Japanese, Turkish, and some dialects of Korean. Consider the following Swedish example ([Cruttenden, 1997](#), p.11).

78
<table>
<thead>
<tr>
<th>Consonant Type</th>
<th>Bilabial</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>p pʰ pʷ</td>
<td>t tʰ tʷ</td>
<td>ᵍ k kʰ kʷ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td>ṭ</td>
<td></td>
<td></td>
<td>ṭ̊</td>
</tr>
<tr>
<td>Tap or flap</td>
<td></td>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td>ṭ̊</td>
</tr>
<tr>
<td>Fricative</td>
<td></td>
<td>s sʰ</td>
<td>h</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td></td>
<td>j</td>
<td>u̍</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>w - labio-velar approximant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>η - labio-palatal approximant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[j] [i̱] [i̱ʰ] - postalveolar affricates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.8: Korean Consonants of the IPA chart

(10) a. ¹anden LHL ‘the duck’
    b. ²anden HLHL ‘the spirit’

There are two pitch accents in almost all dialects of Swedish (and Norwegian), identified by the numerals 1 and 2. The phonetic forms of the two pitch accents differ from dialect to the next.

Finally, consider the following examples from Kyengsang Korean (Kim and Jun, 2009).

(11) a. kaci HL 가지 ‘type’
    b. kaci HH 가지 ‘branch’
    c. kaci LH 가지 ‘eggplant’

For disyllabic words Kyengsang Korean has the tone patterns HL, HH, and LH, which are distributed across the two syllables.

3.6 The Transcription of Korean

We will briefly consider how to transcribe Korean, starting with the consonants, which are less controversial, see Figure 3.8. In addition to the consonants on this chart, Korean also has three affricates: [ʃ] as in 잔, [ʃʰ] as in 차, and [ʃʰ] as in 전.

The vowel inventory of Korean is quite controversial. This is due in part to ongoing changes and vowels shifts (as is common in all languages). In traditional Korean grammars, Korean is described has having ten vowels. In actual speech, however, there are seven vowels. These are shown in Figure 3.9. Korean orthography is much more consistent than English orthography, allowing us to easily draw a correspondence between the orthographic form in Korean and the phonetic transcription.

Let’s begin with a simple illustration. The following words illustrate each of the seven vowels on the chart.
One of the more controversial aspects of Korean phonology concerns the representations of the vowels in the words 개 and 게. Historically, these vowels were distinct and had the values [e] in 개, and [e] in 게. For virtually all speakers in South Korea, now, these two vowels have merged (Shin, 2015). Historically, 위 was pronounced [y] (Shin, 2015). This pronunciation is still advocated by the 국립국어원; however, the pronunciation [qi] is also prescribed. For virtually all speakers, then, 귀 is pronounced [ki] rather than [ky].

Korean also has the following diphthongs: 와 [wa], 워 [wa], 와, 웰, 워 [we], 위 [qi], 야 [ja], 여 [ja], 워, 웰 [je], 요 [jo], 유 [ju], and 의 [ui].

With these basics in place we can begin to transcribe Korean words phonetically. Examine the following examples of transcribed Korean words. Take careful note of how sounds change. We will talk about sounds changes much more when we discuss phonology. For now, we will simply take note of sound changes as we encounter them. Observe that ㅅ normally corresponds to an [s] sound. In the first example, however, it corresponds to a [ʃ] sound. Note also the difference between the two ㄷ’s in 동동주. We will cover these sound changes later.

3.7 The Transcription of English

Let’s consider the inventory of consonants in English, shown in Figure 3.10. As mentioned, there are also two affricates in English: [ʃ] as in chin and [dʒ] as in gin. English consonants are quite stable in all dialects. However, there are some dialects in which [θ] and [ð] have merged with either [s] and [z] or with [t] and [d],
<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Postalveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td>k</td>
<td>g</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>f</td>
<td>v</td>
<td>θ</td>
<td>draulic</td>
<td>s</td>
<td>z</td>
<td>j</td>
<td>3</td>
</tr>
<tr>
<td>Approximant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>j</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral approx.</td>
<td>ɾ</td>
<td>j</td>
<td>ɾ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Others
w - labio-velar approximant
ʃ ʒ - postalveolar affricates

Figure 3.10: English Consonants of the IPA chart

respectively. The sounds [θ] and [ð] can be tricky for some non-English speakers. Here are some examples to practice.

(12) a. words beginning with [θ]: think, thin, through, Thursday, three
   b. words beginning with [ð]: this, that, though, then, they

Another pair of sounds that some people find tricky is [ʃ] and [ʒ]. Note that [ʒ] never appears word-initially in English. Again, here are some words containing these two sounds.

(13) a. words containing [ʃ]: shoe, she, ash, assure, Aleutian
   b. words containing [ʒ]: beige, fission, vision, azure, allusion

Although English consonants are quite stable across dialects, English orthography often belies the phonetic form of words. For instance, "of" has a [v] sound rather than an [f] sound. The words cough, though, through, enough, and bough all end in ough, but all five of these words are pronounced differently.

There are various reasons for the discrepancy between English orthography and the phonetic form. One is that English does not have a prescriptive authority like the 국립국어원 for Korean or the Académie Française for French. Thus, there is no authority to update the spelling system after changes in the sound system. An extreme example of this is words like knife and knee, where the initial <k> has not been pronounced since the 15th century. Another reason is that the orthographic forms often give important clues to the morphological relatedness of words. For instance, the <g> in sign is silent; however, it is audible in the related word signature.

Like Korean, English has a relatively large vowel inventory. The vowels are shown in Figure 3.11. The vowel [ə] is not found in all dialects. Note, too, that of the low back vowels, either [ɑ] or [ɔ] is found, depending on the dialect, but never both. There is a great deal of diversity in the vowel inventory in English dialects, especially in the lower vowels.

In addition to the vowels in Figure 3.11, there are three diphthongs in English: [aɪ], [aʊ], and [ɔɪ]/[ɔi]. The last diphthong varies dialectally.

The examples below illustrate the vowels of English in various dialects. The following abbreviations are
Table 3.3: Vowels of English

<table>
<thead>
<tr>
<th>word</th>
<th>SCE</th>
<th>SAmE</th>
<th>RP</th>
<th>SAusE</th>
</tr>
</thead>
<tbody>
<tr>
<td>bee</td>
<td>[bɪt]</td>
<td>[bɪ:]</td>
<td>[bɪ:]</td>
<td>[bɪ:]</td>
</tr>
<tr>
<td>bit</td>
<td>[bɪt]</td>
<td>[bɪt]</td>
<td>[bɪt]</td>
<td>[bɪt]</td>
</tr>
<tr>
<td>bait</td>
<td>[bɛɪt]</td>
<td>[bɛɪt]</td>
<td>[bɛɪt]</td>
<td>[bɛɪt]</td>
</tr>
<tr>
<td>bet</td>
<td>[bɛt]</td>
<td>[bɛt]</td>
<td>[bɛt]</td>
<td>[bɛt]</td>
</tr>
<tr>
<td>trap</td>
<td>[tɾæp]</td>
<td>[tɾæp]</td>
<td>[tɾæp]</td>
<td>[tɾæp]</td>
</tr>
<tr>
<td>bath</td>
<td>[bæθ]</td>
<td>[bæθ]</td>
<td>[bæθ]</td>
<td>[bæθ]</td>
</tr>
<tr>
<td>palm</td>
<td>[pʌm]</td>
<td>[pʌm]</td>
<td>[pʌm]</td>
<td>[pʌm]</td>
</tr>
<tr>
<td>lot</td>
<td>[lɒt]</td>
<td>[lɒt]</td>
<td>[lɒt]</td>
<td>[lɒt]</td>
</tr>
<tr>
<td>cloth</td>
<td>[klɔθ]</td>
<td>[klɔθ]</td>
<td>[klɔθ]</td>
<td>[klɔθ]</td>
</tr>
<tr>
<td>thought</td>
<td>[θɔt]</td>
<td>[θɔt] or [θɒt]</td>
<td>[θɔt]</td>
<td>[θɔt]</td>
</tr>
<tr>
<td>goose</td>
<td>[ɡuːs]</td>
<td>[ɡuːs]</td>
<td>[ɡuːs]</td>
<td>[ɡuːs]</td>
</tr>
<tr>
<td>foot</td>
<td>[fʊt]</td>
<td>[fʊt]</td>
<td>[fʊt]</td>
<td>[fʊt]</td>
</tr>
<tr>
<td>price</td>
<td>[prɪs]</td>
<td>[prɪs]</td>
<td>[prɪs]</td>
<td>[prɪs]</td>
</tr>
<tr>
<td>prize</td>
<td>[praɪz]</td>
<td>[praɪz]</td>
<td>[praɪz]</td>
<td>[praɪz]</td>
</tr>
<tr>
<td>choice</td>
<td>[ʧɔs]</td>
<td>[ʧɔs] or [ʧɔːs]</td>
<td>[ʧɔs]</td>
<td>[ʧɔs]</td>
</tr>
<tr>
<td>goat</td>
<td>[ɡoʊt]</td>
<td>[ɡoʊt]</td>
<td>[ɡoʊt]</td>
<td>[ɡoʊt]</td>
</tr>
<tr>
<td>strut</td>
<td>[strʌt]</td>
<td>[strʌt]</td>
<td>[strʌt]</td>
<td>[strʌt]</td>
</tr>
<tr>
<td>bird</td>
<td>[bɜːd]</td>
<td>[bɜːd]</td>
<td>[bɜːd]</td>
<td>[bɔd]</td>
</tr>
<tr>
<td>about</td>
<td>[əbɔr]</td>
<td>[əbɔr]</td>
<td>[əbɔr]</td>
<td>[əbɔːr]</td>
</tr>
<tr>
<td>aloud</td>
<td>[əlaʊd]</td>
<td>[əlaʊd]</td>
<td>[əlaʊd]</td>
<td>[əlæʊd]</td>
</tr>
</tbody>
</table>

Table 3.3: Vowels of English

used in Table 3.3: SCE = Standard Canadian English, SAmE = Standard American English, RP = Received pronunciation (Roach, 2004), SAusE = Standard Australian English.

A few comments are in order. First, the examples above give only a fraction of the variation attested in these four varieties of English. Note also that we have left out several other varieties of English such as Singapore English, Indian English, New Zealand English, South African English, just to mention a few. Within each of the four varieties above there are several dialects; however, the most general properties are given above. The following website has some interesting discussion with several audio examples on the vowels for nurse, strut, comma, and goat.

https://www.englishspeechservices.com/blog/strut-for-dummies/

Let’s practice with a few examples of English words. Consider the following examples.
Again, we’ll mention some of the variation here, but go over it in more detail when we discuss phonology. In North American varieties of English, there is a distinction between [l] as in let and [I] as in tell. The initial vowel in entrapment is [ɛ] in careful speech, but [ə] in casual or fast speech. For now, just try to hear the difference, and we’ll discuss it in more detail when we cover phonology.
Answers to Practice Questions

Practice Question 1.
Korean: 3, 3, 2, 3, 3, 4 or 5 [타기] or [달기], 4, 5, 5, 9, 8, 8
English: 3, 4, 3, 3, 3 (diphthong counts as one sound), 5, 4, 4, 3, 5, 4, 4, 6, 6, 8, 5

Practice Question 2.
- v does not belong - the rest are alveolar
- n does not belong - the rest are plosive
- f does not belong - the rest are bilabial
- t does not belong - the rest are voiced
- l does not belong - the rest are nasal
- b does not belong - the rest are fricatives
- j does not belong - the rest are lateral

Practice Question 3.
Korean: [tn], [ʧn], [l], [pm], [kʰn], [tj] or [tlj], [pʃ], [skw], [pʃl], [sʌn], [ʃpʲtʰ], [tʃdʒ]
English: [dg], [kts], [bk], [fl], [km], [ft], [wt], [θt], [dt], [dpt], [ft], [fn], [krs], [fnf], [lng], [dʒf]

Practice Question 4.
- o does not belong - the rest are close
- θ does not belong - the rest are front
- θ does not belong - the rest are rounded
- θ does not belong - the rest are open
- i does not belong - the rest close-mid

Practice Question 5.
Korean: [tn], [ʃoa], [al], [pm], [kʰm], [tai] or [talgi], [apʰ], [sakwa], [pjlo], [sʌn], [ʃp^h am^h ut], [təŋdɔŋdʒu]
English: [dag] or [dɔg] or [dɔg], [karts], [bæk], [ʃel], [kmdʒ], [æft], [wɔt] or [wot], [θt] or [θot] and [θut], [dɔnt], [dapt], [læft], [fani], [krtsi], [fnf], [lüf], [dʒæf]
Exercises

Question 1. Give a narrow transcription for the following English words: cat, dog, mouse, water, sun, of, this, bread, June, am, that, jumped, strong, grief, dive, love, gate, ski, dread

Question 2. Give a narrow transcription for the following Korean words: 판, 물, 알아, 알려, 마늘, 프랑스어, 깨잎, 치과, 소주, 김밥

Question 3. Give a narrow transcription for the following English words: rhinoceros, giraffe, dreadfully, Belgian endive, encyclopedia, microscope, telephone, terrific, festival, ferocious

Question 4. For each set of sounds below, indicate which one of the sounds does not belong with the rest. In some cases, there is more than one possible answer for each set. Give all possible answers and explain. The first one is done for you.

1. a e i o u k (answer: [k] does not belong. It is a consonant and the rest are vowels.)
2. a i y w i
3. d p b g j
4. s f v z k
5. s z ð d v
6. p b m n ß
7. m n ñ n ñ d
8. i e u u
Further Reading

- **Rogers (2014)** - This textbook provides a clear and in depth introduction to phonetics.
- **Ogden (2017)** - This is a short introduction to phonetics that concentrates on English.
- **Catford (2001)** - This practical introduction to phonetics has several exercises for practicing sounds.
- **Shin et al. (2012)** - This is an in depth introduction to the sounds of Korean.
- **Choo and O’Grady (2003)** - This is an introductory discussion of the sounds of Korean.

useful websites:

- Korean Phonetics and Phonology: [https://dx.doi.org/10.1093/acrefore/9780199384655.013.176](https://dx.doi.org/10.1093/acrefore/9780199384655.013.176)
- UCLA Phonetics Lab Data: [http://www.phonetics.ucla.edu/index.html](http://www.phonetics.ucla.edu/index.html)
Chapter 4

Phonology

By the end of this chapter you should:

- understand the difference between a phoneme and an allophone
- determine the distribution of allophones
- write phonological rules using phonological features
- know and understand the main kinds of phonological processes in the world’s languages
- understand basic syllable structure
- be comfortable with determining morphophonological alternations

4.1 Introduction

The previous chapter dealt with the articulatory features of the world’s sounds and signs. It is clear to anyone who has ever studied a foreign language that different languages make use of different subsets of all the available sounds. Furthermore, how these sounds are arranged differ from one language to the next. Whereas phonetics is the study of the linguistic properties of sounds and signs, phonology is the study of how these sounds and signs are organized within a given language or dialect. Research in phonology has revealed that languages tend to avail themselves to the same kinds of phonological rules. Thus, phonology is also the study of the universal grammar of sounds and signs. At the end of this chapter we bring together concepts of morphology and phonology and discuss how they interact.

4.2 The Phoneme

As we just said above, different languages have different sets of sounds. But how do figure out what the set of sounds is in a given language? To this day, numerous languages around the world are lacking a basic description. And even in better studied languages there are rare dialects that have not been adequately recorded. In the early 20th Century, structural linguists such as Ferdinand de Saussure, Leonard Bloomfield and André Martinet
developed a methodology for determining the set of sounds for a given language or dialect (Bloomfield, 1933). They relied on the notion of contrast to guide their inquiry. They proposed the following definition.

**Phoneme:** The smallest segmental unit of sound employed to form meaningful contrasts between utterances

Leonard Bloomfield was an American linguist in the early 20th century who specialized in Indo-European, Austronesian, and Algonquian languages. He was the founder of structuralism, which characterizes the approach to phonology presented in this chapter. He conducted much fieldwork on Algonquian languages in North America.

A “segmental unit of sound” is the smallest isolable bit of sound. Consider the English word “dog”. You can articulate the [d], [a], and [g] separately and still tell that they are the sounds of the word “dog”. Consider just the [d], however. It is a voiced alveolar stop. You can’t articulate the voicing separately from the rest of the sound leaving just a [t] and some vague voiced grunt and still expect people to know that these are the sound components of the word “dog”. The [d] is an atomic package as far as determining the sounds of a word (though we will see below that phonology can manipulate the features of sounds and change them around). Thus, [d] is a segmental unit of sound. Now, what do we mean by ‘meaningful contrasts’? Consider the following pairs of words along with rough phonetic transcriptions.

(1) Examples of Minimal Pairs
   a. Korean: [makt*a] ‘clear, clean’; [nakta] ‘old, used’
   b. English [fli] ‘flea’; [flu] ‘flew’

Notice that each pair of words differs by only one segment. Such pairs are called **minimal pairs** (최소 대립상). Looking at the Korean example, the sounds [m] and [n] give rise to a contrast. That is, exchanging one for the other gives rise to a difference in meaning. Thus, we say that [m] and [n] are distinct **phonemes** (음소) in Korean. In English, the vowels [i] and [u] are contrastive because the give rise to the minimal pair shown. Likewise, the sounds [s] and [z] are contrastive in Portuguese. Figure 4.1 shows three minimal pairs for ASL.\(^1\) The signs for ‘please’ and ‘sorry’ are a minimal pair for hand shape.

The **phonemic inventory**, or set of basic sounds, of a language can be deduced by comparing several minimal pairs of the language. In actual practice, hundreds of minimal pairs should be contrasted to ensure you have an accurate and complete phonemic inventory. Some sounds in a language could be quite rare. For instance, the sound [g] is somewhat rare in English, occurring only in a few words (beige, illusion, vision, etc.). Indeed, it is quite difficult to construct a minimal pair for [f] and [z] in English. There are at least two such pairs, however:

\(^1\)Image from Emmorey et al. (2003)
Figure 4.1: Minimali Pairs in ASL
Aleutian/allusion and mesher/measure. Sometimes, a set of data may be limited, in which case we can rely on **near minimal pairs**.

<table>
<thead>
<tr>
<th>Minimal Pair</th>
<th>Near Minimal Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>[mr̓[f̓]́]</td>
<td>[mr̓[z̓]́]</td>
</tr>
<tr>
<td>[polufan]</td>
<td>[koluṣon]</td>
</tr>
</tbody>
</table>

This difference is far enough away from the segments being tested that we can be fairly certain that [f̓] and [z̓] are contrastive.

Figure 4.2: Minimal and Near Minimal Pairs

**Practice 4.1:** Find as many minimal pairs as you can in the following Ainu data. Try to determine as many phonemes as you can based on minimal pairs.

| [arkə] | ‘to be in pain’ |
| [arki] | ‘to come’       |
| [hani] | imperative particle |
| [kusu] | ‘because’       |
| [sapa] | ‘head’          |
| [tap]  | ‘shoulder’      |
| [kuru] | ‘person’        |
| [mat]  | ‘wife’          |
| [tan]  | ‘this’ (demonstrative) |
| [tup]  | ‘two’           |

| [arpə] | ‘to go’        |
| [tsup] | ‘sun’          |
| [tani] | ‘now’          |
| [tsaka] | ‘to open’     |
| [poro] | ‘big’          |
| [mak]  | ‘mountain side’ |
| [sapo] | ‘older sister’ |
| [poru] | ‘cave’         |
| [etu]  | ‘nose’         |
| [atu]  | ‘rope’         |

### 4.3 Allophonic Variation

Let’s begin by looking at the sounds [s] and [ʃ] in Korean. Consider the following data.

(2) Korean data

| [sawgwa] | ‘apple’    | [ʃiptʰa] | ‘to want’ |
| [kasal]  | ‘temporary installation’ | [ʃip] | ‘ten’ |
| [mɛʃɛɾ]  | ‘oath’     | [kaʃi]  | ‘thorn’  |
| [suwun]  | ‘Suwon’    | [ʃilka] | ‘Silla’  |
| [isa]    | ‘director’ | [maʃiɾa] | ‘to drink’ |
| [sori]   | ‘sound’    | [tuʃiɾa] | ‘to eat (hon.)’ |
Looking through this data you will notice that there are no minimal pairs for [s] and [ʃ]. That is, there is no environment where you could find either one of these two sounds. Examining the data further you will notice that the distribution is predictable. To figure out the distribution, it is helpful to build a chart listing the sounds that appear immediately before and after the sound in question. In the left-hand column we write the sound that immediately precedes the sound in question, and in the right-hand column we write the sound that immediately follows the sound in question. If the sound is at the beginning or end of a word, we use the symbol, #, to mark a word boundary.

<table>
<thead>
<tr>
<th>[s]</th>
<th>[ʃ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>a</td>
</tr>
<tr>
<td>a</td>
<td>η</td>
</tr>
<tr>
<td>η</td>
<td>ε</td>
</tr>
<tr>
<td>#</td>
<td>u</td>
</tr>
<tr>
<td>i</td>
<td>a</td>
</tr>
<tr>
<td>#</td>
<td>o</td>
</tr>
</tbody>
</table>

Figure 4.3: Environments in which [s] and [ʃ] are found

The distribution of [s] and [ʃ] is now quite clear. [ʃ] appears before the vowel [i], and [s] appears elsewhere. Note that the preceding sound does not play a role since there are some environments that are the same for the preceding sound. When two or more sounds occur in a predictable distribution we call them *allophones*. Thus, in Korean [s] and [ʃ] are allophones of a single phoneme. By convention, the label of the phoneme is the allophone with the widest distribution. [ʃ] appears before /i/, and [s] appears elsewhere, so we can also call [s] the elsewhere case. The distribution of allophones is typically represented in a tree as follows. The following tree is read as follows: “/s/ becomes [ʃ] when it appears before /i/ and become [s] when it appears elsewhere.”

```
[s] / elsewhere

[ʃ] / __ i

/s/ / __ i

Figure 4.4: Distribution of allophones
```

This alternation can also be captured by a phonological rule, as shown here. This rule is read as “‘ess’ becomes ‘esh’ when it appears before ‘i’.”

```
/s/ → [ʃ] / __ i

Figure 4.5: Phonological Rule
```

Note carefully that the allophones appear in square brackets and that the phoneme appears between slashes. Please remember this for the discussion below.
Phonological rules

<table>
<thead>
<tr>
<th>Rule</th>
<th>Condition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>/x/ → [y] / __ a</td>
<td>&quot;x becomes y when it appears before a&quot;</td>
<td></td>
</tr>
<tr>
<td>/y/ - input</td>
<td>C - any consonant</td>
<td></td>
</tr>
<tr>
<td>[y]- output</td>
<td>V - any vowel</td>
<td></td>
</tr>
<tr>
<td>abc - conditioning environment</td>
<td># - word boundary</td>
<td></td>
</tr>
<tr>
<td>__ - position of input</td>
<td>Ø - empty</td>
<td></td>
</tr>
<tr>
<td>/x/ → [y] / a __ b</td>
<td>$ - syllable boundary</td>
<td></td>
</tr>
<tr>
<td>“x becomes y when it appears between a and b”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/x/ → Ø / __ a</td>
<td>Note: The symbols C and V will be revised in section 4.4</td>
<td></td>
</tr>
<tr>
<td>”delete x before a”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ø → [x] / __ a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>”insert x before a”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Practice 4.2: For each rule below, determine the output for the various inputs given to the right of the rule.

1. /p/ → [b] / __ V /pat/ /pit/ /tap/ /apt/ /plit/ /apro/
2. /t/ → [s] / C __ i /pti/ /pta/ /kti/ /ti/ /tik/ /pto/
4. Ø → [t] / V __ V /ia/ /tat/ /foa/ /ona/ /beana/ /fiat/

Let us continue by looking at the sounds [t], [ʃ], and [ɕ] in the following (slightly simplified) Japanese data. Although vowel length is indicated in this data, assume that it does not play a role here.

(3) Japanese data

[tatami] ‘mat’ [keisatsu] ‘police’
[hoteru] ‘hotel’ [isuu] ‘when’
[hito] ‘person’ [ʃi:sai] ‘small’
[ʃikai] ‘near’ [takai] ‘expensive’
[toi] ‘far’

We see that there are no minimal pairs, so we make a chart to examine the environments in which these phones are found.
Finally, segments may exist in free variation. A minimal pair may be found; however, there is no difference in meaning. In English plain t, [t], and unreleased t, [t^*], are in free variation in word-final position. Thus, the word hat can be pronounced [hæt] or [hæt^*]. Although segments in free variation do not give rise to differences in lexical meaning, they may give rise to sociolinguistic distinctions such as speech style (casual, monitored, or formal) or speaker relations (honorific or intimate).

Practice 4.3: Consider the following Warao data, a language isolate spoken in Venezuela. Are the phones r and d allophones of the same phoneme or two separate phonemes?

<table>
<thead>
<tr>
<th>Warao</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>[dahi]</td>
<td>‘a kind of bee’</td>
</tr>
<tr>
<td>[hikiri]</td>
<td>‘fly’</td>
</tr>
<tr>
<td>[daisa]</td>
<td>‘another’</td>
</tr>
<tr>
<td>[domu]</td>
<td>‘bird’</td>
</tr>
<tr>
<td>[dani]</td>
<td>‘his/her father’</td>
</tr>
</tbody>
</table>

4.4 Phonological Features and Rules

We will start this section with a brief look at some (slightly simplified) Korean data. Pay particular attention to the oral stop consonants (aside from [t^*]). The distribution charts have been made below the data for your convenience.

(4) Korean Intersonorant Voicing

<table>
<thead>
<tr>
<th>Korean</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>[tadari]</td>
<td>‘monthly’</td>
</tr>
<tr>
<td>[trsagwan]</td>
<td>‘embassy’</td>
</tr>
<tr>
<td>[makt*a]</td>
<td>‘to eat’</td>
</tr>
<tr>
<td>[kada]</td>
<td>‘to go’</td>
</tr>
<tr>
<td>[abu]</td>
<td>‘flattery’</td>
</tr>
<tr>
<td>[kaunga]</td>
<td>‘riverside’</td>
</tr>
</tbody>
</table>
Observe that a pattern emerges only once we consider what appears on both sides of the segment in question. Consider the following allophone distribution charts and their associated phonological rules. The rules seem somewhat complex until we consider what the triggering segments all have in common.

(5) Distribution of allophones of Korean oral stops

<table>
<thead>
<tr>
<th>[t]</th>
<th>[p]</th>
<th>[k]</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>a</td>
<td># a</td>
</tr>
<tr>
<td>#</td>
<td>e</td>
<td>#</td>
</tr>
<tr>
<td>a</td>
<td>t</td>
<td>a</td>
</tr>
<tr>
<td>a</td>
<td>t</td>
<td>a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[d]</th>
<th>[b]</th>
<th>[g]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>n</td>
<td>a</td>
<td>w</td>
</tr>
<tr>
<td>o</td>
<td>a</td>
<td>u</td>
</tr>
<tr>
<td>l</td>
<td>a</td>
<td>a</td>
</tr>
</tbody>
</table>

(6) \( /t/ \rightarrow [d] / \{l,n,V\} \_ V \)  \( /p/ \rightarrow [b] / \{n,V\} \_ V \)  \( /k/ \rightarrow [g] / \{\eta,V\} \_ \{V,w\} \)

Notice that both sides of the triggering environment always have a voiced sound. Rather than list an assortment of sounds as the triggers of voicing, let’s assume that there is a phonological feature \([+\text{voice}]\) that triggers voicing. We can re-write the rules in a much simpler way. We use the abbreviation \([+\text{voi}]\) as an abbreviation.

(7) \( /t/ \rightarrow [d] / [+\text{voi}] \_ [+\text{voi}] \)  \( /p/ \rightarrow [b] / [+\text{voi}] \_ [+\text{voi}] \)  \( /k/ \rightarrow [g] / [+\text{voi}] \_ [+\text{voi}] \)

Now, if we consider these rules more closely we see that they describe the same general process for \(/t/, /p/ and /k/\). Is there a general statement we can make about these three rules? It seems as though voiceless consonants become voiced between two voiced segments. We can write a single rule as follows to account for the distribu-
tion of all six sounds. Such rules, where segment becomes voiced between two voiced segments is extremely common in the world’s languages and is referred to as **intersonorant voicing**.

\[(8) \{t, p, k\} \rightarrow [+v oi] / [+v oi] _ [+v oi]\]

We may ask if the sounds in the input all have something in common. We could come up the following rule to simplify the rule in (8).

\[(9) \{-v oi\} \rightarrow [+v oi] / [+v oi] _ [+v oi]\]

We can test the validity of this rule by looking at the behaviour of other voiceless sounds in Korean. Consider the following additional data (note also the word for ‘embassy’ in the set of data in (4) above).


Here, /s/ does not become voiced to become [z]. Thus the rule in (9) must be revised. What do /t/, /p/, and /k/ have in common in contrast to /s/? /s/ is a fricative in contrast to the other sounds, which are stops or plosives. Phonologists use the feature [+continuant] to describe any sound with a constant flow of air through the oral tract. The rule that captures all these facts is shown below. (We will talk more about phonological features below.) /s/ has a constant flow of air through the oral tract, so it is [+continuant]. /t/, /p/, and /k/ do not have a constant flow of air through the oral tract, so they are [-continuant]. We will discuss these features more below.

\[(11) [-continuant] \rightarrow [+voice] / [+voice] _ [+voice]\]

Let’s consider for a moment why we might want to state the rule for intervocalic voicing in Korean as in (11) as opposed to as in (6). There are two important reasons for this. First, we want to capture a generalization about the phonology of Korean. Specifically, we want to say that intersonorant voicing is a property of oral stop consonants in general. The rule in (11) captures this generalization. Second, we want to be able to write only those rules that are likely to be found in natural language. Consider the following two hypothetical rules.

\[(12) \text{a. } /s/ \rightarrow [z] / _ \{i, a, u\} \quad \text{b. } /s/ \rightarrow [q'] / _ \{r, w, o\}\]

The first rule is one which is quite plausible, while the second is extremely unnatural, and unlikely to be found in any language. Both of these rules are very easy to write in their current forms. However, if we refer only to phonological features, then only the first rule is easy to write. We will come back to this point below.

Let us now examine the various phonological features that are found in natural language. Recall from the previous chapter that phonetic features refer to how sounds are articulated. They are constant across languages. Phonological features, on the other hand, capture the behaviour of sounds in a particular language, and thus are subject to cross-linguistic variation. The descriptions below indicate the expected phonological features.
found on the segments discussed. Furthermore, not all features are active in every language. If a language has no lateral sounds, for example, it would be redundant to label every phoneme in the language [-lateral]. Since these are the expected properties of the sounds listed below evidence to the contrary is needed to propose a feature specification that departs from these expectations. For example, if a given language has the phoneme /l/, we assume that it is [+continuant] unless there is evidence to the contrary. Most of the important features are given below, along with their standard abbreviations. This list of features is not exhaustive, however. Generally speaking, the more phonemes a language has, the more features are necessary to describe the phonology of the language. In the following charts, commonly used abbreviations are presented below each entry.

**Consonant Features - Manner**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>±v</td>
<td>refers to the activity of the vocal folds</td>
</tr>
<tr>
<td>[v]</td>
<td>[±voice] – vocal folds are vibrating</td>
</tr>
<tr>
<td>[-v]</td>
<td>[±voice] – vocal folds are open</td>
</tr>
<tr>
<td></td>
<td>See the discussion of the Korean data in (4) for an illustration of this feature.</td>
</tr>
<tr>
<td>±c</td>
<td>refers to airflow in the oral tract</td>
</tr>
<tr>
<td>[c]</td>
<td>[+continuant] – constant airflow in the oral tract</td>
</tr>
<tr>
<td>[-c]</td>
<td>[-continuant] – airflow is stopped along the oral tract, includes plosives, nasals, and affricates on the IPA chart</td>
</tr>
<tr>
<td></td>
<td>Trills and flaps vary in their specification of [±cont]. See the rule in (11) for an illustration.</td>
</tr>
<tr>
<td>±s</td>
<td>refers to turbulence in airflow in the vocal tract</td>
</tr>
<tr>
<td>[s]</td>
<td>[+sonorant] – non-turbulent air-flow, includes nasals, trills, flaps, approximants, and vowels on the IPA chart</td>
</tr>
<tr>
<td>[-s]</td>
<td>[-sonorant] – strongly turbulent or blocked airflow</td>
</tr>
<tr>
<td></td>
<td>The term <strong>obstruent</strong> refers to those sounds that are [-sonorant]. Sonorants are typically always voiced.</td>
</tr>
<tr>
<td>±n</td>
<td>refers to the position of the velum</td>
</tr>
<tr>
<td>[n]</td>
<td>[+nasal] – lowered velum allowing airflow through the nasal cavity</td>
</tr>
<tr>
<td>[-n]</td>
<td>[-nasal] – raised velum allowing airflow only through the oral cavity</td>
</tr>
<tr>
<td></td>
<td>The nasal sounds on the IPA chart and all other sounds with the [ ] diacritic are [+nasal].</td>
</tr>
<tr>
<td>±c</td>
<td>refers to constriction along the vocal tract</td>
</tr>
<tr>
<td>[c]</td>
<td>[+consonantal] – enough constriction along the vocal tract to produce at least some turbulence</td>
</tr>
<tr>
<td>[-c]</td>
<td>[-consonantal] – no such constriction along the vocal tract, includes vowels, glides and glottal sounds on the IPA chart</td>
</tr>
<tr>
<td>±d</td>
<td>refers to length of time to release airflow</td>
</tr>
<tr>
<td>[d]</td>
<td>[+delayed release] – airflow is released slowly, includes all affricates and fricatives</td>
</tr>
<tr>
<td>[-d]</td>
<td>[-delayed release] – all other sounds</td>
</tr>
</tbody>
</table>
Consider the following carefully transcribed English data. Note that . indicates a syllable boundary in IPA transcriptions. (The symbol $ is used in phonological rules.) Observe that vowels in English become nasal if they precede a nasal consonant in the same syllable. The rule below the data captures this fact. Note that we can informally use the symbol V to indicate the feature set [+voi, -cons]. We can also use the symbol C to indicate [+cons]. Recall that glottal sounds such as /h/ are actually [-cons]. Authors vary as to whether C is meant to include /h/ or not. In practice, you should indicate whether /h/ is included in C or not, if this matters for your discussion. Note that an additional consonant may appear between the nasal consonant and the syllable boundary. We indicate that this consonant is optional by placing it in brackets.

(13) English Nasalization

| [θŋk]  | ‘think’ | [θık]  | ‘thick’ |
| [səŋk] | ‘sank’   | [sæk]  | ‘sack’  |
| [lɛmp] | ‘limp’   | [lɪp]  | ‘lip’   |
| [ʃeɪm] | ‘fame’   | [ʃeɪ.mæs] | ‘famous’ |
| [dəʊm] | ‘dome’   | [hæ.mə] | ‘hammer’ |
| [sʌm]  | ‘some’   | [sæ.mə] | ‘summer’ |
| [dɹnt] | ‘dent’   | [dɹt]  | ‘debt’  |

(14) \( V \rightarrow [+\text{nasal}] / \_\_ [+\text{nasal}](C)$

Here is a final illustration. Consider the following Canadian French data, paying close attention to the alveolar consonants.

(15) Canadian French

| [tu]    | ‘all’     | [tsy]    | ‘you.SG’ |
| [lyt]   | ‘to fight’| [dzı]    | ‘of the.MASC.SG’ |
| [du]    | ‘sweet’   | [tsi]    | ‘to pull’ |
| [dɔrn]  | ‘customs’ | [ɔdʒinatər] | ‘computer’ |
| [dɔnt]  | ‘to give’ | [aktɪsf] | ‘active.MASC’ |
| [itali] | ‘Italy’   | [tsɪp]   | ‘type’   |
| [asaine] | ‘to assassinate’ | [ziʃi] | ‘ziti’ (a type of pasta) |
| [skype] | ‘occupied’| [plkn]   | ‘picnic’ |
From these data, we can determine the following phonological rules. (If you don’t see the pattern right away, construct charts as we did for example (4) above.)

\[ (16) \quad /h/ \rightarrow [\ddot{\text{e}}] / \_\{i,y\} \quad /d/ \rightarrow [\dot{\text{d}}z] / \_\{i,y\} \]

Again, we see that this is the same basic rule, so we restate it as follows. (See below for the feature [+coronal], which restricts the rule to applying only to alveolar consonants.)

\[ (17) \quad \begin{array}{c}
+\text{coronal} \\
-\text{sonorant}
\end{array} \rightarrow [+\text{del rel}] / \_\begin{array}{c}
+\text{hi} \\
+\text{front}
\end{array} \]

Next we turn to place features of consonants. Note that the following definitions represent a consensus; nevertheless, the precise definitions in some cases are a matter of ongoing research.

**Consonant Features - Place**

[±labial] refers to the participation of one or both lips in the articulation of a sound

[+lab] – one or both lips are involved in sound production

[-lab] – lips are not involved in sound production

The bilabial, labiodental, and all other sounds with a labial designation (such as [w]) are [+labial].

[±coronal] refers to the use of the tip or blade of the tongue

[+cor] – the tip/blade of the tongue is raised, this includes all dental, alveolar, postalveolar and retroflex consonants

[-cor] – the tip/blade of the tongue is in neutral position

[±anterior] refers to the place of sounds, using the alveolar ridge as a dividing line

[+ant] – all sounds articulated in front of and including the alveolar ridge

[-ant] – all sounds articulated behind the alveolar ridge

[±lateral] refers to the activity of the sides of the tongue

[+lat] – one or both sides of the tongue is lowered, includes all the laterals on the IPA chart

[-lat] – the sides of the tongue are not lowered

[±dorsal] refers to the activity of the back of the tongue

[+dor] – the back of the tongue is raised, includes palatals, velars and uvulars on the IPA chart

[-dor] – the back of the tongue is not raised

[±palatal] refers to the position of the back of the tongue

[+pal] – the back of the tongue is raised and moved forward, includes palatals

[-pal] – the back of the tongue is not raised and moved forward

Consider the following Turkish data, paying attention to the dorsal consonants. Again, the sign . indicates a
syllable boundary.

(18) Turkish Dorsal Consonants

[bak.la] ‘broad beans’ [stuk.ma] ‘squeezed’
[ok.фа] ‘caress’ [park.ta] ‘in the park’
[kʰa.fa] ‘head’ [stu.kʰu] ‘tight’
[ic.na] ‘convince’ [ec.le] ‘add’
[ec.si] ‘minus’ [cʰe.çi] ‘goat’
[cʰø.myr] ‘coal’ [i.cʰi] ‘two’

From these data we can determine that [k], [kʰ], [c], and [cʰ] are all in complementary distribution and thus are allophones of the same phoneme. The following chart gives the distribution of these four allophones. Note that any one of the four consonants could have been chosen as the else-where case; however, /k/ is the least marked (the most common cross-linguistically, roughly speaking), so it is chosen as the basic phoneme.

```
/k/ $ __ {ø,i,e} 
[c] / $ {i,e} __ $ 
[k] / elsewhere 
```

As a final illustration, consider the following Cantonese data (tones omitted), paying attention to the high vowels.

(19) High vowels in Cantonese

[sì] ‘poetry’ [tʰim] ‘sweet’
[lip] ‘elevator’ [sìk] ‘to eat’
[luŋtsʰ] ‘interest’ [tʰit] ‘iron’
[bin] ‘flat’ [dksi] ‘taxi’
[bun] ‘minus’ [jok] ‘meat’
[guwak] ‘sneaky’ [cʰoŋ] ‘heavy’
[nam] ‘male’ [bak] ‘white’

From the data presented above we can determine the following rules for [i] and [u].

(20) a. /i/ → [i] / _{ŋ,k} 
   b. /u/ → [u] / _{ŋ,k}
These two rules can be conflated into the following single rule. Note that the vowel features are introduced below, and that /g/ is absent in the conditioning environment because it is never found at the end of the word in Cantonese.

\[(21) \ [+\text{hi}] \rightarrow [-\text{tense}] / \_+[\text{dor}] \]

<table>
<thead>
<tr>
<th>Vowel Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>([\pm \text{high}]) refers to vowel height</td>
</tr>
<tr>
<td>([+\text{hi}]) – phonetically high (close) vowels: i y i u u u o u</td>
</tr>
<tr>
<td>([-\text{hi}]) – all other vowels</td>
</tr>
<tr>
<td>([\pm \text{low}]) refers to vowel height</td>
</tr>
<tr>
<td>([+\text{lo}]) – phonetically low (open) vowels: æ a æ æ æ æ u æ</td>
</tr>
<tr>
<td>([-\text{lo}]) – all other vowels</td>
</tr>
<tr>
<td>([\pm \text{back}]) refers to the highest part of the tongue</td>
</tr>
<tr>
<td>([+\text{back}]) – phonetically back and central vowels, and [a]</td>
</tr>
<tr>
<td>([-\text{back}]) – phonetically front vowels except [a]</td>
</tr>
<tr>
<td>([\pm \text{round}]) refers to lip rounding on vowels</td>
</tr>
<tr>
<td>([+\text{rnd}]) – phonetically round vowels: y y u u o o æ æ æ æ æ u æ</td>
</tr>
<tr>
<td>([-\text{rnd}]) – all other vowels</td>
</tr>
<tr>
<td>([\pm \text{tense}]) refers to vowel tenseness</td>
</tr>
<tr>
<td>([+\text{tns}]) – described as involving greater muscular tension, but precise characterization unclear</td>
</tr>
<tr>
<td>([-\text{tns}]) – described as involving less muscular tension, but precise characterization unclear</td>
</tr>
</tbody>
</table>

This feature is poorly understood. It was thought to be related to another feature, \(*\text{advanced tongue root} [\pm \text{ATR}]*. In English, the following vowels are \([-\text{tns}]\): i u æ. This feature varies from one language to the next, but see the Cantonese example in \((19)\) for an example.

Phonological rules tend to be sensitive to *natural classes*, which can be described by a small number of phonological features - often just one. In the Cantonese example above we saw that vowel laxing was triggered by \([+\text{dorsal}]\) consonants. The use of features here captures the generalization that languages refer to natural classes rather than to random sets of sounds. Consider the set of vowels in English in Figure 3.11. In English, is \{i, u, o\} a natural class? These, and only these vowels are all \([+\text{hi}]\) in English, so it is a natural class, which is identified by the feature \([+\text{hi}]\). How about the set \{i, u\}? This is not a natural class. There is no set of phonological features that can uniquely identify this set of vowels. Finally, consider the set \{i, t\}. This is a natural class. It is captured by two phonological features, \([+\text{hi}, -\text{back}]\).
Practice 4.4: What feature or features are necessary to identify the following natural classes? Refer to the consonant and vowel inventories in Chapter 3.

English: {i, i, e, e, æ}, {o, o, o, a, a}
English: {m, f, v, p, w} {f, ʃ, ɹ, ɹ}, {b, d, g, v, ð, z, ð, ɹ}
Korean: {i, u, u}, {i, r}, {r, a, a}
Korean: {p, t, k, h}, {r, j, i, w}, {t, t, ð, s, s}

Let’s look at some examples to illustrate how these features are used. Consider the following Canadian English data. These data represent a phenomenon known as the Canadian Raising (Chambers, 1973).

(22) Canadian Raising

| [aɪ] | ‘ice’ | [aɪ] | ‘eyes’ |
| [hæt] | ‘height’ | [hæd] | ‘hide’ |
| [læf] | ‘life’ | [læ] | ‘lie’ |
| [klæt] | ‘clout’ | [klæd] | ‘cloud’ |
| [læs] | ‘house’ | [læyzə] | ‘houses’ |
| [pæt] | ‘pout’ | [pæt] | ‘pow’ |

Examining these data closely we see that the diphthongs [aɪ] and [aʊ] become [aɪ] and [aʊ], respectively, before voiceless consonants. The head of the diphthong, /a/, has the feature [+low] and the head of the raised diphthong, /æ/, has the feature [-low], so we can formulate the following rule to account for the Canadian Raising.

(23) V \rightarrow [-low] / V\{-voi}\]

Practice 4.5: Consider the following Purépecha data (Foster, 1969). There is a single rule to determine when certain consonants become voiced. Please state the rule using features. Consider all consonants.

| [ʒaŋkʰu] | ‘just this much’ | [aʃu] | ‘man’ |
| [ɔŋndʒakwa] | ‘back of the neck’ | [tuɾeri] | ‘snake’ |
| [ʃipiti] | ‘alive’ | [kʰɛɾɾ ndarray] | ‘cliff’ |
| [pukuroni] | ‘to remain’ | [puruatoni] | ‘to boil something’ |
| [əmbe] | ‘thing’ | [ŋqunakwa] | ‘meat’ |
| [kanani] | ‘to be born’ | [niŋimb̥qa] | ‘his/her/their brother’ |
| [pamskwaɾɾni] | ‘to touch one’s mouth’ | [kʰwineɾʃekwa] | ‘festival’ |
| [jumindu] | ‘you make’ |
4.5 Syllable Structure

You are probably already aware that a stream of speech can be divided into units called syllables. Most of us have an intuitive sense of what a syllable is, but if you actually try to define a syllable to someone, it can be quite tricky. If we have to explain the concept, we usually resort to giving several examples to get the point across. A syllable is a phonological unit of organization, usually consisting of at least a vowel and possibly one or more consonants. More precisely, a syllable can be defined as a sonority peak, where sonority refers to vocal energy—roughly equivalent to loudness. Figure 4.8 contains an illustration of sonority peaks for the English sentence, the bird flew away. Observe that there are five sonority peaks corresponding to the five syllables of this sentence.

In many languages a syllable may consist of a sequence of a consonant-vowel-consonant, CVC. Consider the English word beet.

\[ \sigma \]
\[
\begin{array}{c}
C \\
V \\
C
\end{array}
\]

We should ask ourselves whether this structure accurately represent the facts of the organization of the syllable. We will review several lines of evidence that bear on the matter. First, try saying the sequence [ba] quickly several times in succession: ba ba ba ba. . . How about [ab]? ab ab ab ab. . . Which is easier? You will likely notice that the sequence [ba] is much easier to pronounce several times in succession than [ab].
Language games often provide interesting evidence for the underlying structure of language. Specifically, they often elucidate the structure of the syllable. Consider Pig Latin, an English based language game. This game operates on the initial consonant or consonant cluster of the first syllable of each word, although there are different “dialects” of Pig Latin. Typically the initial consonant (or consonant cluster) is placed at the end of the of the word and “ay” [e] is added (Nevins and Vaux, 2003). Consider the following examples.

(25) dog → og-day; butter → utter-bay; snow → ow-snay; banana → anana-bay

Language games around the world tend treat the initial segment of the syllable as one unit and the vowel plus final consonant as another unit.

Another property distinguishing the initial segments of a syllable from the final segments concerns sound neutralizations, which tend to happen in the final consonant much more commonly than in the initial consonant in a syllable. For example, recall that Korean distinguishes /t/, /tʰ/, /tʰʷ/, /s/, /sʰ/, /sʰʷ/, and /ʃʰ/. This contrast appears only in the initial portion of the syllable, however. In the final portion of the syllable, these sounds all become neutralized to [tʰ]. Cantonese distinguishes /p/ and /pʰ/, but again only in the initial portion of the syllable. In the final portion, this contrast is neutralized to [pʰ]. Languages typically have a smaller number of consonants available in the final portion of the syllable than in the initial portion. For example, Mandarin allows /s/, /n/, /t/ and /tʰ/ in initial portion of a syllable, but allows only /n/ in the final portion.

Speech errors often involve transposition of the initial consonant in the syllable, but rarely involve transposition of the final consonant. Normally, such errors result in gibberish; however, they can give rise to humorous alternatives called spoonerisms.

(26) You missed my history lectures → You hissed my mystery lectures.

The observations above lead to the conclusion that the vowel and final consonant together form a constituent. We say that the syllable is divided into an onset and a rhyme. The rhyme consists of the nucleus – the central sonority peak of the syllable – and the coda.

(27)

Languages clearly differ in how they structure their syllables. The description of the possible syllable structures for a language is called its phonotactics. A descriptive grammar for a language may give a complete syllabary. This is the case for Japanese, where the writing system is syllabic and the number of possible syllables is small. It is somewhat impractical for English, given the large number of possible syllables. Often, a descriptive
grammar will simply describe the possible onsets, nuclei and codas for the language under discussion.

Next we discuss the procedure for syllabification. Although the procedure below is universal, it relies on phonotactics, which varies from language to language. All languages have obligatory nuclei. Normally, this is the only required component of a syllable, although some languages have obligatory onsets, too. In many languages, the nucleus must be a vowel; however, in English, the nucleus can also be a liquid or a nasal. In Blackfoot, an aboriginal language spoken in North America, a nucleus can be either a vowel or /s/. Another universal property of syllable structure is the Maximize Onset Principle. This principle states that consonants prefer to appear in onset position rather than in coda position (recall how much easier it is to say [ba-ba-ba] rather than [ab-ab-ab]). Thus, if you have a consonant or set of consonants between two syllables, it is preferable to place them in the onset of the following syllable rather than in the coda of the preceding syllable, as long as it does not violate the phonotactic constraints of the language to do so.

1. Identify nuclei – obligatory
2. Form onsets \(\rightarrow\) "Maximize Onset Principle"
3. Form codas

Let’s look at an example with the English word *extreme*. The first step is to identify the nuclei. Although liquids and nasals can be the nucleus of the syllable, vowels must be. Also, a nasal cannot be the nucleus if it adjacent to a vowel. A liquid can only in some circumstances. Here are the nuclei for the English word *extreme*.

\[
\begin{align*}
&\text{N} \\
&\text{\small \_\_\_\small k s t i m}
\end{align*}
\]

Next, we form the onsets. The rule “Maximize Onset Principle” states that as much material as possible should appear in the onset of the syllable. Note that in English [\(t\), [\(t\), and [\(st\) are all possible onsets, but [\(kst\)] is not. Thus, [\(st\)] forms the onset of the second syllable.

\[
\begin{align*}
&\text{N} \\
&\text{\small \_\_\_\small k s t i m}
\end{align*}
\]

Finally, we form the codas. Since [\(k\)] cannot be a part of the onset of the second syllable, it must be the coda of the first syllable. Of course [\(m\)] must be the coda of the second syllable. The nucleus and the coda together form the rhyme, and the rhyme and the onset together form the syllable.
Practice 4.6: Syllabify the following English words. Consult a native speaker of English if you can to verify the pronunciations. There is some variation in the number of syllables in the word *snorkelling*.

- horse
- snorkel
- snorkelling
- mechanization
- steaming
- reflection
- barbers
- temperature
- latchstring

Depending on the phonotactics of the language, further phonological operation may be necessary to syllabify a word. Recall that in Korean there may be a maximum of one C in coda position and one C or one C plus glide in onset position. If there is a sequence of consonants that violates this constraint, then one of the consonants must be deleted according to specific deletion rules in Korean. One such rule is the following, where $ refers to a syllable boundary.

(28) Deletion rule

/ps/ $ \rightarrow [p^*] /__$

Consider, now, the following Korean data. The root /ps/ means ‘not have’. What is the syllable structure for the phonetic forms? What happened to the /s/ in 냐다? (Note that there is fortition in some of the consonants, identified by an *. This does not play a role in the questions of syllabification here.)

(29) Derivations of 냐다 and 냐어 요.

<table>
<thead>
<tr>
<th>Underlying Representation</th>
<th>/aps-ta/</th>
<th>/aps-jo/</th>
</tr>
</thead>
<tbody>
<tr>
<td>not.have-DECL</td>
<td>not.have-INFORMAL-POLITE</td>
<td></td>
</tr>
</tbody>
</table>

| Surface Representation  | [ap’t*a] | [ap’s*jo] |

To answer this question, let us look at the underlying forms and attempt to syllabify them. First we identify the nuclei in the two forms.
Next, we form the onsets. Again, the rule “Maximize Onset Principle” states that as much material as possible should appear in the onset of the syllable. Note that in Korean only a single consonant can appear in onset position. This gives the following results, so far.

\[
\begin{array}{ccccccc}
N & O & N & N & O & N & O \\
\lambda & p & s & t^* & a & \lambda & p \\
\end{array}
\]

Finally, we form the codas. Since \([ps]\) is not an acceptable coda in Korean, the deletion rule in (28) applies.

\[
\begin{array}{ccccccc}
\sigma & \sigma & \sigma & \sigma & \sigma \\
R & O & R & O & R & O \\
\lambda & p & s & t^* & a & \lambda & j \\
\end{array}
\]

**Practice 4.7:** Syllabify the following Korean words.

- 사과 (サガポ) (apple)
- 먹어요 (メキヨヨ) (eating)
- 깨끗해졌다 (ケクチャキタ) (cleaned)
- 날아가요 (ナカバキヨ) (flying)

As a final illustration of how syllable structure can affect phonological rules and play a role in our understanding of the phonological forms, we will consider devoicing in German. Note that in the following data there is a great deal of variation in the shape of the plural marker; however, it plays no role in the discussion. Syllable boundaries are marked with a period. Consider the following data.

(30) **German Plural and Genitive Markers**

<table>
<thead>
<tr>
<th>singular</th>
<th>plural</th>
<th>genitive</th>
<th>translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ʃulp]</td>
<td>[ʃulp.ən]</td>
<td></td>
<td>‘cuttlebone’</td>
</tr>
<tr>
<td>[typ]</td>
<td>[ty.pən]</td>
<td>[typs]</td>
<td>‘type’</td>
</tr>
<tr>
<td>[lump]</td>
<td>[lump.ən]</td>
<td></td>
<td>‘scoundrel’</td>
</tr>
<tr>
<td>[hip]</td>
<td>[hi.ə]</td>
<td></td>
<td>‘blow’</td>
</tr>
<tr>
<td>[dip]</td>
<td>[di.ə]</td>
<td>[dips]</td>
<td>‘thief’</td>
</tr>
<tr>
<td>[zip]</td>
<td>[zi.ə]</td>
<td></td>
<td>‘sieve’</td>
</tr>
<tr>
<td>[flut]</td>
<td>[flu.ən]</td>
<td></td>
<td>‘flood’</td>
</tr>
<tr>
<td>[frlt]</td>
<td>[frl.ə]</td>
<td>[frlts]</td>
<td>‘field’</td>
</tr>
<tr>
<td>[ait]</td>
<td>[ai.ə]</td>
<td></td>
<td>‘oath’</td>
</tr>
</tbody>
</table>

Let’s consider the various forms for ‘thief’. It looks as though it is possible to account for the surface forms by assuming a rule of intervocalic voicing as we did for the Korean data above. However, assuming such a rule
is problematic for much of the other data. Consider the following putative derivations for ‘thief’ and ‘type’ based on a rule of intervocalic voicing.

<table>
<thead>
<tr>
<th>UR</th>
<th>/dip/</th>
<th>/dip-ə/</th>
<th>/typ/</th>
<th>/typ-ən/</th>
</tr>
</thead>
<tbody>
<tr>
<td>C → [+voi] / V __ V —</td>
<td>dibø —</td>
<td>tvbøn</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| SR       | [dip] | [dibø] | [typ] | [tvbøn] |

Such a rule gives rise to the wrong surface form for the plural of [typ], so we must look for an alternative explanation. Looking through the data we see that obstruents (all sounds that are [-sonorant]) are always voiceless in coda position. So instead let us explore the possibility that voiced obstruents become voiceless word-finally. Note we must restrict this rule to obstruents since /n/ and /ŋ/ appear in coda position, yet are still voiced. In order for this analysis to work, we must assume that the underlying form for ‘thief’ is /dib/. Here, then, are the derivations with the correct rule. The two genitive forms for ‘thief’ have been added to illustrate the analysis further.

<table>
<thead>
<tr>
<th>UR</th>
<th>/dib/</th>
<th>/dib-ə/</th>
<th>/typ/</th>
<th>/typ-ən/</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-son] → [-voi] / __ #</td>
<td>dip — — —</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| SR       | [dip] | [dibø] | [typ] | [tvbøn] |

These German data offer another illustration of neutralization (see the discussion on flapping in English above). Although /p/ and /b/ are distinct phonemes in German, this distinction is found only in onset position. In coda position these two sounds neutralize to [p]. Neutralizations in coda position are extremely common cross-linguistically. In Korean the sounds /t/, /th/, /tʰ/ , /ʃ/, /ʃʰ/, /ʃʰ/, /s/, and /sʰ/ are constrastive in onset position, but all neutralize to [tʰ] in coda position.

4.6 Major Phonological Rules of Korean

This section covers some phonological rules of Korean, concentrating on properties that may cause confusion for the Korean speaker learning English. Let’s start with nasalization. Consider the following Korean data carefully.

(31) Nasalization in Korean

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>있어</td>
<td>[is*ʌ]</td>
<td>있는</td>
<td>[innum]</td>
<td>받아</td>
</tr>
<tr>
<td>핏빛</td>
<td>[hɾt<em>p</em>it’]</td>
<td>핏무리</td>
<td>[hɾtu:nur]</td>
<td>앞으로</td>
</tr>
</tbody>
</table>
Let’s consider the underlying representation of ‘ten’. Let’s assume for now that it’s /ʃp/. When /p/ is in coda position in Korean, it becomes an unreleased stop [p']. Notice that for ‘ten metres’, however, /p/ becomes [m]. Likewise, the /p/ in ‘chew’ and ‘cold’ becomes [m] before /n/. We also see that underlying /s*/ /s/, and /t/ become [n] before /n/ or /m/. What happens to /k/ before /n/ or /m/? We could write three individual rules as follows.

(32)  

a. /k/ → [ŋ] / __ {m, n}  
b. /t/ → [n] / __ {m, n}  
c. /p/ → [m] / __ {m, n}

These rules all look very similar. It seems to be the same kind of change that happens in each rule. A consonant changes into a nasal consonant with the same place of articulation before another nasal consonant. We can rewrite these three rules with one simple rule.

(33)  

C → [+nasal] / __ [+nasal]

---

Nasalization in L2 English

The Kpop song Pick Me by 김창한 provides a good example of L1 interference. When the performers sing "pick me up" quickly, the /k/ in pick becomes an /ŋ/ because of the following /m/. Is it especially noticeable at 1:32 in the following video.

https://www.youtube.com/watch?v=BiorIyrjTHc

Consider now the following Korean data.

(34)  

Intervocalic Voicing in Korean

먹다 [maŋt*a]  
먹어 [maŋa]  
먹었다 [maŋat*a]  
어기 [agi]  

입다 [ip’t*a]  
입었어 [ipas*a]  
입 [ip']  
입이 [ibi]  

어디 [adi]  
이따 [it*a]  
담배 [tambe]  
읽어 [ipbi]  

받다 [pat’t*a]  
받아 [pada]  
앞으로 [apbaru]  
앞나 [anmi]  

거짓부령 [kadʒi’t*p*urang]  
거짓말 [kadʒinmal]  
국어 [kugJa]  
국민 [kuŋmin]  

예쁘다 [jep’uda]  
가까워 [kak*aupa]  
반 [pan]  
생각 [sengak*]

For this data we will look at [p] and [b]; [t] and [d]; and [k] and [g]. We construct our charts as we did above. We can ignore the difference between plain and stopped consonants, [p] versus [p’], since we know that consonants in coda position are predictably stopped.
Let’s examine [p] and [b] first. Checking the charts above we see that there is no pattern looking just at the left side or the right side. Rather, we must look at both sides to spot the pattern. We see that [b] appears when there is a voiced sound both on the left and on the right; otherwise, [p] appears. Observe further that /pʰ/ and /pʰ/ are not affected. So, we come up with the following distribution chart for the two allophones of /p/.

\[
\begin{align*}
&\text{[b]} / [+\text{voice}] _[+]\text{[voice]} \\
&\text{[p]} / \text{elsewhere}
\end{align*}
\]

In fact, if we look at the distributions of [t] and [d], as well as [k] and [g], we see the same pattern. Note that this rule affects plain consonants only [-CG, -SG]. Fortis and aspirated consonants are not affected. We can come up with the following rule to account for these changes.

\[
\begin{align*}
&\begin{bmatrix}
\text{-cont} \\
\text{-CG} \\
\text{-SG}
\end{bmatrix} \rightarrow [+\text{voice}] / [+\text{voice}] _[+]\text{[voice]}
\end{align*}
\]

**Palatalization:** s → S / _i (see page 228 for more examples of palatalization in Korean.)

**Intersonorant Voicing:**

\[
\begin{align*}
&\begin{bmatrix}
\text{-cont} \\
\text{-CG} \\
\text{-SG}
\end{bmatrix} \rightarrow [+\text{voice}] / [+\text{voice}] _[+]\text{[voice]}
\end{align*}
\]

**Nasalization:** C → [+nas] / _[+nas]

## 4.7 Major Phonological Rules of English

This section covers some of the common phonological rules of English, in addition to those covered previously. We will see that many of these rules depend on syllable structure. Consider the following data.

(36) Allophones of /l/ in English
Can you predict the distribution of [l] and [ı] (often called dark-l)? The distribution is predictable, however, it is not determined by neighbouring sounds, but rather by the position of this phoneme in the syllable. When /l/ is in onset position, it surfaces as [l]. When is appears in the rhyme (either in the nucleus or in the coda), it surfaces as [ı]. Thus, we can state the distribution of these two allophones informally as follows.

\[
\begin{array}{c}
/l/ \\
\arrow{\rightarrow} \\
\text{[l] / elsewhere} \\
\end{array}
\]

\[
\begin{array}{c}
/l/ \\
\text{[ı] / in nucleus or coda position (= rhyme)} \\
\end{array}
\]

More formally, we can state this as a phonological rule as follows. Note carefully how syllable structure is referred to. Specifically, that part of the syllable in question points to the underscore in the rule.

\[
\begin{array}{c}
/l/ \\
\text{→ [ı] /} \\
\end{array}
\]

Consider the following Canadian English data. This data represents a phenomenon known as the Canadian Raising. Canadian Raising is widespread throughout Canada and is also found in parts of the northern US.

- ice /aɪs/  eyes /aɪz/
- height /hɑɪt/  hide /hɑɪd/
- life /laɪf/  lie /laɪ/  live /laɪv/
- clout /klɔʊt/  cloud /klɔʊd/  cow /kɔʊð/
- house /hɔʊs/  houses /hɔʊzəz/
- pout /pɔʊt/  pow /pɔʊ/  

Examining this data closely we see that the diphthongs /aɪ/ and /aʊ/ become [ʌ] and [ʌʊ], respectively, before voiceless consonants. The head of the diphthong, /aɪ/, has the feature [+low] and the head of the raised diphthong, /ʌ/, has the feature [-low], so we can formulate the following rule to account for the Canadian Raising.

(37) \( V \rightarrow [-\text{low}] / \_\_\_V[-\text{voi}] \)
Next, we look at aspiration. Consider the following data. Try to determine if the distribution of [t] and [tʰ] is predictable or not. Do the same for /p/ and /k/.

<table>
<thead>
<tr>
<th>Word</th>
<th>Pronunciation</th>
<th>Word</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>atom</td>
<td>[æ.ˈrəm]</td>
<td>atomic</td>
<td>[ə.ˈtʰə.mək]</td>
</tr>
<tr>
<td>pot</td>
<td>[pʰət]</td>
<td>spot</td>
<td>[spət]</td>
</tr>
<tr>
<td>top</td>
<td>[tʰəp]</td>
<td>stop</td>
<td>[stʰəp]</td>
</tr>
<tr>
<td>repopulate</td>
<td>[rə.pʰə.pjuˈleɪt]</td>
<td>Gestapo</td>
<td>[ˈgə.ʃta.pəʊ]</td>
</tr>
<tr>
<td>ferment</td>
<td>[ˈfɜrnənt]</td>
<td>fermentation</td>
<td>[ˈfɜrnənt.ˈtʰə.ʃən]</td>
</tr>
<tr>
<td>accomplish</td>
<td>[ə.ˈkʰəm.plɪʃ]</td>
<td>talked</td>
<td>[ˈtʰəkt]</td>
</tr>
</tbody>
</table>

We see that voiceless stops become aspirated when they appear at the left edge of a stressed syllable. Crucially, observe that in the words ‘spot’, ‘stop’, and ‘Gestapo’ the /t/ is not aspirated. For convenience, we can notate a stressed syllable as $\hat{\sigma}$.

\[
\begin{bmatrix}
-\text{son} \\
-\text{voi} \\
-\text{cont}
\end{bmatrix} \rightarrow [+SG] / \underline{\$}
\]

**4.8 Morphophonology**

Until now, we have just been looking at the properties of sound systems of language – phonology. We know, however, that phonology interacts with morphology in many ways. The interaction between phonology and morphology is called **morphophonology**. For instance, certain kinds of allomorphic variation is triggered by
phonological properties. We call this **phonologically conditioned allomorphy**. Consider the following Korean data.

<table>
<thead>
<tr>
<th>base noun</th>
<th>nominative form</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>[sagwa]</td>
<td>[sagwaga]</td>
<td>‘apple’</td>
</tr>
<tr>
<td>[kogi]</td>
<td>[kogiga]</td>
<td>‘meat’</td>
</tr>
<tr>
<td>[minho]</td>
<td>[minhoga]</td>
<td>‘Minho’ (a name)</td>
</tr>
<tr>
<td>[mul]</td>
<td>[muri]</td>
<td>‘water’</td>
</tr>
<tr>
<td>[pap^]</td>
<td>[pabi]</td>
<td>‘cooked rice’</td>
</tr>
<tr>
<td>[hak’s*ra]</td>
<td>[hak’s*raj]</td>
<td>‘student’</td>
</tr>
</tbody>
</table>

There are two allomorphs for the nominative marker in Korean: [-i] and [-ga]. The choice between these two is triggered by the phonological properties of the base to which it attaches. We can state the insertion frames for the lexical entries as follows.

(38) Insertion frames for the Korean Nominative Case Suffix
   a. \( \text{NOM} \leftrightarrow [-i]/V \) __
   b. \( \text{NOM} \leftrightarrow [-ga]/C \) __

Likewise, the conditioning environment for the allomorph in question may be lexically specified. Recall the German data in (30) above. Observe that there are three plural allomorphs represented in these data. (In reality, there are more plural allomorphs, but we will restrict ourselves to the three in these data.) The choice of allomorphs depends on the root it attaches to, so we call this **lexically-conditioned allomorphy**. Here are the insertion frames for the plural allomorphs in the German data.

(39) Insertion frames for the German Plural Suffix
   a. \( \text{PL} \leftrightarrow [-@n]/\{\text{ulp, t yp, lump, flut}\} \) __
   b. \( \text{PL} \leftrightarrow [-@]/\{\text{hib, dib, zib, aid}\} \) __
   c. \( \text{PL} \leftrightarrow [-@K]/\{\text{feld, kind}\} \) __

Lexically-conditioned allomorphy is not necessarily sensitive just to roots but also to semantic features of roots. Recall that the choice of classifier in Korean and in Chinese languages is conditioned by various semantic features.
Practice 4.8: Consider the following (slightly simplified) Tarascan data (Foster, 1969). The morpheme /te/ indicates volitional movement and has two allomorphs. Determine the distribution of these two allomorphs. Is this an example of phonologically-conditioned allomorphy or lexically-conditioned allomorphy?

[jo-tsi-ta\sp\^h eni] ‘to put a great deal on someone’s table’
[ni-tsi-kwareni] ‘to go all alone’
[xu-tsi-kwareni] ‘to come all alone’
[anta-tse-nuni] ‘to be born’
[kura-tse-ni] ‘to be ashamed’
[wanta-tse-ni] ‘to talk about another’

Sometimes, however, the actual rules of phonology interact with the rules of morphology. Consider the following European Portuguese data, paying close attention to the plural forms of the masculine definite determiners [uf], [u\^], and [uz].

(40) Portuguese Morphophonological Alternations

[u pra\^u] ‘the plate’
[u kw\art\^u] ‘the bedroom’
[u \tet\^u] ‘the ceiling’
[u k\ant\^u] ‘the song’
[u \^ivr\^u] ‘the book’
[u ri\u] ‘the river’
[u bar\ku] ‘the boat’
[u m\^\nu\^r\^u] ‘the piece of furniture’
[u a\^u] ‘the eye’
[u a\^\^u] ‘the garlic’
[u i\u] ‘the anthem’

[u\^] ‘the plates’
[u kw\art\^u] ‘the bedroom’
[u \tet\^u] ‘the ceilings’
[u\^ k\ant\^u] ‘the songs’
[u\^ \^ivr\^u] ‘the books’
[u\^ ri\u] ‘the rivers’
[u\^ bar\ku] ‘the boats’
[u\^ m\^\nu\^r\^u] ‘the pieces of furniture’
[u\^ a\^u] ‘the eyes’
[u\^ a\^\^u] ‘the garlic (plants)’
[u\^ i\u] ‘the anthems’

We see here that there are three forms with the following distribution.

(41) Distribution of Portuguese masculine plural definite determiner

/uf/ /uz/ /uf/

[u\^] / __ (+cons) __ (+voi)

[+cons] __ (+voi) __ (-cons) __ (+cons) __ (-voi)

It should be clear by now that should try to discern a natural pattern from this distribution. Note that the consonants in the first two allomorphs are voiced ([\^] and [z]). Note further that the triggering environment in
both cases contains the feature [+voi]. (Vowels are always voiced unless specified otherwise.) Thus, we can naturally state a voicing assimilation rule as follows.

(42)  /f/ → [+voi] / _[+voi]

We still have to account for the [z] in the second allomorph above, so we need one more rule. Here, we must change the [z] into a [z]. The only difference between these two sounds is the value of the feature [±anterior]. So we come up with the following rule.

(43)  /z/ → [+ant] / _[-cons]

Now, the astute reader should have three questions at this point. (1) How did we determine that /uf/ is the basic allomorph? (2) The phonological rules, as stated, apply across the board, not just to the morpheme /uf/. That is, it predicts that /f/ should always change to [z] when it appears before a vowel—not just in the morpheme /uf/. Is the prediction borne out or falsified? (3) Are the rules ordered? Let’s answer these questions in order.

Choosing other forms as the basic morpheme yields the following results.

In both cases, rules of equal complexity can be formulated, which doesn’t help settle the matter. However, the form /uf/ when uttered in isolation surfaces as [uf]. Also, when this morpheme is used as a pronoun (meaning ‘them’), it can often appear at the end of a phrase and also surfaces as [uf]. So, [uf] appear both before voiceless consonants and in phrase-final position (where ¯ marks a phrase boundary). Thus, the conditioning environment for [uf] is shown in (44). Observe that it does not form a natural class, and so must be stated as the elsewhere case, giving rise to the distribution in (45).
The second question requires us to consider additional data. There are several other morphemes with the same distribution as /uʃ/. Here are some examples to show that the phonological rules generalize to other morphemes.

(46) Additional Portuguese data

[ʊʃ praʧuf] ‘some plates’   [ʊʃ jɪvruʃ] ‘some books’   [ʊz əʌuf] ‘some eyes’
[poʊkʊf praʧuf] ‘few plates’   [poʊkʊʒ] ‘few books’   [poʊkuz əʌuf] ‘few eyes’
[ɛlɡɨf praʧuf] ‘some plates’   [ɛlɡɨʒ jɪvruʃ] ‘some books’   [ɛlɡɨz əʌuf] ‘some eyes’
[mʊiʧuf praʧuf] ‘many plates’   [mʊiʧuʒ jɪvruʃ] ‘many books’   [mʊiʧuz əʌuf] ‘many eyes’

However, the rules as stated also suggest that they hold inside words and inside morphemes. Consider the following data.

(47)  [ʃuˈtu] ‘kick’; [ʃadɾef] ‘chess’; [ʃæti] ‘people’

Thus, these rules do not hold inside words. Note, finally that in the data in (40) and (46) that the rules we proposed hold between words or between morphemes. That is because the boundary between the two elements in question is both a word boundary and a morpheme boundary.

(48)  [ʊʃ bɑɾkuʃ]  
the.M.PL boat-PL  
‘the boats’

[ʊʃ bɑɾkuʃ] ‘the boats’
#ʊʃ#bɑɾkuʃ# word boundaries
+ʊʃ+bɑɾku+ʃ+ morpheme boundaries

So, although we are certain that our phonological rules do not operate across the board (since they do not
affect the data in (47), we cannot tell based on the data so far whether they operate at word boundaries only or morpheme boundaries, too. The following data show clearly that they operate at morpheme boundaries. The phones in question are highlighted in red.

(49) Portuguese data with prefixes

[\[w\]] ‘grandfather’  [\[biz+\]] ‘great-grandfather’
[\[bi\] ‘great-grandson’
[\[\]] ‘post-modern’
[\[\]] ‘post-syntactic’
[\[\]] ‘transport’
[\[\]] ‘beyond the Amazon

We restate the rules as follows.

(50) Morphophonological rules for Portuguese

a. /\[\]/ \rightarrow [+voi] / + [+voi]
b. /\[\]/ \rightarrow [+ant] / + [-cons]

Finally, we consider whether the rules need to be ordered or not. Observe that /\[\]/ is the output of the first rule. In order to get the correct output, the two rules must apply in the order shown. To show this, let’s consider the reverse order first.

Underlying Representation /uf pra\[\]/ /uf livru\[\]/ /uf akuf/
/\[\]/ \rightarrow [+ant] / + [-cons] – – –
/\[\]/ \rightarrow [+voi] / + [+voi] – u\[\] livru\[\] u\[\] akuf
Surface Representation [uf pra\[\]] [u\[\] livru\[\]] [u\[\] akuf] !!!!

Observe that we get the wrong result with the rules as stated in the order above. If we reverse the order of the rules, the we get the right result.

Underlying Representation /uf pra\[\]/ /uf livru\[\]/ /uf akuf/
/\[\]/ \rightarrow [+voi] / + [+voi] – u\[\] livru\[\] u\[\] akuf
/\[\]/ \rightarrow [+ant] / + [-cons] – – u\[\] akuf
Surface Representation [uf pra\[\]] [u\[\] livru\[\]] [uz akuf]

As a final illustration of the interaction between morphology and phonology, consider the following Plains Cree data (Ahenakew 1987). There are two lexically-conditioned allomorphs for the plural morpheme. /-a/ is the plural marker for inanimate nouns and /-ak/ is the plural marker for animate nouns. Recall the discussion of noun classes based on animacy from page 41. Our task is to determine the phonological forms of the roots and to determine what noun class they belong to (animate or inanimate). There is also a locative morpheme in these data (a morpheme meaning ‘in’ or ‘on’). What is the phonological form of this morpheme? Does it have any allomorphs? Finally, we need to determine what phonological rules are necessary to account for these data.
Looking at the first four lines, we see that the addition of the plural suffix does not present any difficulties. The words for ‘child’ and ‘sock’ take the plural morpheme -ak, and the words ‘shoe’ and ‘town’ take the plural morpheme -a. The following five words have an additional complication. Let’s compare ‘sock’ and ‘dog’.

- ‘socks’ asikan  -ak
- ‘dogs’ atim w  -ak

The plural form for ‘dog’ has an unexplained /w/, as do the other four words from ‘town’ to ‘tanned hide’. There are three possibilities to consider.

1. /w/ is inserted by some phonological rule
2. /w/ is deleted by some phonological rule
3. There are really four plural allomorphs, and the /w/ forms are lexically specified.

The third option essentially says we have unpredictable, irregular forms. In an of itself, this is not impossible. We have seen irregular plurals in English and in German. However, we should try to look for a systematic explanation first before giving up. The /w/-insertion rule should be triggered by the sound to the left of the insertion site. Looking at these sounds, however, there doesn’t seem to be any pattern.

Let’s consider, then, a /w/-deletion rule. If we assume the underlying representation of ‘dog’ in Cree is /atimw/, then we can propose a /w/-deletion rule as follows.
Delete /w/ at the end of a word if it is preceded by a consonant. Note we must specify that the /w/ is preceded by a consonant because of the word for ‘town’. This gives us the following underlying representations so far.

(53) Underlying Representations for the Cree forms in (51)

| /awaːsis/  | ‘child’     | /-ak/  | animate plural |
| /maskisin/ | ‘shoe’      | /-a/   | inanimate plural |
| /asikan/   | ‘sock’      |        |                 |
| /oteːnaw/  | ‘town’      |        |                 |
| /mistikw/  | ‘tree’      |        |                 |
| /atimw/    | ‘dog’       |        |                 |
| /waːposw/  | ‘rabbit’    |        |                 |
| /pahkeːkinw/ | ‘tanned hide’ | | |

Finally, we consider the locative morpheme. We observe that the form [-ohk] appears after roots ending in /w/. The form [-ihk] appears elsewhere. We could argue that this is phonologically-conditioned allomorphy followed by /w/-deletion, or we could propose a vowel coalescence rule. Since the second option is simpler, we take that option and propose that the underlying representation of the locative morpheme is /-ihk/. We also propose the following rule.

(54) /wi/ → [o]

This gives us the complete list of underlying representations for the Cree forms in (51).
(55) Final list of Underlying Representations for the Cree forms in (51)

| /awāsis/  | ‘child’ | /-ak/ | animate plural |
| /maskisin/ | ‘shoe’ | /-a/ | inanimate plural |
| /asikan/ | ‘sock’ |   |   |
| /o:te:naw/ | ‘town’ | /-ihk/ | locative |
| /mistikw/ | ‘tree’ |   |   |
| /atimw/ | ‘dog’ |   |   |
| /waːposw/ | ‘rabbit’ |   |   |
| /pahkeːkinw/ | ‘tanned hide’ |   |   |

We end by running through the derivations of ‘tree’, ‘trees’, and ‘in a tree’. Note that since these rules do not interact, we do not have to worry about rule ordering.

<table>
<thead>
<tr>
<th>Underlying Representation</th>
<th>/mistikw/</th>
<th>/mistikw-ak/</th>
<th>/mistikw-ihk/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/w/ → Ø / C_#</td>
<td>mistik</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>/wi/ → [o]</td>
<td>–</td>
<td>–</td>
<td>mistikohk</td>
</tr>
</tbody>
</table>

| Surface Representation | [mistik] | [mistikwak] | [mistikohk] |
Answers to Practice Questions

Practice Question 1.
The following vowel phonemes are found: a, e, i, o, u
The following consonant phonemes are found: p, t, k, s, h, ɬ, m, n, r

Practice Question 2.
1. [bat], [bit], [tap], [apt], [plit], [apro]
2. [psi], [pta], [ksi], [ti], [tik], [pto]
3. [ap], [af], [asp], [astf], [sastf], [ostfa]
4. [ita], [tat], [fota], [ona], [betana], [fitat]

Practice Question 3.

\[ \text{[d]} / \_ \# \]

\[ /r/ \]

\[ [r] / \text{elsewhere} \]

Practice Question 4.
English vowels: 1. [-back], 2. [+hi], 3. [+back]
English consonants: 1. [+labial], 2. [+coronal, -anterior], 3. [-sonorant, +voiced]
Korean vowels: 1. [+hi], 2. [-back], 3. [-hi]
Korean consonants: 1. [+SG], 2. [+sonorant, -nasal], 3. [+coronal]

Practice Question 5.

\[ \begin{array}{c}
\text{-son} \\
\text{-SG} \\
\text{-cont}
\end{array} \rightarrow [+\text{voi}] / [+\text{nas}]_{\text{—}} \]

Practice Question 6.
Practice Question 7.

Practice Question 8.
Exercises

Question 1. What feature or features capture the following natural classes in English? The vowels in brackets indicate that only some dialects of English have that vowel. For those dialects, the vowels in question are part of the natural class.

1. p, b, t, d, k, g
2. p, b, f, v, w, m
3. p, b, m, t, d, n
4. v, ð, z, ð, i, l, j, w
5. t, θ, s, f, ʃ
6. k, g, ʝ
7. b, d, g, v, ð, z, ð, ʃ
d3
8. i, i, e, r, ə
9. i, i, u, o
10. i, e, u, u, (ə)
11. u, u, o, (ə), (ŋ)

Question 2. What feature or features capture the following natural classes in Korean?

1. t, tʰ, t*, n, r, s, s*
2. m, n, η, r, w, j, u
3. s, s*, h
4. i, u, u
5. u, a, a
6. o, r, a, a

Question 3. State the UR of the following English words and give the derivations necessary to derive the SR: pen, wink, thimble, powder, pouting

Question 4. State the UR of the following Korean words and give the derivations necessary to derive the SR: 먹다 (mek-ta ’to eat’), 먹는다 (mek-nun-ta ’eats’), 닭이 (talk-i, ’chicken.NOM’)
**Question 5.** Transcribe and syllabify the following English words: textured, impossible, travelled, importance

**Question 6.** Transcribe and syllabify the following Korean words: 맞있어요, 없습니다, 생각해요, 철각형, 목적어

**Question 7.** Consider the following Ajagbe data. Are [s] and [ʃ] allophones of the same phoneme? If they are allophones of the same phoneme state the distribution. If they are separate phonemes give appropriate evidence.

<table>
<thead>
<tr>
<th>pronunciation</th>
<th>English</th>
<th>pronunciation</th>
<th>English</th>
<th>pronunciation</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ahiʃi]</td>
<td>‘dishonesty’</td>
<td>[ezɔ]</td>
<td>‘fire’</td>
<td>[aʃi]</td>
<td>‘woman’</td>
</tr>
<tr>
<td>[zɔn]</td>
<td>‘walk’</td>
<td>[zeqeka]</td>
<td>‘directly’</td>
<td>[ʒinkpin]</td>
<td>‘chair’</td>
</tr>
</tbody>
</table>

**Question 8.** Consider the following Nepali data. Are p, pʰ and ʃ allophones of the same phoneme or separate phonemes? If they are separate phonemes state your evidence. If they are allophones of the same phoneme state their distribution.

<table>
<thead>
<tr>
<th>pronunciation</th>
<th>English</th>
<th>pronunciation</th>
<th>English</th>
<th>pronunciation</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>[səfa]</td>
<td>‘clean’</td>
<td>[поі]</td>
<td>‘husband’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[pʰas]</td>
<td>‘slow’</td>
<td>[kapі]</td>
<td>‘monkey’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[kaʃi]</td>
<td>‘coffee’</td>
<td>[пал]</td>
<td>‘moment’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[pʰul]</td>
<td>‘flower’</td>
<td>[lапʰ]</td>
<td>‘jump’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[qəpʰla]</td>
<td>‘a band’</td>
<td>[saфal]</td>
<td>‘successful’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[pʰilmi]</td>
<td>‘related to a film’</td>
<td>[афу]</td>
<td>‘self’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[pʰal]</td>
<td>‘fruit’</td>
<td>[pʰіr]</td>
<td>‘Turn on!’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Question 9.** Consider the following Ilocano data. State the distribution of [i] and [ɪ]. Write a phonological rule to capture this alternation. Stress is marked by the symbol ‘ to the left of the syllable. Syllable boundaries are indicated by a period.

<table>
<thead>
<tr>
<th>pronunciation</th>
<th>English</th>
<th>pronunciation</th>
<th>English</th>
<th>pronunciation</th>
<th>English</th>
</tr>
</thead>
</table>

**Question 10.** Consider the following Hungarian data. Observe that consonants sometimes become voiced. State the rule for consonant voicing.
Consider the following Onondaga data. State the distribution of the sounds [s], [dʒ], and [ʃ].

[soːw] ‘duck’
[soːw] ‘duck’
[soːw] ‘duck’
[soːw] ‘duck’

Question 12. Consider the following Oneida data (an Iroquoian language spoken in central North America on the border between Canada and the US). What are the two allomorphs of the 3rd person singular masculine prefix? What is their distribution? Use phonological features if necessary.

[ɑː-ha-nohale-ʔ] [ɑː-ha-tsali-ʔ]
FUT-3SG.M-wash-PERF FUT-3SG.M-find-PERF
‘He will wash it.’ ‘He will find it.’

[wa-ha-nohale-ʔ] [je-ha-atawjaʔ-haʔ]
FACT-3SG.M-wash-PERF TRANSLOC-3SG.M-enter-IMPF
‘he washed it.’ ‘he goes in.’

[la-hjatu-heʔ] [la-nohale-heʔ]
3SG.M-write-IMPF 3SG.M-wash-IMPF
‘he is writing.’ ‘he is washing it.’

Consider now the following data. What are the two allomorphs of the factual morpheme (FACT)? What is their distribution? Use phonological features if necessary.
<table>
<thead>
<tr>
<th>Verb Form</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>wa-ha-nohale-?</td>
<td>3SG.M-wash-PERF</td>
</tr>
<tr>
<td>wa?-k-nohale-?</td>
<td>1SG-wash-PERF</td>
</tr>
<tr>
<td>wa?-t-ha-ashto-?</td>
<td>3SG.M-cry-PERF</td>
</tr>
<tr>
<td>wa-s-nohale-?</td>
<td>2SG-wash-PERF</td>
</tr>
<tr>
<td>wa?-tni-nohale-?</td>
<td>1DU.INCL-wash-PERF</td>
</tr>
<tr>
<td>wa?-jakni-nohale-?</td>
<td>1DU.EXCL-wash-PERF</td>
</tr>
</tbody>
</table>
Further Reading

- [Kennedy (2016)](#) - a workbook containing phonology problems and discussions
- [Carr (2012)](#) - an introduction to English phonetics and phonology
- [Carr (2015)](#) - an often used textbook for undergraduate phonology courses
- [Shin et al. (2012)](#) - an overview of Korean phonetics and phonology
Chapter 5

Syntax

By the end of this chapter you should:

- be familiar with the following structural relations of syntax trees: dominance, sisterhood, immediate dominance, exhaustive dominance, c-command
- be able to explain why the underlying structure of language is hierarchical rather than flat
- understand the difference between SVO and SOV order and how to capture it
- understand basic concepts of constituency and be able to execute various tests for constituency
- understand the basic structure of NP, VP, AdvP, AdjP and PP
- be able to draw trees for multiclausal structures
- understand some basic transformations, including passivization, wh-movement, and scrambling.

5.1 Introduction

In the previous chapters we have investigated word structure and sound structure in a number of languages around the world. We have seen that the same set of underlying principles can be used to investigate, describe and analyse phonology and morphology. Another universal property of language is that words are organized into strings called sentences. Syntax is the study of sentence structure, which is the topic of this chapter. We will see how words are put together into larger and larger phrases, and then into sentences. We will also look at differences in word order in various languages. Finally, we will consider how phrases move in sentences, for instance in passives and in questions. Word order is clearly important for language. The study of syntax dates back to the days of Pāṇinī, who published a set of 3, 959 rules of Sanskrit morphology and syntax. Pāṇinī’s grammar is still considered an elegant piece of scholarship for its clarity and logical structure.

Consider the following pair of sentences.
These two sentences contain exactly the same set of words, but in a different order. These two different orders have two different meanings, thus illustrating the importance of word order. Like the morphological structures we saw before, syntactic structures are also organized into hierarchies called phrase structure.

We will start our discussion of syntax with a few important properties of phrase structure.

### 5.1.1 Basic Properties of Phrase Structure

Although syntactic inquiry in the 20th century has generated tremendous insight into the hierarchical organization of language, the idea that language is organized hierarchically, rather than as a linear string of words can be traced back to at least Aristotle and Pāṇini. In this chapter, we will motivate the structures that we assume for sentences so that we can understand why we draw trees the way we do.

Before we begin, let’s introduce some basic terminology. The structures and terminology here have been common since Chomsky (1965) and have been developed in Stowell (1981). Consider the following syntactic tree.
is marked by the line starting from the bottom of X and going down to the top of A. X **exhaustively dominates** A, B, D and E. (We don’t need to include C, since D and E are contained in C, although it would not be incorrect to say the X exhaustively dominates A, B, C, D and E.) X **immediately dominates** A, B and C, but X does not immediately dominate D and E – it only dominates them. The relation of immediate domination holds between two elements only if there are no elements intervening between the two elements. A, B and C are **sisters**, as are D and E. C is the **mother** of D. D is the **daughter** of C. This structure also indicates **hierarchical** relationships among the elements. For instance, it shows that D and E behave as a unit that is called C here. It also shows that A, B, D, and E behave as an even larger unite, which is labelled X here. A string of words that behave as a unit is called a **constituent**. In our hypothetical tree above, D and E form a constituent labelled C. Here are formal definitions of these relations. We assume the relation of dominance is an axiom. Although these definitions may be tricky to follow at first pass, you should try to work through the logic to convince yourself that they work. (Note: iff = if and only if.)

\[(4) \]

a. **α immediately dominates** β iff α dominates β and there is no γ such that α dominates γ and γ dominates β.

b. **α exhaustively dominates** a set of terminals A iff for all β, β ∈ A, α dominates β and there is no γ such that α dominates γ and γ /∈ A.

Constituency is an important concept in any theory of syntax. So, what exactly do we mean by ‘constituent’? A constituent is an isolable syntactic unit, which behaves as an independent entity [Harris, 1970]. Before we consider some actual data, let’s be sure how we will represent the notion of constituency in terms of phrase structure. In the syntactic trees we are considering, a constituent is all and only the material dominated by a single node. Recall that a node, X, exhaustively dominates a string, xyz, if xyz are all dominated by X and X dominates nothing else.

**Constituency:** To represent a string xyz as a constituent it is exhaustively dominated by a single node.

**Practice 5.1:** Consider the following tree:

```
      A
     /\  
    B   C
   /\     
  D   E   c
 /\     |
 F G H I J
 /\     |
 f g h i j
```

Which of the following strings are constituents? fg, hij, ijc, fhg, fghij, ghij, hi, f
These are most of the basic structural relations. For now, let’s continue by considering the following sentence. We will ask ourselves how we can understand the hierarchical structures contained within it.

(5) The smart student is writing a tough exam.

We’ll start our analysis of this sentence by assuming it consists of a linear string of words, not grouped together into phrases of any kind. We can represent this analysis as follows. We can consider this our null hypothesis since it has the least amount of structure.

(6) \[
\text{S} \quad \text{Det} \quad \text{Adj} \quad \text{N} \quad \text{Aux} \quad \text{V} \quad \text{Det} \quad \text{Adj} \quad \text{N} \\
\text{the} \quad \text{smart} \quad \text{student} \quad \text{is} \quad \text{writing} \quad \text{a} \quad \text{tough} \quad \text{exam}
\]

This analysis shows the linear order of the lexical items, but nothing more. According to this tree, there are no constituents inside the sentence, although it indicates that the entire sentence is a constituent labelled S (for sentence). Just thinking about the sentence intuitively for a moment, we suspect that the noun phrases the bright student and a tough exam are constituents. The tests for constituency will be presented in the next section; however, pre-empting this discussion, we note that the bright student can be replaced by a pronoun such as she, and a tough exam can be replaced by another pronoun, it. This behaviour shows us that we should represent these two noun phrases as units in our tree. So we revise the tree as follows.

(7) \[
\text{S} \quad \text{NP} \quad \text{Aux} \quad \text{V} \quad \text{NP} \\
\text{Det} \quad \text{Adj} \quad \text{N} \quad \text{is} \quad \text{writing} \quad \text{Det} \quad \text{Adj} \quad \text{N} \\
\text{the} \quad \text{smart} \quad \text{student} \quad \text{a} \quad \text{tough} \quad \text{exam}
\]

The node, NP, stands for ‘noun phrase’ and dominates the two strings of words the smart student and a tough exam, respectively. Pronouns are common proforms in many of the world’s languages, and identify NP constituents. Consider the following Mandarin examples. The red pronoun in (8-b) replaces the string of words in red in the (8-a) example, showing that these strings of words behave as a constituent.

(8) a. zhè ge gāodà de xuéshēng qù dàxué. [Mandarin]  
   this LNK tall DE student go university  
   ‘This tall student is going to university.’

b. tā qù dàxué.  
   s/he go university  
   ‘S/he is going to university.’
Practice 5.2:
Re-write the following sentences with a pronoun to illustrate that the underlined strings of words are constituents.

The man in the moon ate some cheese.
Mary talked to the student who is eating an apple about the test.
Fred saw someone eating the apples that Mary bought.

Practice 5.3:
Consider the following French data. Assume that the object pronouns *le* (‘him’), *la* (‘her’), and *les* (‘them’) pick out nominal constituents as we saw for English and for Mandarin. Note that the object pronouns are homophonous with the definite determiner in French. This does not play a role in this practice set. Given the data below, identify the nominal constituents in the a examples. For all three examples assume that the b examples are accurate paraphrases of the a examples.

(9) a. Marie veut manger les pommes rouges demain.
   Marie wants to eat the apples red tomorrow
   ‘Marie wants to eat the red apples tomorrow.’

   b. Marie veut les manger demain.
   Marie wants them to eat tomorrow
   ‘Marie wants to eat them tomorrow.’

(10) a. Marie a bu le vin d’Italie avec Gianni.
    Marie has drunk the wine from Italy with Gianni
    ‘Marie drank the wine from Italy with Gianni.’

    b. Marie l’a bu avec Gianni
    Marie it has drunk with Gianni
    ‘Marie drank it with Gianni.’

    Marie has studied the radioactivity of isotopes of polonium in France.
    ‘Marie studied the radioactivity of isotopes of polonium in France.’

    b. Marie l’a étudiée en France
    Marie it has studied in France
    ‘Marie studied it in France.’

Looking at the tree in [7] we see that the subject NP and the object NP are sisters. This structure suggests that the subject and the object are somehow equivalent in hierarchical terms. There are many subject-object asymmetries found in natural language, however, suggesting that the subject and object are not hierarchically equivalent. Subject/object asymmetries have been investigated as far back as Aristotle. Let’s consider some of these asymmetries.

In the sentence *John and Mary like each other* the reference of the object *each other* is determined by the subject *John and Mary*. Thus, when the object is a reciprocal pronoun, it can take its reference from the
subject. Consider, now the ungrammatical sentence *Each other like John and Mary. We see that the opposite scenario does not hold. Namely, when the subject is a reciprocal pronoun, it cannot take its reference from the object. This is our first subject-object asymmetry. That is, a subject can determine the reference of an object pronoun, but an object cannot determine the reference of a subject pronoun.

The next asymmetry involves how the role of the subject is understood in the sentence. In the following examples, the role of the subject changes depending on the object. In the sentences on the left, the subject is understood as in control of the event. In the sentences on the right, the subject is understood as not being in control. Thus, the verb and the object together determine the role of the subject. Languages don’t seem to group the subject and the verb together to determine the role of the object.

    b. Mary threw a ball. Mary threw a fit.

In example (12-a), John is in control of his actions in the first sentence but not in the second. Likewise, in example (12-b) Mary is in control of her actions in the first sentence but not in the second. There are no examples, however, where the role of object depends on verb and subject together. Thus, the verb and the object function together as a unit to the exclusion of the subject (Marantz, 1984).

The final subject-object asymmetry we will look at concerns idioms. An idiom is a phrase whose meaning cannot be determined from its parts. Consider the following example.

(13) John decided to hit the books because he wanted an A+ on his exam.

Here, the string of words hit the books is not understood literally. Rather, it means to ‘study hard’. There are many idioms are composed of a verb and an object together. Table 5.1 lists several idioms from different languages.

There are no idioms, however, that are composed of a subject and verb together to the exclusion of the object (Marantz, 1997). Be careful not to become confused by idioms that are complete sentences such as The cat’s out of the bag or The horse has left the barn. Although these idioms include the subject and the verb, they do not do so to the exclusion of the object. Idioms like The cat’s out of the bag tell us that the whole sentence (S) is a constituent.

There are many other subject-object asymmetries. Some we will see in later discussions and others you will see if you undertake further studies of syntax. Taken together, these asymmetries suggest that the verb and the object form a constituent to exclusion of the subject.
<table>
<thead>
<tr>
<th>Language</th>
<th>Idiom</th>
<th>Literal Meaning</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>kick the bucket</td>
<td>die</td>
<td>have a conversation</td>
</tr>
<tr>
<td></td>
<td>chew the fat</td>
<td>have a conversation</td>
<td>tell a secret</td>
</tr>
<tr>
<td></td>
<td>spill the beans</td>
<td>work late into the night.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>burn the midnight oil</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portuguese</td>
<td>lavar a égua</td>
<td>to wash the mare</td>
<td>to finally do something after waiting a long time</td>
</tr>
<tr>
<td>Korean</td>
<td>sichimi tteyta</td>
<td>to take off the tail-feather</td>
<td>to play dumb/feign ignorance</td>
</tr>
<tr>
<td>Northern Sotho</td>
<td>go bea pelo</td>
<td>to place a heart</td>
<td>to satisfy someone</td>
</tr>
<tr>
<td>Cantonese</td>
<td>doei3 ngau4 taan4 kam4</td>
<td>to play the lute for the cow</td>
<td>say something people can’t understand or appreciate</td>
</tr>
<tr>
<td></td>
<td>caat3 haai6</td>
<td>to polish shoes</td>
<td>to flatter someone</td>
</tr>
<tr>
<td>Persian</td>
<td>pol-a-ro shekastan</td>
<td>to break bridges</td>
<td>to break off ties/to burn bridges</td>
</tr>
<tr>
<td>Mohawk</td>
<td>tusayuthaahkwe’</td>
<td>someone picked up the road</td>
<td>someone led the way</td>
</tr>
</tbody>
</table>

Table 5.1: Idioms from different languages

(14) 

```
(14) S
    | NP | Aux | VP |
    | The smart student | is | V | NP |
    | writing | a tough exam |
```

We say that the verb is the **head** of the VP and takes the object NP as a **complement**, terms borrowed from traditional grammar. The head of a phrase is the word that determines the category of the phrase. Thus, the head of a VP is the V. In the example above, the transitive verb *write* requires a direct object. We can say that the verb *write* takes an NP as a complement. We will explore complements of other heads below.

Let’s look at another example now. Consider the following Korean sentence. For now, the abbreviations in small caps aren’t important here, but you can look them up in the list of abbreviations.

(15) `Minswu-ka sakwa-lul mek-ess-ta
    Minsoo-NOM apple-ACC eat-PST-DECL
    ‘Minsoo ate an apple.’`

We notice one striking difference right away—the word order is not the same between English and Korean. The basic order of the English sentence is subject-verb-object, while the Korean sentence is subject-object-verb. We say that English is an SVO language, while Korean is an SOV language. This is a remarkably consistent word order pattern in both English and in Korean, and we capture this property by using a left-headed VP in
English and a right-headed VP in Korean. In a left-headed phrase, the head is to the left of its complement. In a right-headed phrase, the head is to the right of its complement. Here is the tree for the Korean sentence.

(16)

We will now turn to some specific diagnostics or tests for constituency. In particular, we will adduce more evidence for the VP node.

### 5.2 Constituency Tests

Recall example (14) from above. In this tree, the entire sentence is a constituent, and so are the NP *the bright student* and the VP *writing a tough exam*. Although we will cover the concepts more thoroughly in the next section, let’s take a brief preview of some of the properties of the NP and VP that would lead us to conclude that they are indeed constituents. First, we intuitively feel the NP *the bright student* is a constituent, since it is a participant in the sentence. When asked who is writing a tough exam, an appropriate response would be, "the bright student". Also, this NP can be replaced by a pronoun: *She is writing a tough exam*. As for the VP, it might feel less intuitive that it forms a constituent; however, when asked the question what the bright student is doing an appropriate response would be, "writing a tough exam". These preliminary observations lead us to suspect that the NP and VP are constituents, as the tree structure indicates. Remember, since the string of words *the bright student* is exhaustively dominated by NP and *writing a tough exam* is exhaustively dominated by VP, these two strings of words are represented as constituents and, hence, are predicted to behave as such. On the other hand, the string of words *the bright student is writing* is not exhaustively dominated by a single node, so we do not expect this string of words to behave as a constituent in this sentence. If we are asked what is happening with the tough exam, a response such as, "The bright student is writing," (with the intended meaning of, "The bright student is writing it") would not be an appropriate response. Thus, we are led to believe that the string of words *the bright student is writing* is not a constituent, which is consistent with the structure in (14).

This discussion has illustrated an important point that students often miss. As generative syntacticians, we are trying to develop a model of syntax that predicts how language behaves. Our model consists of the tree structures and the notion of representing constituency with exhaustive domination. We test the model against observed data and the results of the constituency diagnostics (which are given in detail below). If our model is accurate, then those strings of words that act as constituents according to the diagnostics should be exhaustively dominated by a single node. In turn, we can look at our model to see which strings of words it predicts to behave as constituents. The tree in (14) predicts that the following strings of words to be constituents:
writing a tough exam  a tough exam  the smart student

but not (for example) the following:

writing a tough

So far, our model appears to represent the data faithfully. We will continue to assume this basic model and adjust it from time to time to account for additional data that we uncover. We will now continue with various tests for constituency.

5.2.1 Proforms

The first test is often called the replacement test. Many constituents can be replaced by proforms. The notion that a constituent can be replaced by a proform is called the Proform Criterion. Not all XP’s have appropriate proforms available to them, so, as with all tests for constituency, care must be taken in interpreting the results of this test. If this test reveals no appropriate proform for the string of words being tested for constituency, you may wish to try some other tests, first. So, while it is true that if a given string of words can be replaced by a proform is a constituent, the converse is not necessarily true. That is, it is not necessarily the case that if a given string of words cannot be replaced by a proform it is not a constituent.

Common English proforms:

<table>
<thead>
<tr>
<th>Category</th>
<th>Proforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td><em>I, you, he, she,</em> etc.</td>
</tr>
<tr>
<td>VP</td>
<td><em>so</em></td>
</tr>
<tr>
<td>PP</td>
<td><em>there, then</em></td>
</tr>
<tr>
<td>AdvP</td>
<td><em>thus</em></td>
</tr>
<tr>
<td>AdjP</td>
<td><em>such, so</em></td>
</tr>
</tbody>
</table>

NPs can be replaced by pronouns as in the following examples. *The bright student* is replaced by *she* and *a tough exam* is replaced by *it*.

(17) **The bright student** is writing a tough exam.

**She** is writing **a tough exam**.

**She** is writing **it**.

NPs proforms are common throughout the world’s languages, but they are by no means universally applicable. Care must be taken to establish adequate proform replacement tests for different languages. Here is an example from Portuguese that illustrates the pronoun replacement test.
(18) a. Vou falar com a menina que o Pedro viu.
   go.1SG speak with the girl that the Pedro see.PST.3SG
   ‘I will speak with the girl that Pedro saw.’

   b. Vou falar com ela
   go.1SG speak with 3SG.F
   ‘I will speak with her.’

Just as in English, the Portuguese pronoun *ela* replaces an entire noun phrase, showing that it behaves as a constituent.

In English, VPs can be replaced by the proform *so*, as in the following example.

(19) John is writing a tough exam, and so is Mary.

Here, *so* replaces the string of words *writing a tough exam*, indicating that this string of words is a constituent and, thus, that it must be exhaustively dominated by a single node. This provides further evidence for the structure in (14) above – namely where the string *writing a tough exam*, the verb and the object, form a constituent labelled VP, which does not include the subject (or the auxiliary). VP proforms are not as common cross-linguistically. Here is an example from Korean, however (Shim Den Dikken, 2008).

(20) yengcin-i chenchenhi pap-ul mek-ess-ko minho-nun kuleh-ess-ta
   Youngjin-NOM slowly rice-ACC eat-PST-CONJ Minho-TOP PROFM-PST-DECL
   ‘Youngjin ate the rice slowly, and so did Minho.’ (영진이 천천히 밥을 먹고 민호는 그랬다.)

In this example, the preform *kuleh* (labelled PROFM) replaces the string *chenchenhi pap-ul mek* (*eat rice slowly*).

PPs in English can often be replaced with the adverbs *there* or *then* (and often by their equivalents in other languages). Note that with is a tricky preposition, but PPs formed with with can sometimes be replaced by thus. Some English proforms are illustrated in the following examples.

(21) a. Terry peeled the potatoes in the kitchen. Terry peeled the potatoes there.

   b. Terry peeled the potatoes in the morning. Terry peeled the potatoes then.

   c. Terry peeled the potatoes with a knife. Terry peeled the potatoes thus.

AdvPs can sometimes be replaced by *thus*, while AdjPs can sometimes be replaced by *such* (for attributive AdjPs) or *thus* (for predicative AdjPs). Here are some examples.

(22) a. Maria sings rather beautifully. Maria sings thus.

   b. a very tall man such a man

   c. Mr. Jones is very tall ...and so is Mrs. Smith.
**Practice 5.4:**
Re-write the following sentences with a proform to illustrate that the underlined strings of words are constituents.

The man in the moon ate some cheese.
Fred put the apple on the large table that Mary bought.
I met a man very proud of his daughter.
John is reading a book.
Mary is fixing the car in the garage.
Fred read a very interesting book.
That book is very interesting.

### 5.2.2 Stand Alone

If a string of words can stand alone as a single utterance, then it is a constituent. Many (but not all) constituents are well-formed utterances by themselves. Consider the following question/answer pairs.

(23) Stand alone test with question/answer pairs.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who is writing a tough exam?</td>
<td>the bright student</td>
</tr>
<tr>
<td>What is the bright student</td>
<td>writing a tough exam</td>
</tr>
<tr>
<td>What’s going on?</td>
<td>The bright student is writing a tough exam.</td>
</tr>
<tr>
<td>Where is the student writing the exam?</td>
<td>in the classroom</td>
</tr>
<tr>
<td>턱 먹고 싶어?</td>
<td>그 맛있는 사과</td>
</tr>
<tr>
<td>누구를 만났어요?</td>
<td>영화의 아버지</td>
</tr>
<tr>
<td>책을 어디서 얻었어요?</td>
<td>지 쪽 나무 옆에서</td>
</tr>
</tbody>
</table>

Observe that all but one of the constituents shown in (14) satisfies the stand alone test. The auxiliary *is* is technically speaking a constituent, but cannot stand alone. Again, we must stress that not all constituents can stand alone, but all strings of words that can stand alone are constituents.

Cross-linguistically, the stand alone test is quite robust. Again, many languages have restrictions on minimal utterances. For instance, in French it is generally unacceptable to utter a bare noun. It must always have some kind of determiner. Compare the following Onondaga and French examples. In Onondaga one can answer the given question with just one word. This is not possible in French.

(24) a. nwadę wa’hék John?
    what he.ate.it John
    “What did John eat?”
b. swahýona
apple
‘an apple’

(25) a. Qu’est-ce que Jean a mangé?
what’is-DEM that John has eaten
‘What did John eat?’

b. une pomme
an apple
‘an apple’

c. *pomme
apple
(‘an apple’)

5.2.3 Coordination Tests

Only constituents can be coordinated. This test works well for virtually all types of constituents, unlike some of
the other diagnostics, which only work well for certain types of constituents. Consider the following examples.

(26) a. John ate [[NP a cake] and [NP an apple pie]].

b. Mary walked [[PP over the hill] and [PP under the bridge]].

c. Minjoon [[VP ate an apple] and [VP drank some tea]].

In (26-a) the two NPs are conjoined into one large NP, and in (26-b) the two PPs are likewise conjoined.
Example (26-c) shows more evidence that a verb and an object together form a constituent. A further property
of coordination is that only constituents of the same category can be coordinated. Consider the following data.

(27) a. Minsoo is a doctor.
b. Minsoo is a singer.
c. Minsoo is reading a book.
d. Minsoo is drinking some tea.

(28) a. Minsoo is a doctor and a singer.
b. Minsoo is reading a book and drinking some tea.
c. *Minsoo is a doctor and drinking some tea.
d. *Minsoo is [reading a book] and [a singer].
Observe that the coordination of two constituents of the same category is possible in (28-a) and (28-b) but that coordination of two constituents of different categories as in (28-c) and (28-d) is impossible. Note in passing that (28-d) has a bizarre reading in which Minsoo is reading two things. The first is a book, and the second is a singer. This is strange, of course, because one doesn’t normally read people except figuratively, when referring to facial expressions, or unless perhaps they’re wearing clothes with lots of writing on them. This other reading has the following structure.

(29) Minsoo is reading [a book] and [a singer].

Until now, we have just been assuming that a verb and an object together form a verb phrase or VP, but what do we assume this? Using the coordination test, we are in a better position to answer this question. Since only constituents of the same category can be conjoined, we can test to see what kinds of things can be conjoined with a phrase such as reading a book. Here is the test sentence. Substitute X for other phrases whose categories are not in question.

(30) a. Carlos is [[X] and [reading a book]].
   b. *Carlos is [[a magazine] and [reading a book]]. → attempted conjunction with an NP
   c. Carlos is [[laughing] and [reading a book]]. → successful conjunction with a VP

Since the phrase laughing contains only a verb, there is no doubt that it is a verb phrase. Since the phrase reading a book can be conjoined with a VP and not with an NP, it must be a VP, too.

**Practice 5.5:**

Use the coordination diagnostic to show that the underlined string of words is a constituent.

The man in the moon ate some cheese.
Fred put an apple on the large table that Mary bought.
Susan bought a very expensive laptop.
John is reading a book.
Mary is fixing the car with a wrench.
Fred read a very interesting book.
Alice put the books in the attic.

### 5.2.4 Movement Tests (advanced)

In this section, we consider diagnostics in which a string of words is placed in different locations within a sentence. We will start by analyzing some actual data. Consider two possible structures for the sentence John put the cake in the oven. Note that we use triangles here for convenience to show the NP and PP constituents.
How do we choose between (31-a) and (31-b)? We see that the NP the cake and the PP in the oven do not form a single constituent in example (31-a) because there is no one single node that exhaustively dominates the string the cake in the oven. In example (31-b), however, this same string does form a constituent because it is exhaustively dominated by NP node shown. We will use the movement tests described below to determine the correct structure for this sentence.

I Clefting

The first type of movement test we explore is called clefting. A clefted sentence has the following form, such that X is always a constituent. There are some examples to the side in which the constituent in the X position is underlined. As you can see from the examples, NPs and PPs can be clefted in English.

(32) It was X that Y
a. It was the apple that Mary ate.
b. It was the magazine that Minsoo read.
c. It was into the kitchen that Juma walked

Let us use this test in the following two sentences to see if the underlined string of words are constituents or not.

(33) a. John likes the lamp in the living-room.
b. John put the cake in the oven.

To execute this test, we put the underlined string of words in the X position in the cleft. This gives us the following results, showing us that the underlined string of words in the first sentence is a constituent, but that the underlined string of words in the second sentence is not. This results of this test lead us to conclude that (31-a) is the correct structure for the sentence in (33-b).
(34)  a.  It was the lamp in the living-room that John likes.
     b.  *It was the cake in the oven that John put.

Now, let us consider the following ambiguous sentence, with the two paraphrases given below.

(35)  Mary saw the child with binoculars.
     a.  The child had the binoculars, and Mary saw him.
     b.  Mary used binoculars to see the child.

To understand these two readings a bit better, consider the following two questions.

(36)  a.  Which child did Mary see?
     b.  Who did Mary use the binoculars to see?

Both of these questions can be answered by the statement in (35) with the appropriate meanings as given. Now, let us consider the structures for this sentence that distinguish between these two meanings. First let us see what happens when we try to cleft the relevant parts of the sentence.

(37)  a.  It was the child with binoculars that Mary saw.
     b.  It was the child that Mary saw with binoculars.

Here, we see that the range of possible meanings has changed. In (37-a) we only get the reading in (35-a); however, in (37-b) we only get the reading in (35-b). Thus, in order to get the reading in (35-a) the string of words the child with binoculars must be a constituent. To get the reading in (35-b) we see that this string cannot form a constituent, otherwise (37-a) would still be ambiguous. (37-b) shows us that the child is a constituent, of course. Now we are in a position to draw the appropriate structures for the two readings of this sentence.

(38)  

Here, with binoculars modifies the VP, not the NP the child. So, we get the meaning where Mary is using the binoculars to see the child. This is made clear in the following unambiguous paraphrases.

(39)  a.  It was with binoculars that Mary saw the child.
     b.  It was the child that Mary saw with binoculars.
To get the meaning where the child has the binoculars as in the paraphrase in (37-a), the string the child with binoculars must be a constituent. This is shown in the following structure, where the child with binoculars is exhaustively dominated by a single node, NP. Note that for now we use triangles to show the constituency. In the next section we will discuss the internal structure of NP and PP.

(40)  

Here, the child with binoculars is a constituent, so it is available for clefting. Note that the child by itself is no longer a constituent, so it is not available for clefting. (with binoculars is a constituent – but there are other reasons why it cannot be clefted).

The reason (35) is ambiguous is because the PP can adjoin either to the NP or to the VP in English. In many languages, however, a PP can adjoin only to a VP. Consider the following Korean example.

(41)  

II Pseudoclefting

Pseudoclefting is another reliable test for constituency. A pseudoclefted sentence has the following form, again such that X is always a constituent. There are some examples to the side in which the constituent in the X position is underlined. As you can see from the examples, NPs and PPs can be clefted in English. A pseudocleft must select the appropriate wh-word (question word – what for things, who for people, etc.). Not only can NPs and PPs be pseudoclefted, but VPs can be pseudoclefted, too. When a VP is pseudoclefted some form of the verb do is required.

(42)  

Now let’s consider another example with a larger constituent. The following sentence strongly favours the
reading in which the string of words the tea in the blue teapot is a single NP constituent. (Ignore the alternative, rather bizarre reading in which Sandy was inside the teapot drinking tea.)

(43) Sandy sipped on the tea in the blue teapot.
   a. [The tea in the blue teapot] is what Sandy sipped on.

Returning to the examples we were discussing above, again we have evidence that favours the structure in (31-a) over that of (31-b) for the sentence John put the cake in the oven.

(44) a. *The cake in the oven is what John put.
   b. The cake is what John put in the oven.

Again, this test shows that the string of words the cake in the oven is not a constituent, while the shorter string the cake is a constituent. Now for the ambiguous sentence we had above let us apply the pseudocleft test. In (45-a), the string of words the child has been pseudoclefted. This is possible with the structure in (38) but not with the structure in (40). This is because only in (38) is the child a constituent. As a consequence, (45-a) can only mean that the binoculars were used to see the child. Conversely, in (45-b) the child with binoculars must be a constituent, which consistent only with the structure in (40). These are the same results we saw with the clefting diagnostics above.

(45) a. The child is who Mary saw with binoculars.
   b. The child with binoculars is who Mary saw.

III Passive

We will look at passivization in more detail later in this chapter. For now, let us concentrate on the main characteristics of passive constructions and how we can use them as constituency diagnostics. A passive construction is one in which what we conceptually understand to be the direct object of the sentence appears in the position of the subject. Now, observe that only NPs can be passivized. Thus, if a string of words can be passivized, it is a constituent. But, if string of words cannot undergo passivization, it may either be because it is not a constituent or because it is not an NP.

(46) a. John bought a radio.
   b. A radio was bought.
   c. Susan wrote on the paper.
   d. *On the paper was written.

In the first example above, the string of words a radio is the direct object of buy. It is also an NP, and so can undergo passivization as in the sentence that follows it. The string of words on the paper is a constituent, too; however, it is a prepositional phrase (discussed later) not an NP and therefore cannot undergo passivization.
Let’s return to the sentence *John put the cake in the oven*. The structure in (31-a) predicts that the string of words *the cake* can undergo passivization, while the structure in (31-b) predicts that the string of words *the cake in the oven* can undergo this process. Again, the results show that the tree in (31-a) is on the right track.

(47) a. The cake was put in the oven.
    b. *The cake in the oven was put.*

As expected, the passivization test converges with the clefting and pseudoclefting test indicating that (31-a) is the correct structure for this sentence. Let’s consider another ambiguous sentence.

(48) Mary ate the cookies in the kitchen.

This sentence can mean that the kitchen is the place where Mary ate the cookies, or it can mean that it was the cookies in the kitchen that Mary ate (and not the cookies in the living-room, say). Consider, now, whether the following passivized sentence is ambiguous.

(49) The cookies in the kitchen were eaten.

Here, we only get the reading where *in the kitchen* modifies the cookies, not the eating. That is, we get the reading in which it was the cookies in the kitchen that Mary ate (and not the cookies in the living-room). Thus, the string the cookies in the kitchen must be a constituent. With this much in mind, we can sketch a rough tree for this reading of the sentence.

(50)

In this structure the NP *the cookies in the kitchen* is a constituent and so is available for passivization. Convince yourself that this is the correct phrase structure for this reading by applying the other constituency tests above. Now, let us consider the reading where the kitchen is the location in which Mary ate the cookies. The passivized sentence above cannot have this reading. This tells us that the string of words the cookies in the kitchen does not form a constituent under this reading. The following tree captures this fact.
The string *the cookies in the kitchen* is not a single NP constituent. As such, this string of words cannot undergo passivization. The NP *the cookies*, of course, is available for passivization. At this point, you should be able to explain to someone why the passivized sentences above are not ambiguous, while the active sentence Mary ate the cookies in the kitchen is.

### 5.3 The Content of Lexical Categories

#### 5.3.1 Noun Phrases

NPs in English commonly appear with determiners (*the, a*), demonstratives (*this*), question words (*which book*), quantifiers (*many books, some water*), and so forth. Here are some examples.

\[
(52) \quad \text{NP} \quad \text{NP} \quad \text{NP} \quad \text{NP} \\
\text{Det} \quad \text{N} \quad \text{Dem} \quad \text{N} \quad \text{Det} \quad \text{N} \quad \text{Q} \quad \text{N} \\
\text{the} \quad \text{book} \quad \text{these} \quad \text{books} \quad \text{which} \quad \text{zebra} \quad \text{many} \quad \text{people}
\]

NPs can be modified by adjectives, relative clauses and, in some languages, PPs. Here are some English examples with all three.

\[
(53) \quad \text{Examples of English NPs} \\
\text{the big house} \quad \text{the red apple} \quad \text{the happy child} \\
\text{the house that Jack built} \quad \text{the apple that I bought} \quad \text{the child who speaks Swahili} \\
\text{the house on the corner} \quad \text{the apple on the table} \quad \text{the child in the playground}
\]

Here are some examples of Cantonese NPs. Note that in Cantonese a noun cannot be directly modified by a PP. Instead a relative clause is used to describe the same property. The structure of relative clauses is an advanced topic covered later in this chapter. Note that *ge3* is a modifier element (MOD) that links an adjective to a noun. CL is a classifier, which appears in Cantonese whenever a demonstrative or a number is present in the noun phrase.)
(54) a. ni1 go3 gwai1 ge3 sai3 lou6zai2
   DEM CL obedient MOD child
   ‘this obedient child.’

   b. ngo5 sik6 zo2 go2 go3 ping4 gwo2
      I eat PERF DEM CL apple
      ‘the apple that I ate’

Here are the structures for English noun phrases with adjective and a PP as a modifier.

(55)

```
NP
  Det AdjP N
     the big house
```

(56)

```
NP
  Dem Cl AdjP N
    ni1 go3 gwai1 ge3 sai3 lou6zai2
```

Here is a tree for the first Cantonese example.

Consider, now, a noun phrase that has just one word in it, such as John or water. Consider the following sentences.

(57) a. John went to the movies.
    b. John and his friend went to the movies.
    c. Fish need water.
    d. Fish need water and a little food.

The coordination tests tells us that the b and d examples above have the following structure.

(58) a. [[John] and [his friend]] went to the movies.
    b. Fish need [[water] and [a little food]].

Since the strings his friend and a little food are clearly NPs, their associated conjuncts must also be NPs, as follows.

(59) a. [[NP John] and [NP his friend]] went to the movies.
    b. Fish need [[NP water] and [NP a little food]].
Here, then, are the structures for John and a friend. Note that we will not cover possessives here.

NP  NP
  |   /
N  Det  N
  |   |
John  a  friend

**Practice 5.6:**

Draw the structures for the following noun phrases.
this apple  those pears  a big apple
a very interesting book  the man in the moon  water

### 5.3.2 Verb Phrases

Verbs can be classified by the number and kinds of objects they take as a complement. An intransitive verb has no objects. A transitive verb takes a direct object, and a ditransitive verb takes a direct object and an indirect object. Consider the following examples.

(60)  a. cough
     b. eat an apple
     c. give Susan a book

It is quite common to classify verbs by the number of objects they take. In some languages transitive and intransitive verbs have different affixes. Here is a Korean example. Observe the morphological difference in the verbs.

(61)  a. mwul-i kkulh-ess-ta 물이 끓었다.
      water-NOM boil-PST-DECL
      ‘The water boiled.’ (intransitive)

  b. minswu-ka mwul-ul kkulh-i-ess-ta 민수가 물을 끓였다.
     Minsoo-NOM water-ACC boil-CAUS-PST-DECL
     ‘Minsoo boiled the water.’ (transitive)
Adjuncts to the VP express concepts such as manner, location and instruments, etc, and usually take the form of a PP or AdvP. As with all adjuncts, they are considered extra information that is not vital to the rest of the sentence. The adjuncts in the following examples are underlined. Adjuncts almost always appear to the right of the verb and its objects in English, but some adverbs may appear to the left of the verb. In English, an adverb may not appear between the verb and the direct object, although this is possible in other languages. Consider the following examples, where the adjuncts are in italics. The following trees offer a couple examples.

(62)  
   a. eat a turkey on Tuesday  
   b. pass an exam with ease  
   c. paint a picture quickly  
   d. always use a napkin  
   e. sweep the floor with a broom  
   f. read a book on the couch

5.3.3 Adpositional Phrases

Adpositions come in two types: **prepositions** (전치사) and **postpositions** (후치사). Adpositions take an NP complement and generally locate their complement in either time or space. Sometimes an adposition marks a special role on the complement such as benefactive or source. A preposition appears to the left of its noun phrase complement, and a postposition appears to the right of its noun phrase complement. Here are the structures for prepositions and postpositions. Both prepositional phrases and postpositional phrases are abbreviated PP.

(64)  
   PP  
   |  
   P  
   |  
   preposition  
   PP  
   |  
   P  
   |  
   postposition

Here are some examples of prepositional phrases in English.
Japanese and Bengali have postpositions. Here are some examples with trees below.

(66) a. kuruma to [Japanese]
car with 'with a car'

b. babar kache [Bengali]
father near 'near father'

c. PP PP
   NP P NP P
   N to N kache
   kuruma babar

The PP can be modified by a degree word (Deg) such as nearly, almost, or just. In English, degree words appear to the left of the preposition.

(67) PP
    Deg P NP
    just around Det N
    the corner

Degree words:
Degree words in English (and in many other languages) exhibit much variation and change quickly with time. For instance really is now much more common than very. For a good discussion, see [Rika and Tagliamonte](2003).
5.3.4 Adjective Phrases

A number of adjectives can take complements. As with nominal and verbal complements above, there is a lexically encoded semantic relation between the adjective and its complement. For instance, consider the word *fear*, which can be either a noun or a verb, and the related word *fearful* and its more common synonym *afraid*, which are adjectives. All four of these forms can take *mice* as a complement (sometimes with the help of a preposition) and all four convey the same lexical meaning.

(68)  
   a. John fears mice.  
   b. John’s fear of mice.  
   c. John is fearful of mice.  
   d. John is afraid of mice.

Like PPs, AdjPs can appear with a degree phrase. Here are some examples.

(69)  
   AdjP  AdjP  AdjP  
   |      |      |      |  
   Adj  Deg  Adj  PP  
   |      |      |      |  
   quiet  very  quiet  curious  P  NP  
   |      |      |      |      |  
   about  N  syntax
5.3.5 Adverbial Phrases

Not all adverbs are created alike. There is actually quite a large diversity in the type of adverbials found in language. We will focus on two major types here. These are sentence-level adverbs and VP-level adverbs. Sentence-level adverbs modify the entire sentence. They often depict the speaker’s attitude towards the event or state described by the sentence or otherwise modify the event as a whole, including the subject’s participation in it. VP-level adverbs, on the other hand, modify only the predicate – that is the verb and its object. They usually describe the manner in which the event is carried out or the frequency with which it happens. In the following examples, happily is used as a sentence-level adverb and a VP-level adverb. In the first sentence, the adverb describes the manner in which the leaving of the building took place.

(71)  
  a. Happily, John left the building (sentence-level adverb)  
  b. John left the building happily (VP-level adverb)  

Other common sentence-level adverbs include (un)fortunately, presumably and typically. AdvPs can also contain degree phrases. Here are some examples.

(72)  

(70) 

\[ \text{AdjP} \quad \text{AdjP} \]  
\[ \text{Deg} \quad \text{Adj} \]  
\[ \text{PP} \quad \text{PP} \]  
\[ \text{of} \quad \text{of} \]  
\[ \text{N} \quad \text{NP} \]  
\[ \text{Mary} \quad \text{mice} \]  
\[ \text{big} \quad \text{scary} \]  

\[ \text{Adj} \quad \text{Adj} \]  
\[ \text{really} \quad \text{totally} \]  
\[ \text{proud} \quad \text{afraid} \]  
\[ \text{Deg} \quad \text{Deg} \]  
\[ \text{of} \quad \text{of} \]  
\[ \text{N} \quad \text{NP} \]  
\[ \text{big} \quad \text{scary} \]
Practice 5.8:
Draw the full structures for the following noun phrases and verb phrases.

- a very delicious apple
- run very quickly
- push the ball really hard
- the picture on the wall

5.3.6 One-replacement and nested VPs (advanced)

In this section we will look more closely at the structure of NP and of VP. Let’s start with NPs and the one-replacement test. Consider the following examples.

(73) a. this book about linguistics and that one
    b. those travellers from Thailand and these ones
    c. the picture of Mary in the kitchen and the one in the living room

The preform one replaces a constituent inside the NP. Here are the same data again with the portions replaced by one underlined.

(74) a. this book about linguistics and that one
    b. those travellers from Thailand and these ones
    c. the picture of Mary in the kitchen and the one in the living room

Observe that one-replacement never targets the determiner or demonstrative. Looking at the first example, we see that the string book about linguistics is a constituent as it can be replaced by one. Work through the rest of the data to convince yourself of the constituency of the other two examples. Here are the structures for these NPs. Since one-replacement never targets the determiner or demonstrative, determiners must branch off the highest node in the tree.

To represent intermediate nodes, we use an X’, where X stands for whatever phrase we’re looking at. Here are the NPs for the first conjunctions in (73).
Recall that the proform so replaces a VP. Consider, now, the following data.

(76)  
a. Mary ate the cake in the garden with a fork, and so did John.  
b. Mary ate the cake in the garden with a fork, and so did John with chopsticks.  
c. Mary ate the cake in the garden with a fork and so did John in the kitchen with chopsticks.

In these three examples, the proform so replaces different strings of words, indicating that each string of words below is a constituent.

(77)  
a. so = eat the cake in the garden with a fork  
b. so = eat the cake in the garden  
c. so = eat the cake

How can we capture these effects? We use nested VP as follows.
5.4 Clauses

Until now, we have only considered phrases in the sentence such as VP and NP. Let us now see how to construct full clauses. We have to understand first how auxiliaries fit into the tree. Consider the following English sentences.

\[\begin{align*}
\text{(79) a. Mary is reading a book.} \\
\text{b. John has eaten dinner.} \\
\text{c. Fred will fly to Beijing tomorrow.}
\end{align*}\]

Let’s run some constituency diagnostics to determine the structure of sentences with auxiliaries. Consider the following data.

\[\begin{align*}
\text{(80) a. Mary is reading a book, and so is John.} \\
\text{b. John has eaten dinner, and so has Fred.} \\
\text{c. Fred will fly to Beijing tomorrow, and so will Mary.}
\end{align*}\]

\[\begin{align*}
\text{(81) a. Mary is reading a book.} \\
\text{b. What is Mary doing?} \\
\text{c. reading a book}
\end{align*}\]

These data all show that the string of words reading a book is a constituent. Which of the following structures captures this fact?
Only the tree in (82-a) captures the constituency facts we determined. Let move on to different kinds of clauses.

We have examined declarative sentences only, so far. We will look at other kinds of sentences here. We need to distinguish between a proposition and illocutionary force. A proposition is statement that can be true or false. Once you have a proposition, a speaker utters it to an addressee. How the speaker intends the addressee to react to the proposition is called illocutionary force. In a declarative sentence the speaker expects the addressee to accept the proposition as new information. In an interrogative sentence the speaker expects the addressee to fill in the missing information from the proposition.

We present four types of illocutionary force here, but only discuss the first two in detail.

**declarative** A statement: the speaker gives new information to the addressee

**interrogative** A question: the speaker asks the addressee for new information

**imperative** A command: the speaker tells the addressee to do something.

**hortative** A suggestion: the speaker suggests to the addressee to do something together.

Here are examples of these from various languages.

(83) English

a. John ate an apple. [declarative]
b. Did John eat an apple? [interrogative]
c. What did John eat? [interrogative]
d. Eat this apple! [imperative]
e. Let’s eat an apple. [hortative]
The data above shows that illocutionary force is encoded in a number of ways. We will discuss these in more details in the following section.

We now need a way to represent the illocutionary force of a clause. Let’s assume that S encodes the proposition. Once we have a proposition, it needs illocutionary force to be a complete sentence. Looking at the Korean data, we see that illocutionary force is encoded at the extreme right edge of the clause (-ta for declaratives and -ni for interrogatives). In fact, illocutionary force is encoded on a head called COMP under a projection called S’ (S-bar). Let’s examine the trees for the Korean sentences in (84-a) and (84-b) first.
Turning back to English, now, consider the tree for the simple declarative in (83-a).

Unlike Korean, English does not have an overt morpheme that corresponds to declarative force in main clauses. Sentences can also consist of more than one clause. Consider, now, the following examples.

(88)  

a. Mary knows Fred.  

b. Mary knows that Fred ate an apple.

The verb *know* can take an NP as a complement, the object. It can also take a clause as a complement. We say
that the clause *that Fred ate an apple* is the object of the verb *know*.

Now, consider the following examples. The red portion in both sentences is the complement of the verb *know*.

(89)  
(a) John knows Swahili.  
(b) John knows that Susan speaks Swahili.

The two entities in red are similar in many respects. They both express the object of *know*. Both strings express what it is that John knows. Both elements can be resumed by the same pronoun as follows.

(90)  
(a) John knows Swahili, and Mary knows it, too.  
(b) John knows that Susan speaks Swahili, and Mary knows it, too.

In example (89-a) the red portion is the embedded or subordinate clause (종속절). The main clause (주절) is in black and is structurally the highest clause. That is, it is not introduced by another verb or by another complementizer. Let’s investigate the constituency of the embedded clause in red in (89-a). The pronoun replacement test in (90-a) shows that the string that Susan speaks Swahili is a constituent. Consider now the following examples. The second example uses the interrogative complementizer, *if*.

(91)  
(a) John thinks that either [Susan speaks Swahili] or [Sally speaks Māori].  
(b) John wonders if [Susan speaks Swahili] or [Sally speaks Māori].

This example shows that the string *Susan speaks Swahili* is a constituent, too. How can we capture these constituency facts? Consider the following tree.

(92)

A common diagnostic for the matrix subject (the subject of the main clause) in English is found in tag questions.
A tag question consists of the auxiliary of the main clause and a pronoun that refers to the matrix subject. If there is no auxiliary in the main clause, dummy do is used. Here are some examples.

(93)  

(a) John thinks that Mary is smart, doesn’t he?  
(b) *John thinks that Mary is smart, doesn’t she?  
(c) *John thinks that Mary is smart, isn’t he?  
(d) *John thinks that Mary is smart, isn’t she?

Observe that the auxiliary must be do, reflecting the lack of an auxiliary in the main clause. Also, the pronoun must refer to the matrix subject, John, and cannot refer to the subject of the embedded clause or to anyone else.

In a complex sentence, each clause has its own illocutionary force. We will look at declaratives first. Consider the following examples. The embedded clause is in square brackets.

(94)  

(a) John thinks [that Mary left the party].

(b) 민수는 [상빈이 사과를 먹었다고] 생각한다.  

       Minsoo-TOP Sangbin-NOM apple-ACC eat-PST-DECL-COMP think-do-PRS-DECL  
       ‘Minsoo thinks that Sangbin ate the apple.’

(c) Zhangsan shuo le  [Lisi mai le shu]  [Mandarin]  

       Zhangsan shuo PERF Lisi buy PERF book  
       ‘Zhangsan said that Lisi bought some books.’

In English, the complementizer that is used to introduce a declarative clause. In Korean, the embedded clause is introduced with the affix –ko and it is marked as a declarative clause by the affix –ta. Note that this affix appears in the matrix clause, too. Also, note in English that the complementizer that introduces the clause is at the left edge of the clause, while the complementizer that introduces the clause in Korean is at the right edge of the clause. Finally, note that Mandarin does not have any special marking to indicate declarative force. Many language do not have a dedicated marker for declarative force. Observe that in English the declarative complementizer that is often optional.

The embedded clauses we have looked at so far are known as object clauses in traditional grammar. The embedded clause can also be a subject, as in the following examples. Note that while the complementizer is often optional when it introduces an object clause, it is obligatory when it introduces a subject clause.

(95)  

(a) That John left the party early surprised everyone.  
(b) That it rained on Saturday made the picnic-goers sad.
Practice 5.9: Identify the main clauses and the embedded clauses in the following examples.

1. Mary thinks that John ate an apple.
2. Fred said that Alice will win the race.
3. That Kim lost the race disappointed the spectators.
4. 영화는 민수가 사과를 샀다고 말했다.
5. 그 학생은 통사론이 재밌다고 생각한다.
6. 민지는 응답이 왔는지 궁금하다.

Let us know move on to embedded interrogatives. First, recall the tree for (88-b) which contains an embedded declarative.

(96)

In this example, both the main clause and the embedded clause are declarative. In English, embedded declarative clauses are introduced by *that*. Here’s an example in which the embedded clause is interrogative. Note that embedded interrogative clauses are introduced by *if* in English.

(97) Mary wonders if Fred ate an apple.
To conclude, the important points of this section are that clauses consist of a proposition and illocutionary force. The proposition is expressed by $S$ and the illocutionary force is encoded in the COMP head. The core structure of a clause, then, is the following.

$$S' > S > VP$$

### 5.5 Word Order Typology

We have noted above that English has SVO word order and that Korean has SOV word order. In this section we will investigate the word order properties of human language. There are six logically possible word orders, all of which are found in natural language. Here are some examples of each type, followed by some specific examples.

1. **SOV**: Korean, Japanese, Dakota
2. **SVO**: Mandarin, Cantonese, English, Swahili
3. **VSO**: Irish, Berber, Jakaltek
4. **VOS**: Malagasy, Tzotzil, Halkomelem
5. **OVS**: Hixkaryana, Urarina
6. **OSV**: Warao, Apurinã

#### (101) a. Ịjmu ska mini ciŋ [Dakota]  
cat white water wants  
‘The white cat wants water.’
b. Kariuki ameenda sinema [Swahili]
Kariuki has gone cinema
‘Kariuki has gone to the cinema.’ (Mohamad 2001, p. 108)

c. Ú-dïïny Juààny bèh’ cw re’. [San Dionisio Ocotepec Zapotec]
COMPL-hit Juan dog that
‘Juan hit that dog.’ (adapted from Broadwell 2001 ex. 1)

d. Nividy mofo izahay [Malagasy]
bought bread we.EXCL
‘We(excl) bought bread.’ (Keenan 1976, p. 250)

e. biryekomo y-otaha-no wosi [Hixkaryana]
boy AGR-hit-PST woman
‘The woman hit the boy.’ (Derbyshire 1979, p. 38)

f. Anana nota apa [Apuriña]
pineapple I fetch
‘I fetch pineapple.’

The following pie chart shows the distribution of the various word order types. As you can see SOV and SVO are the two most common word types.

For the rest of this section we will concentrate on SOV and SVO languages only. Greenberg (1966) ana-
lyzed several properties of numerous languages around the world and noticed that properties of languages are not distributed randomly. Rather, the properties cluster. We will talk about three properties of SOV and SVO languages here. It is important to note that the following properties are strong tendencies, not universals.

The first property we discuss is the formation of polarity questions, also known as ‘yes-no questions’ as they request an answer of either ‘yes’ or ‘no’. SOV languages tend to rely on sentence-final particles to form polarity questions. Here is a Korean example.

\[(102)\] yenghui-ka sakwa-lul mek-ess-ni? 영희가 사과를 먹었니?
Younghui-NOM apple-ACC eat-PST-INTER
‘Did Younghui eat an apple?’

In this example the affix -ni indicates that this sentence is a question. In English, polarity questions are formed by subject/auxiliary inversion. If the base sentence has no auxiliary, then a “dummy-do” is used. Here are two examples. The first one has an auxiliary in the declarative form, so it inverts with the subject. The second one has no auxiliary in the declarative form, so a dummy-do is inserted.

\[(103)\]
\[(104)\]
\[(105)\]

\[(103)\] a. Mary will read that book.
   b. Will Mary read that book?

\[(104)\] a. John ate an apple.
   b. Did John eat an apple?

The next correlate of SOV and SVO word order is the placement of adpositions, which have seen previously. The placement of adpositions is very tightly correlated with with word order. SOV languages nearly always have postpositions, and SVO languages nearly always have prepositions. Again, let’s illustrate with Korean and its English translation.

\[(105)\] minswu-ka phihayngki-lo pwusan-ey ka-ss-ta
Minsoo-NOM airplane-with Busan-to go-PST-DECL
‘Minsoo went to Busan by plane.’

We can capture the differences between postpositions and prepositions as we saw above, by placing the head of the PP on either the right or the left. Here are the trees for the Korean and English examples.
The third correlate concerns the placement of relative clauses with respect to the noun it modifies. In SOV languages relative clauses are prenominal, and in SVO languages relative clauses are postnominal. Again, let’s illustrate this with Korean and its English translation.

(108)  

\textbf{haksayng-i} \textbf{mek-un} sakwa

\textit{student-NOM eat-ADN apple}

‘the apple that the student ate.’

Here, -ADN means \textit{adnominal}, which refers to a morpheme used to attach relative clauses to nouns in Korean. In the example above the relative clause is in bold-face and the noun it modifies is underlined. Observe the difference in order between Korean and English.
Correlates of SOV and SVO order:

1. Polarity Questions
   - SOV: sentence-final question particles
   - SVO: Subject/Aux inversion or sentence-initial question particles

2. Adpositions
   - SOV: postpositions
   - SVO: prepositions

3. Relative Clause Order
   - SOV: pre-nominal relative clauses
   - SVO: post-nominal relative clauses

We saw above that prepositions and postpositions differ in the order of the head and the complement. Thus, the head of a prepositional phrase is on the left and the head of a postpositional phrase is on the right. We can use this as a common schema for all phrases in SVO and SOV languages; however, we will only look at the VP. In and SVO language the head of the VP is on the left and in SOV languages, the head is on the right. Let’s see this with two examples.

(109) a. eat the cake with a fork
     b. phokhu-lo kheyikhu-lul mek-ess-ta
        fork-with cake-ACC eat-PST-DECL
        ‘ate the ate with a fork’

(110) a. VP
     V  NP  PP
  / \\  /  \
 eat  Det    N    P    NP
 /   |     |   |
the cake with Det N
    a      fork

b. VP
  PP  NP  V
 /  |  |
NP  P  N  mek-ess-ta
  |  |  |
N  lo  kheyikhu-lul
   phokhu

Thus, in general we can say that SOV languages are right-headed and that SVO languages are left-headed.
5.6 Transformations

One of the advances in linguistics in the middle of the 20th century was the observation that certain sentence types seem to be built from more basic sentence types. For instance, the difference in word order between questions and statements is remarkably consistent—so much so that second language textbooks typically give clear instructions on how to form questions from statements. Instructions such as *Place the question word at the beginning of the sentence* are typical of many pedagogical English grammars. In this section, we will look at some typical transformations found in the world’s languages, including polarity questions, passives, scrambling, and wh-movement.

5.6.1 Polarity Questions

Recall that one of the diagnostics for distinguishing SVO and SOV languages dealt with yes/no questions or *polarity questions*. SVO languages typically have Subject-Aux inversion (or Subject-Verb inversion in some languages), while SOV languages have sentence-final question particles. We will examine the syntax of this phenomenon here.

Let’s look at some examples of subject/aux inversion.

(111) a. Is Mary reading a book?
    b. Has John eaten dinner?
    c. Will Fred fly to Beijing tomorrow?

Let us propose that above $S$ is another node, $S’$ (S-bar), that hosts information such as whether the sentence is declarative or interrogative. The head of this phrase is called Comp. We say that the auxiliary *raises* from Aux to Comp as follows, leaving a *trace*. 
We distinguish two representations: Deep Structure and Surface Structure. Deep Structure (DS) is the representation that shows the initial position of the elements in the tree. Surface Structure (SS) is the representation that shows that final position of the elements in the tree, as well as the traces of movement.

5.6.2 Passivization

The next transformation that we investigate is passivization. Passivization is a process by which an object (usually the direct object) is promoted to the subject position. In traditional grammar, at least three voices are traditionally recognized. These include active voice, passive voice, and middle voice. For now, we will concentrate on the first two. Compare some examples.

English

a. Mary ate the apple. [active voice]
b. The apple was eaten by Mary [passive voice]
c. Mary gave a book to John [active voice]
d. The book was given to John [passive voice - direct object passive]
e. John was given a book [passive voice - indirect object passive]
Consider the first example. Notice that *the apple* plays the same role in both sentences. In both cases, it’s the thing that was eaten. The same property holds for the house in the Korean examples and the chopsticks in the Cantonese examples. At this point, we need to distinguish *logical arguments* from *grammatical arguments*. Logical arguments of a verb are what we understand to fill the role normally indicated by the verb. For example, the logical subject of *eat* is the person or animal doing the eating, and the logical object of *eat* is the piece of food that was eaten. Thinking back to our first example above, the apple is the logical object in both the active and the passive sentences. Grammatical arguments are those that are related to some grammatical property of the sentence. Consider subject agreement in the following English sentences.

(117)  
(a) John has stolen a book from the library.  
(b) John and Mary have stolen a book from the library.  
(c) A book was stolen from the library.  
(d) Some books were stolen from the library.

Note that the grammatical subject controls agreement on the auxiliary in English (and in many languages). The grammatical subject also receives nominative case in many languages. The notion of the grammatical object is a bit more difficult to pin down in English. English does not have object agreement (although some languages do). The grammatical object typically appears with accusative case; however, case is marked only on pronouns in English. One property of objects in English is that they must be adjacent to the verb (or verb + particle with phrasal verbs). For this discussion, we will be concerned more with logical objects. In a passive construction, the logical subject appears in an optional *by*-phrase in English. (Note that in Cantonese the *by*-phrase is obligatory.) For our purposes we will simply assume that the *by*-phrase (if present) is attached to the VP like any other PP.

How do we represent the fact that *the apple* is both the logical object and the grammatical subject. Recall
from above that we have two representations: Deep Structure and Surface Structure, abbreviated DS and SS, respectively. Let’s make more precise these two notions. DS is where logical relations are assigned, and SS is where grammatical relations are assigned. With this in mind, let’s consider the DS and the SS for (114-b) ignoring the by-phrase for now.

(118) Deep Structure for (114-b)

\[
\begin{array}{c}
S' \\
\text{COMP} \\
\text{S} \\
\text{Aux} \quad \text{VP} \\
w\text{as} \quad \text{V} \quad \text{NP} \\
\text{eaten} \quad \text{Det} \quad \text{N} \\
\text{the} \quad \text{apple}
\end{array}
\]

In this sentence, the apple is the logical object of the verb eat, so it appears in object position at DS. It is the grammatical subject, however, so at SS it raises to subject position.

(119) Surface Structure for (114-b)

\[
\begin{array}{c}
S' \\
\text{COMP} \\
\text{S} \\
\text{NP}_1 \quad \text{Aux} \quad \text{VP} \\
\text{Det} \quad \text{N} \quad \text{was} \quad \text{V} \quad t_i \\
\text{the} \quad \text{apple} \quad \text{eaten}
\end{array}
\]

Deep Structure
- the level of representation that encodes logical relations and the base position of auxiliaries

Surface Structure
- the level of representation that encodes grammatical relations, the raised position of auxiliaries, and the traces of movement

Cross-linguistically, passives are extremely common. Most languages, however, have only direct object passives. Indirect object passives are quite rare. Two English examples were shown above. They are repeated here, along with some German examples.
(120) English Passives
   a. Mary gave a book to John [active voice]
   b. The book was given to John [direct object passive]
   c. John was given a book [indirect object passive]

(121) German Passives
   a. Petra gab dem Student einen Apfel. [active voice]
      Petra gave the.DAT student a.ACC apple
      'Petra gave the student an apple.'
   b. Der Apfel wurde dem Student gegeben. [direct object passive]
      the.NOM apple was the.DAT student given
      'The apple was given to the student.'
   c. Dem Student wurde einen Apfel gegeben. [indirect object passive]
      the.DAT student was a.ACC apple given
      'The student was given an apple.'

In German, the indirect object receives dative case. Unlike in English, when the indirect object is passivized it does not receive nominative case, but rather retains dative case.

We end this discussion with an illustration of active and passive voice in Korean. Here are the DS and SS structures for (115-b).

(122)
Answers to Practice Questions

Practice Question 1.
fg – yes, the string fg is exhaustively dominated by D

  hij – yes, the string hij is exhaustively dominated by E

  ijc – no, the string ijc is dominated by A, but A dominates other material as well

fgh – no, the string fgh is dominated by B, but B dominates other material as well

fghij – yes, the string fghij is exhaustively dominated by B

ghij – no, the string ghij is dominated by B, but B dominates other material as well

  hi – no, the string hi is dominated by E, but E dominates other material as well

  f – yes, the string f is dominated by F.

Practice Question 2. He ate some cheese.
Mary talked to him about the test.
Fred saw someone eating them.

Practice Question 3. les pommes rouges (‘the red apples’)
le vin d’Italie (‘the wine from Italie’)
la radioactivité des isotopes du polonium (‘the radioactivity of isotopes of polonium’)

Practice Question 4. The man in the moon did so.
Fred put the apple there.
I met such a man.
So is John.
Mary is fixing the car there.
Fred is reading such a book.
So is that book.

Practice Question 5. The man in the moon ate some cheese and drank some milk.
Fred put an apple on the large table that Mary bought and on the counter.
Susan bought a very expensive laptop and an inexpensive computer bag.
John is reading a book and eating a snack.
Mary is fixing the car with a wrench and with a jack.
Fred is reading a very interesting and extremely compelling book.
Alice put the books in the attic and under the table.
Practice Question 6.

NP
Dem N Dem N Det AdjP N
this apple those pears a big apple

NP
NP
Det N PP N
a very interesting book the man in the moon water

Practice Question 7.

VP
V NP NP V NP
eat Det N N 먹다 write P NP
an apple 사과를 with Det N

VP
PP V V PP
NP P 가다 push Det N P NP
N 로 the ball with Det N
기차

VP
PP NP V
NP P N 먹다
N 으로 국수를
순가락
Practice Question 8.

NP
  Det a
  AdjP very
  N delicious

NP
  Dem 그
  N 책
  V run
  AdvP very
  Adv

VP
  AdvP Adv
  V run

AdvP
  Adv very

AdvP
  Adv very

AdvP
  Adv very

VP
  N the
  AdvP really
  Deg hard
  Adv
  N the
  V push

AdvP
  Adv

AdvP
  Adv

RevP
  Adv

VP
  N the
  AdvP Adv
  N the
  V push

AdvP
  Adv

AdvP
  Adv

AdvP
  Adv

Practice Question 9. The main clauses are underlined and the embedded are boldface.

1. Mary thinks that John ate an apple.

2. Fred said that Alice will win the race.

3. That Kim lost the race disappointed the spectators.

4. 영화는 민수가 사과를 샀다고 말했다.

5. 그 학생은 통사론이 제맛다고 생각한다.

6. 민지는 용남이 왔는지 궁금하다.
Exercises

Question 1. Draw SS trees for the following English sentences. If any sentence is ambiguous, draw the trees for both meanings and give unambiguous paraphrases.

1. The very hungry student ate the really tasty apple.
2. John thinks that Mary will win the race.
3. Will the children eat the cake with a fork?
4. Did the woman with the microscope quickly identify the bacterium?
5. John told Mary that Bill spoke to the boy with a loud voice.
6. The note was scribbled in the book with a pen. Was Mary told that the letters for Bill were mailed to Busan.

Question 2. Draw SS trees for the following Korean sentences.

1. 민수가 그 사과를 영희에게 꿰웠다.
   Minswu-ka ku sakwa-lul Yenghui-eykey cwu-ess-ta
   Minsoo-NOM DEM apple-ACC Younghui-to give-PST-DECL
   'Minsoo gave that apple to Younghui.'

2. 영희는 규민이 그 학생이 책을 읽었다고 말했다.
   Yenghui-nun kyumin-eykey ku haksayng-i chayk-ul ilk-ess-tako malha-yess-ta
   Younghui-TOP Kyumin-to DEM student-NOM book-ACC eat-pst-decl read-PST-DECL
   'Younghui told Kyumin that that student read a book.'

3. 민효는 집이 칠해졌다고 생각한다.
   Minho-nun cip-i chilhaycy-ess-tako sayngkakha-n-ta
   Minho-TOP house-NOM paint.PASS-PST-DECL think-PRS-DECL
   'Minho thinks that the house was painted.'

Question 3. How is interrogative force encoded on yes-no questions in the following Cantonese data?

a. John sik6 zo2 go3 ping4gwo2
   John eat PERF CL apple
   ‘John ate the apple.’

b. John sik6 m6 sik6 zo2 go3 ping4gwo2
   John eat NEG eat PERF CL apple
   ‘Did John eat the apple?’
c. John za1 ce1
   John drive car
   ‘John drives.’

d. John za1 m6 za1 ce1
   John drive NEG drive car
   ‘Does John drive?’

**Question 4.** Consider the following Mongolian data (Guntsetseg, 2016). Determine whether Mongolian is an SVO or an SOV language. Explain your answer and draw a tree for the first sentence.

(124)  a. bid akt bolgon-uud-yg šalga-dag
       we document every-PL-ACC examine-HAB
       ‘We examine every document.’

b. ˇci Mongol ruu jav-na
    you Mongolia to go-FUT
    ‘You are going to Mongolia.’

c. Tujaa ene nom-yg xen-d og-sön be?
   Tujaa this book-ACC who-DAT give-PST Q
   ‘Who did Tujaa give this book to?’

d. Tujaa Dorž-ijn biˇc-sen nom-yg unš-san
   Tujaa Dorž-GEN write-PST book-ACC read-PST
   ‘Tujaa read the book that Dorž wrote.’
Further Reading

- Adger (2003) - A slightly advanced textbook introducing current syntactic theory
- Lee (2019) - A comprehensive introduction to Korean syntax and semantics. Note that the theoretical background is rather advanced
- Haegeman (2006) - An introductory syntax textbook that emphasizes argumentation
- van Gelderen (2010) - An introductory textbook focusing on English syntax
Chapter 6

Semantics

By the end of this chapter you should:

• be familiar with different properties of word and sentence meaning
• understand the different semantic relationships between words and between sentences
• understand entailment and presupposition
• be familiar with the concept of R-Expressions
• understand the basic principles of denotational and compositional semantics

6.1 Introduction

In the previous chapters we have investigated word structure and sentence structure. Semantics is the study of meaning. We can ask ourselves how a sentence obtains a meaning once the sentence is built. The following examples illustrate how meaning and sentence structure interact.

(1) a. The editor didn’t discover many errors.
    b. Many errors weren’t discovered by the editor.

(2) a. 누군가 많은 책을 읽었다.
    b. 많은 책을 누군가 읽었다.

In the English example, the first sentence leads us to believe that there are few errors in the text in question; however the second sentence leads us to believe there are many. The first Korean sentence is ambiguous between the following two readings: (i) There is one person who read many books, and (ii) there are many books, each of which was read by someone; however, it could be a different person who read each book. The second Korean sentence has only the first reading in (i).

How do we determine if a sentence has a meaning? Do the following sentences have a meaning?
(3) a. The King of France is bald.
    b. There’s a unicorn in my backyard eating my flowers.
    c. The fireplace at the front of the classroom needs to be repaired.
    d. How long has it been since you’ve stopped smoking? (asked to someone who has never smoked)

These sentences are challenging from a semantic perspective for several reasons. France is a republic, so of course has no King. Unicorns don’t exist. It is almost certainly the case that whatever classroom you’re in has no fireplace. If you ask someone when they stopped smoking, it presupposes that they used to smoke. What if they never smoked? Although these observations make it challenging to decide what the meanings of these sentences are, we do not wish to say that they have no meaning. We will touch on some various aspects of the study of meaning in this chapter.

6.2 Word and Sentence Meaning

6.2.1 Semantic Properties of words

In this section we discuss the semantic properties of words, also known as lexical semantics. We start with the concept of synonymy. If two words have the same or nearly the same meaning in some or all contexts, then these words are said to be synonyms (동의어). Consider the following example.

(4) a. We went to Halifax for our vacation.
    b. We went to Halifax for our holidays.

The words vacation and holidays are synonymous in this situation. Note, however, that in a pair of synonymous words one of the two words could have additional senses or meaning. Holidays, for instance, can also refer to concepts such as Christmas, New Years, and so forth. The word vacation does not have this additional meaning. Nevertheless, they are still considered synonyms since they have nearly identical meanings in at least one context. It is somewhat rare for two words to have identical meanings in all possible contexts. The following examples, however, are nearly completely synonymous. Note that spit and expectorate differ only in level of formality.

(5) a. 흑색, 검은색
    b. spit, expectorate

The following examples also show that kill, murder, and assassinate are also synonymous in some contexts. There are differences in meaning, however, in that the object of murder must be human and the object of assassinate must be a famous human. The Korean words 죽다 and 살인하다 exhibit the same properties as kill and murder.
Moving on, if two words have partly opposite meanings in at least some contexts, then these two words are said to be **antonyms** (반의어). In the following example, *dark* and *light* are antonyms.

(9)  
(a) This room is very dark.  
(b) This room is very light.

These two words are not antonyms in all contexts, however. Again, consider the following examples. Although we can use *dark* to describe someone’s personality, we cannot use *light* to express the opposite meaning in this context.

(10)  
(a) John has a dark personality.  
(b) #John has a light personality.

Antonyms can differ with respect to a subset of features of the two words. Consider the following examples.

(11)  
(a) Pat is a boy.  
(b) Pat is a girl.  
(c) 규민은 남자야  
(d) 규민은 여자야

The words *boy* and *girl* differ in gender, but they both refer to human children. They are still antonyms despite the fact that they share some parts of meaning.

For the next concept it is useful to review what we mean by the **form** of a word. The form of a word refers to the phonological content of a word. Consider the following examples. The words *write* and *right* have the same form. The words *bow* [바우] (=bend over) and *bow* [보이] (hair ornament) do not have the same form. Thus, it is vital not to become confused by orthography.

If two meanings of a given form have related meanings, then the form associated with each meaning is
polysemous (다의어). Consider the following example.

(12) a. This lamp is very bright.
    b. That student is very bright.

The forms *bright* (=light) and *bright* (=intelligent) have related meanings in that one is a metaphorical extension of the other. These two senses of bright are **polysemes**. We say the word *bright* has two polysemous senses or meanings. Consider now the following examples.

(13) a. John works in a bank. (=financial institution)
    b. John docked his boat at the bank. (=edge of a river)

These meanings are completely unrelated. We say that the two words *bank* and *bank* are **homophones** (동음이의어). Note the difference here. There is one word *bright*, which has two polysemous senses, and there are two words *bank* and *bank*, which are homophonous. In Korean 눈 meaning 'eye' and 눈 meaning 'snow' are homophones.

The distinction between polysemy and homophony is often difficult to make. The distinction is important in lexicography in order to record entries in the dictionary. For lexicographers some knowledge of the etymology of the word or words is required. Homophonous words are often listed separately in the dictionary, while a single word with two polysemous senses is given one entry. Observe closely the entries for *bank* and *bright* in the Merriam Webster dictionary.

Ultimately, there is often no clear distinction as the appropriate historical information is lacking, and there are many grey areas. Note that some homophonous words actually derive from a common source, but so much time has passed that the words are typically categorized as distinct but homophonous words. The words *flour* and *flower* derive from the same origin.

It is suggested that the different homophonous words for *mug* may have arisen by metaphorical extension. A mug can refer to a coffee cup, and a mug shot refers to a picture taken by the police when someone is arrested. Typically, mug shots are not very flattering. The word "mug" is sometimes used as a colloquial word to mean "face". To mug someone means to steal their possessions in the street. Often, the process of mugging involves hitting the victim to incapacitate them. This sometimes mean being hit in the face, leaving a bruise or mark on the face. Finally, there is a practice of drawing ugly faces on coffee mugs. These other senses of *mug* may all be historically related. Nevertheless, given the lack of information, they are listed as separate entries in most dictionaries.
bright  adjective
\b brit \n
Definition of bright (Entry 1 of 3)
1  a : radiating or reflecting light : SHINING, SPARKLING
   // bright lights
   // bright eyes

 b : SUNNY
   // a bright day
   also : radiant with happiness
   // bright smiling faces
   // bright moments

2  : ILLUSTRIOUS, GLORIOUS
   // brightest star of the opera

3  : BEAUTIFUL

4  : of high saturation or lightness
   // bright colors

5  a : LIVELY, CHEERFUL
   // be bright and jovial among your guests
   — William Shakespeare

 b : INTELLIGENT, CLEVER
   // a bright idea
   // bright children

6  : AUSSPICIOUS, PROMISING
   // bright prospects for the future

Figure 6.1: Entry for bright
bank  noun (1)

\ˈbæŋk\ 

Definition of bank (Entry 1 of 5)

1  : a mound, pile, or ridge raised above the surrounding level: such as
   a  : a piled-up mass of cloud or fog
       // a fog bank
       // a bank of dark clouds
   b  : an undersea elevation rising especially from the continental shelf

2  : the rising ground bordering a lake, river, or sea or forming the edge of a cut or hollow
   // We lived along the banks of the Mississippi River.

3  a  : a steep slope (as of a hill)
       // climbed a steep bank up to the cabin
   b  : the lateral inward tilt of a surface along a curve
       // The engineers hadn't given the road enough bank.
       : the lateral inward tilt of a vehicle (such as an airplane) when turning
       // The bomber crossed the target area in a sharp bank.

4  : a protective or cushioning rim or piece

Figure 6.2: Entry for bank
bank noun (2)

Definition of bank (Entry 3 of 5)

1 a : an establishment for the custody, loan, exchange, or issue of money, for the extension of credit, and for facilitating the transmission of funds
   // paychecks automatically deposited into the bank
   // went to the bank to make a withdrawal
   // open a bank account

   b obsolete : the table, counter, or place of business of a money changer

2 : a person conducting a gambling house or game
   specifically : DEALER

3 : a supply of something held in reserve: such as
   a in games : the fund of supplies (such as money, chips, or pieces) held by the banker (see BANKER entry 1 sense 2) or dealer

   b in games : a fund of pieces (such as dominoes) from which the players draw
   // select another domino from the bank

4 : a place where something is held available
   // memory banks
   especially : a depot for the collection and storage of a biological product
   // a blood bank

Figure 6.3: Entry for bank
While lexicographers are concerned with writing dictionaries and recording historical uses of words, psycholinguists are interested in how the mental lexicon is organized. Thus, while the three uses of *mug* described above may have some historical relationship, it is extremely unlikely that these three meanings are connected in the same way in the mind of the contemporary English speaker.

### 6.2.2 Semantic Relationships among Sentences

Two sentences that have the same meaning in at least one context are *paraphrases* of each other. Two sentences have the same meaning if they have the same *truth conditions* - the set of things that must hold in order for a sentence to be true. Consider the following examples.

(14)  
- a. The police chased the burglar.  
- b. The burglar was chased by the police.

(15)  
- a. 민수가 사과를 먹었다.  
- b. 사과를 민수가 먹었다. (scrambling – change in order of phrases)

Both of sentences in each pair have the same truth conditions. In other words, it is impossible for one of these sentences to be true and for the other to be false. Sometimes, a sentence can be ambiguous, in which case a paraphrase can be equivalent to one of the meanings. Recall the following example from the chapter on syntax.

(16)  
- a. John ate the apple in the kitchen.  
- b. John ate the apple that was in the kitchen. (paraphrase 1)  
- c. John’s eating of the apple took place in the kitchen. (paraphrase 2)

These two paraphrases disambiguate the original sentence. If two sentences have exactly the same truth conditions, they are said to be truth-conditionally equivalent. This is the case for the burglar examples on the previous slide, but not the apple in the kitchen examples above.

Note that not all examples of passives give rise to paraphrases. Consider the following examples from the beginning of this chapter.

(17)  
- a. The editor didn’t discover many errors.  
- b. Many errors weren’t discovered by the editor.

These two sentences have quite divergent meanings. The first sentence means strongly suggests that the manuscript the editor is reviewing contains very few errors. The second sentence indicates that there are many errors in the manuscript and that these errors weren’t discovered by the editor.

Likewise, scrambling in Korean can give rise to different meaning. Consider the following pair of examples.
Ambiguity refers to the property of having two or more distinct meanings. An expression can either be lexically ambiguous or structurally ambiguous. Lexical ambiguity holds when a word has two or more meanings. This is the same concept as polysemy above. For instance, the word pen can refer to a writing instrument or to a small cage to hold an animal. Structural ambiguity holds when an expression has more than one possible morphological or syntactic structure. Recall from the chapter on syntax that we worked with several examples of ambiguous sentences. These sentences are structurally ambiguous. Let’s remind ourselves of an example.

(19) Mary ate the cookie in the kitchen.

Recall that this sentence is structurally ambiguous between the following two trees.

paraphrase: It was the cookie in the kitchen that Mary ate.
paraphrase: It was the cookie that Mary ate in the kitchen.

Compounds can also be structurally ambiguous. The compound noun *steel-knife-sharpener* can mean either a sharpener for steel-knives or a knife-sharpener made out of steel. Here are the structures.

Newspapers are notorious for ambiguous headlines. The following examples are from a 2010 article in The New York Times: [https://www.nytimes.com/2010/01/31/magazine/31FOB-onlanguage-t.html?ref=magazine&_r=1&](https://www.nytimes.com/2010/01/31/magazine/31FOB-onlanguage-t.html?ref=magazine&_r=1&)

(20) a. British left waffles on Falklands
   b. Gator Attacks Puzzle Experts
   c. Squad Helps Dog Bite Victim

*Garden Path Sentences* are sentences that start out leading the reader to conjure up one structure but end up having another. They are not ambiguous. Consider the following two examples. In the first example, it appears that *ducks* is a noun and is the direct object of *hunts*; however, then the rest of the sentence doesn’t make sense. *Ducks* is actually a verb, forming a phrasal verb (*duck out*). The subject of the sentence contains a relative clause, *the man who hunts*.

(21) a. The man who hunts ducks out on weekends.
   b. The cotton clothing is usually made of grows in Mississippi.
The final two properties are related to truth conditions. A **tautology** is a statement that is logically true under every possible situation. Consider the following sentence. There is no situation in which this sentence could be false (assuming both instances of *John* refer to the same person).

(22) If John is sick and Mary is sick, then John is sick.

A **contradiction** is a statement that is logically false under every possible situation. The following sentence can never be true in any context.

(23) Some people that are sick are not people.

**Entailment** is a relation between propositions. Entailment can be defined as follows.

(24) A entails B iff every truth condition that satisfies A also satisfies B.

Entailment is also known as a *logical consequence* (논리적 귀결), but the term ‘entailment’ is more common in semantics. Let’s work through an example.

(25) a. Every student read chapter six.
   b. Every Korean student read chapter six.

In the example above (25-a) entails (25-b). Entailments have the following truth conditions.

(26) A → B (read as "A entails B")
   a. If A is true then B is true.
   b. If B is false then A is false.
   c. If A is false, then B is either true or false.
   d. If B is true, then A is either true or false.

Let’s apply this to the example above. If it is true that every student read chapter six, then it must be true that every Korean student read chapter six. If it is false that every Korean student read chapter six, then it must be false that every student read chapter six. The property of entailment can also be illustrated with Venn diagrams. Figure 6.5 illustrates entailment in (25) above. C is the set of all students. D is the set of all Korean students. If everyone in C (the larger circle) has read chapter six, then everyone in D (the smaller circle) must have read chapter six.
6.3 A Brief History of Semantic Thought

In the late 19th and early 20th centuries, the study of semantics was dominated by Structuralist linguistics as developed by Jakobson and Trubetzkoy in Europe and by Bloomfield in North America. Many of these methods are what we saw in earlier chapters. Recall the precise methodologies for studying phonology and morphology, where phonemes and morphemes could be determined by comparing minimal pairs. The same methodologies were not easily carried over to the study of meaning. As such, early studies in semantics came from philosophical traditions rather than linguistic traditions. Early grammars until the mid 20th century contained virtually no discussion of semantics.

Modern semantic thinking arguably started with Gottlob Frege. In 1892, Frege introduced the terms Sense and Reference (German ‘Sinn und Bedeutung’ 늘과 지시체) (Frege, 1892). Reference can be understood easily with proper nouns. The reference of the singer 조수미 is the actual opera singer herself. The sense of 조수미 includes the notion "Korean coloratura soprano". The sense of a name differs from speaker to speaker. Some people have never heard of 조수미, so have no sense of the name. Her family members, on the other hand, have a much more detailed sense of the DP "Jo Sumi".
Let’s take another example. Consider the name ‘Elizabeth Alexandra Mary Windsor’ and the title ‘Sovereign of the Royal Victoria Order’. Both of these proper nouns have the same reference, namely, Queen Elizabeth II. They have different senses, however. The sense of ‘Elizabeth Alexandra Mary Windsor’ includes her being the Queen of the United Kingdom and the Commonwealth. It also includes her numerous roles within the British Royal Family. The sense of ‘Sovereign of the Royal Victoria Order’ includes the set of official duties of this person.

A referring expression, or R-expression, is an expression that has a reference in the universe of discourse. Note that the universe of discourse is not necessarily the real world. Consider the following example, assuming the speaker is discussing the real world.

(27) John thinks there’s a unicorn and a rabbit in his backyard. He said the unicorn ate his daisies.

Unicorns don’t exist in the real world that we know, so there is no reference for them. The phrase “the unicorn” in this example has a sense, though. Likewise, a phrase such as "the King of France" and 도깨비 (the mythical creature, not the TV show) both have a sense, but no reference. Note furthermore that if John says, “There’s a unicorn eating daisies in my backyard.” then the phrase a unicorn has both a sense and a reference in the universe of discourse since in the universe that John is talking about the unicorn exists.

6.4 Denotational and Compositional Semantics

Recall that syntax is the study of sentence structure. Syntax studies the rules for putting together words to produce sentences. In addition to the structural properties of the sentence, the sentence also has to make sense. That is meaning derived from the way the bits of the sentence compose. Compositional semantics started with Richard Montague starting in the 1960’s (Montague, 1970b,a, 1973).

Within Montagovian semantics R-espressions denote entities. We will restrict ourselves to definite noun phrases as the denotation of indefinite noun phrases is tricky. Denotation is expressed by double square brackets as in the following examples.

(28) a. \[ [Jo Sumi] = \text{the opera singer by the name of Jo Sumi} \]
    b. John bought a book and a magazine. He decided to read the book. \[ [ \text{the book} ] = \text{the book that John bought} \]

The denotation of a declarative sentence is a truth value. By convention, the truth values are 1=true and 0=false. Consider the following examples.

(29) a. \[ [ \text{Seoul is the capital of the Republic of Korea.} ] = 1 \]
    b. \[ [ \text{Busan is the capital of the Republic of Korea.} ] = 0 \]
We now move on to verbs and predicates. Let’s consider an intransitive verb first. Consider the sentence *Fred sings*. This sentence is true if and only if Fred sings. The denotation of *Fred* is Fred himself, and the denotation of *Fred sings* is 1, if Fred does indeed sing. What is the denotation of *sing(s)*? For the purposes of this discussion we will ignore inflectional morphology on the verb. The denotation of *sing* must be something that can take *Fred* and return a truth value. Let’s assume that the denotation of *sing* is the set of people who sing. If Fred is a member of that set, then a truth value of 1 is returned. Let’s assign the following denotation to *sing*. In the following notation the | symbol means ‘such that’.

\[
[[ \text{sing} ]] = \{ x | x \text{sings} \}
\]

Let’s consider a simplified tree for *Fred sings*. We will ignore the S’ node here.

\[
\begin{align*}
S & \rightarrow NP \quad VP \\
NP & \rightarrow Fred \\
VP & \rightarrow \text{sings}
\end{align*}
\]

We can use the following semantic notation to assign a meaning to this tree.

\[
\{ x | x \text{sings} \} (Fred) = \text{Fred sings} = 1
\]

The formalism works as follows. The variable x is filled in by an argument. Thus, we can understand predicates as functions that take arguments.

\[
\begin{align*}
& a. \quad \{ x | x \text{sings} \} (Fred) \\
& b. \quad = \text{Fred sings} \\
& c. \quad = 1
\end{align*}
\]

Let’s consider now a transitive verb. Consider the sentence *Mary read this book*. The denotation of the VP is determined the same way. The denotation of *read this book* is \{ x | x \text{read this book} \}. We can use the same structure as above.
Ultimately, we want to know the denotation of *read*. Let’s look inside the VP. *Read* is a transitive verb that takes an object. Thus, we can think of a transitive verb as a set of sets. It includes the set of people who read ‘this book’, the set of people who read a magazine, the set of people who read a menu at a restaurant, and so forth. Here is the denotation of *read*, again ignoring inflection.

\[
[[\text{read}]] = \{y | \{x | x \text{ reads } y\}\}
\]

This gives us the following tree.
Exercises

**Question 1.** What is the semantic relation between each pair of words below. Construct a pair of sentences to show the relationship.

tall short

difficult hard

happy joyous

dull boring

**Question 2.** What is the semantic relationship between each of the following pairs of sentences?

Mary read that book.
That book was read by Mary.

The tall students read the book.
The students read the book.

Everyone likes tuna.
Someone likes tuna.

Fred ate the soup quickly.
Fred consumed the soup quickly.

**Question 3.** For each sentence below create one sentence that is entailed by the given sentence and one sentence that entails the given sentence. The first one is done for you.

1. Yenghui lives in Gyeonggi-do
   - 1 → Yenghui lives in South Korea
   - Yenghui lives in Suwon → 1

2. Fred ate an apple.
3. A boy sneezed.
4. Mary is a doctor.
5. Susan flew to Manila.
6. 영화가 생물학을 공부한다.
7. 학생 한 명이 잠을 잡다.
8. 그 사람은 대학교에서 일을 한다.

**Question 4.** State all the entailment relationship among the following sentences.

1. A student studies a Slavic language.
2. Someone studies a Slavic language.
3. A student studies an Indo-European language.
4. Someone studies Russian.
5. A student studies Russian.
6. A tall student studies an Indo-European language.
7. A tall student studies Russian.

**Question 5.** What is the denotation of each verb in the following sentences?

Mary ate the apple.

John gave Bill a book.

Alice sneezed.

John thinks that Mary likes apples.
Further Reading

- **Kempson (1977)** - an older textbook that covers the foundational material discussed here in much greater depth.

- **Cann et al. (2009)** - an updated version of the textbook above

- **Saeed (2011)** - a popular textbook for undergraduate semantics classes. It should be accessible after covering the concepts here.

- **Lee (2019)** - a recent textbook on Korean syntax and semantics

- online resources:
  
  - The Stanford Encyclopedia of Philosophy: [https://plato.stanford.edu/](https://plato.stanford.edu/)
  
Chapter 7

Pragmatics

By the end of this chapter you should:

• understand Gricean Maxims
• understand the notions of illocutionary acts and perlocutionary acts
• understand and diagnose the difference between implicatures, assertions, presuppositions, and entailments
• be familiar with different kinds of implicatures
• understand the difference between an implicature and an impliciture
• understand the notion of Common Ground

7.1 Introduction

Pragmatics is the study of a speaker’s intentions and meanings. It crucially focusses on how language is used in particular situations. We typically distinguish between the sentence meaning and the speaker meaning. The sentence meaning is derived from the structure of the sentence. The speaker meaning is the idea that the speaker wishes to convey. Let’s consider a small example.

John: Are you going to Fred’s Party tonight?

Barb: I have to work late.

Barb’s answer doesn’t appear to have anything to do with John’s question. What she means, of course, is that she is unable to go to Fred’s party. Thus, the speaker meaning and the sentence meaning of Barb’s statement are quite different. Such differences between speaker meaning and sentence meaning are common cross-linguistically. Here is a Korean example.

철수: 영화를 보려고 갈래?
영화: 공부해야 되는데.
Again, Younghui’s answer indicates she either cannot or does not want to go to the movies with Cheolssoo. This chapter discusses how speakers use language to convey thoughts and meaning.

### 7.2 Implicature

Consider again the brief dialogue in the introduction. When Barb answers John’s question, she does not directly say "yes" or "no". She *implicates* that she cannot go to Fred’s party because of work obligations. We say that her statement contains an implicature (Grice, 1975). Implicature is notated as follows.

\[
(1) \quad A \Rightarrow B = A \text{ implicates } B
\]

Gazdar (1979, p.38) defines implicature as "a proposition that is implied by the utterance of a sentence in a context even though that proposition is not a part of nor an entailment of what was actually said." Consider now the following examples.

(2)  
   a. Mary ate a green apple, but she didn’t eat an apple.  
   b. Barb: I have to work tonight...but I’ll go to the party anyways.

The first sentence is a contradiction, but the second sentence is not. Recall the notion of entailment from the previous chapter. The sentence *Mary ate a green apple* entails that Mary ate an apple. However, the sentence goes on to assert that she did not eat an apple. This is a contradiction. The second sentence, however, is not a contradiction. Although the first part of the sentence contains an implicature that Barb will not go to the party, implicatures are *cancellable*. This is an important property of implicatures. Consider the next example.

(3)  
   Mary went skiing and broke her arm.

This sentence implicates that she went skiing and then (as a result of a skiing accident) broke her arm. The implicature is cancellable, however.

(4)  
   Mary went skiing and broke her arm...but not in that order.

This contrasts with the following sentence.

(5)  
   Mary went skiing and then broke her arm.

This sentence asserts that she broke her arm after she went skiing. Notice, however, that it still merely implicates that she broke her arm as a result of a skiing accident. Thus, (6-a) is anomalous as it contains a contradiction; however, (6-b) is fine as it merely cancels the implicature that Mary broke her arm as a result of skiing.
(6) a. Mary went skiing and then broke her arm...but not in that order.
b. Mary went skiing and then broke her arm. She didn’t break her arm as a result of a skiing accident, however.

Let’s consider one more example in Korean. In the following pair of sentences [7-a] implicates that Minsoo washed his own hands. Example [7-b] asserts that he washed his own hands. Likewise, we can also say that [7-b] entails [7-a].

    Minsoo-NOM hand-ACC wash-PST-DECL
    ‘Minsoo washed (his) hands.’ 민수가 손을 씻었다.

    Minsoo-NOM self-GEN hand-ACC wash-PST-DECL
    ‘Minsoo washed his own hands.’ 민수가 자기의 손을 씻었다.

Recall presuppositions from the previous chapter. Recall that if A presupposes B, then B must be true for A to be felicitously uttered.

(8) The King of France is bald. \( \overline{\exists} \) France has a King.

If a presupposition is false, then the sentence that contains the presupposition suffers from a presupposition failure. Presuppositions also survive negation and questions. Thus, the following sentences still presuppose that France has a King.

(9) a. The King of France is not bald.
b. Is the King of France bald?

\[
\begin{align*}
A &\Rightarrow B = A \text{ implicates } B \\
A &\rightarrow B = A \text{ entails } B \\
A \overline{\exists} B & = A \text{ presupposes } B
\end{align*}
\]

To sum up this brief section, let’s consider the notations for the various entailments and implicatures. We label the two sentences A and B.

A: Mary went skiing and broke her arm.
B: Mary went skiing and then broke her arm.

(10) a. A \( \Rightarrow \) Mary broke her arm after she went skiing.
b. A \( \Rightarrow \) Mary broke her arm as a result of a skiing accident.
c. $B \rightarrow \text{Mary broke her arm after she went skiing.}$  
d. $B \Rightarrow \text{Mary broke her arm as a result of a skiing accident.}$

### 7.2.1 Conversational Implicature

In the conversation between John and Barb at the beginning of this chapter, Barb’s response is an example of **conversational implicature**. A conversational implicature is one that is dependent on the context of the conversation. Consider the following three possible conversations.

\[(11)\]

a. John: Are you going to Fred’s Party tonight?

b. Barb: I have to work late.

\[(12)\]

a. John: Are you doing anything today?

b. Barb: I have to work late.

\[(13)\]

a. John: What do you have to do today?

b. Barb: I have to work late.

Barb’s response in example \((11)\) has the implicature discussed at the beginning of the chapter. Her response in \((12)\) has an implicature, too; however, here her response merely implicates a yes answer to John’s question. Finally, Barb’s response in \((13)\) carries no implicature, but merely asserts the answer to John’s question. In all three cases we must evaluate against the background of the conversation whether Barb’s answer carries an implicature and, if so, what the implicature is.

### 7.2.2 Conventional Implicature

A **conventional implicature** is not dependent on the conversation, but is inherent to the lexical item or syntactic structure that gives rise to the implicature. Unlike other implicatures, conventional implicatures cannot be cancelled. As such, they are controversial, and we will not discuss them very much here. Potts (2005) argues that a defining property of conventional implicatures is that they are speaker oriented. Compare the following examples.

\[(14)\]

a. John ate a rotten apple.

b. Mary reported that John ate a rotten apple.

Here, the proposition contained in the embedded clause is part of Mary’s report. We can ascribe John’s eating of a rotten apple to Mary’s report of the event. Consider, now the following example.

\[(15)\]

a. John, the idiot, ate a rotten apple.
Example (15-a) contains an *epithet*, a (usually) negative comment by the speaker. Note that in (15-b) the feeling of the epithet is ascribed to the speaker and not to Mary. That it, the epithet is not part of Mary’s report. Thus, (15-b) contains an assertion that John ate a rotten apple and it contains a conventional implicature that John is an idiot.

### 7.2.3 Scalar Implicature

Finally we discuss *scalar implicatures*. Scalar implicatures, as the name suggests, relate to a scale of some kind. Consider the following example.

(16) John has three children.

This sentence implicates that John does not have four (or more) children. Again, this implicature is cancellable, as the following example shows.

(17) Does John have three children? [answer:] Yes, in fact he has four.

Note that if we say *John has exactly three children* or *John only has three children*, then the scalar implicature no longer holds. Consider another example now.

(18) a. Some students study hard.
   b. All students study hard.

Example (18-b) asserts that every single member of the group of contextually relevant students studies hard. Example (18-a) asserts that there is some subset of students that study hard. Consider the circles in Figure 7.1. The black circles indicate students who study hard. The white circles indicate students who don’t study hard. The statement in (18-b) is true only for the set on the left, where all the dots are black. The statement in (18-a) is certainly true for the set on the right, but is it also true for the set on the left?

![Figure 7.1: Illustration of Quantified Sets](image)

Recall that *some X* indicates that the predicate needs to be true for only a subset of X in order for the sentence containing *some X* to be true. Thus, in order for (18-a) to be true only a subset of students needs to
study hard. Here are the same sets as in the previous diagram, with the subsets identified.

![Figure 7.2: Illustration of Scalar Implicatures](image)

Observe in the right set that there is a circled subset of students who study hard. Thus, the truth conditions for (18-a) are satisfied and this sentence is true, as predicted. Note, however, the same truth conditions are satisfied for the set on the left. A subset of students who study hard is identified, thus (18-a) is also true for the set on the left. This conflicts with most people’s judgements, however. This is because of the following scalar implicature.

(19) Some students study hard ⇒ Not all students study hard.

<table>
<thead>
<tr>
<th>A ⇒ B = A implicates B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A → B = A entails B</td>
</tr>
<tr>
<td>A /\ B = A presupposes B</td>
</tr>
</tbody>
</table>

### 7.3 Speech Acts

Speakers use language to accomplish tasks, obtain information, offer information, and to affect changes. For example, if you are eating dinner and you are thirsty and the jug of water is out of reach, you might say, "Could you please pass me the water?" The purpose of uttering this is to get someone to pass the water to you. Thus, speech acts are concerned with the use of language rather than with the meaning of language (Austin, 1975). In a series of lectures in the 1950s (published posthumously in 1962, then again in 1975) Austin identified speech acts as ways of changing the world. Consider the following examples.

(20)

a. I now pronounce you man and wife.

b. I name this ship Queen Elizabeth II.

c. 메리아는 세례명을 줍니다.

By merely uttering these sentences the speaker brings about a change in the world. Austin extended this notion to virtually all utterances. He distinguished between the notions of *illocutionary force* and *perlocutionary force*. These notions were sharpened by Searle (1969) and later by Searle and Vanderveken (1985) as follows. An illocutionary act is the intention of the speaker by uttering something. A perlocutionary act is the effect the utterance has on the listener.
Let’s go through an example. John and Bill are sitting in a room together. The room is a bit chilly, and
the window is open. John says, "It’s cold in here." Then Bill gets up and closes the window. Clearly, John’s
intention is not to give a report on the relative temperature in the room, but is an indirect order to Bill to close
the window. Thus, the illocutionary act here is an order to shut the window and the perlocutionary act is Bill’s
decision to shut the window.

Certain verbs name illocutionary acts and perlocutionary acts. Consider the following examples.

(21) a. John urged Bill to shut the window.
    b. John persuaded Bill to shut the window.

The first sentence describes John’s intention and manner of bringing about the act of shutting the window. The
verb urge here describes the illocutionary act. The second sentence describes the psychological effect on Bill.
This is the perlocutionary act.

Illocutionary acts can be introduced by higher predicates; however, these are usually optional.

(22) a. I order you to shut the window. = Shut the window!
    b. I promise you to buy lunch tomorrow. = I’ll buy lunch tomorrow.

An illocutionary act is the intended act of the speaker
A perlocutionary act is the effect of an utterance on the listener.

7.4 Gricean Maxims

Although speakers often engage in thought, speakers also engage in conversation. Paul Grice was interested in
how speakers negotiate the exchange of information in conversation. Assuming speakers engage in cooperative
conversation (which is not always the case), Grice (1975) proposed the following conversational maxims.

1. The maxim of quantity, where one tries to be as informative as one possibly can, and gives as much
    information as is needed, and no more.

2. The maxim of quality, where one tries to be truthful, and does not give information that is false or that
    is not supported by evidence.

3. The maxim of relation, where one tries to be relevant, and says things that are pertinent to the discussion.

4. The maxim of manner, when one tries to be as clear, as brief, and as orderly as one can in what one
    says, and where one avoids obscurity and ambiguity.

The maxim of quantity states that, when being cooperative, speakers do not include more information than
is needed or less information than is required. Consider the following snippets of conversation.
A: How many apples and oranges should I get at the supermarket?
B: four

B’s answer violates the maxim of quantity because it’s unclear whether B intends four apples and four oranges or some combination of apples and oranges that totals four (say two apples and two oranges). Consider the following conversation. Here, speaker B offers too much information, and includes too many precise details.

A: Where did you get that orange juice?
B: At the CU 84 meters southwest of exit 6 of Gongdeok station. There’s a GS25 beside it and a 7-11 across the street, but I didn’t get my orange juice at either of them.

The Maxim of Quality states that responses are truthful and helpful to the conversation. Responses also not given for which evidence is lacking. Again, we stress that we are talking about the properties of cooperative conversation. Lying and cheating are the subject of more in depth studies. Consider the following examples.

A: Have you seen my keys?
B: They’re in the kitchen.

Speaker B clearly follows the maxim of quality, assuming B knows that A’s keys actually are in the kitchen.

A: Have you seen my keys?
B: You left them at work. [B has no idea where A’s keys actually are.]

B’s answer violates the maxim of quality because B lacks the evidence for claiming that A’s keys are at work.

The maxim of relation simply says to be relevant. Consider the following conversation.

A: Have you seen my box of chocolates?
B: They’re in your room.

Again, this conversation obeys the Maxim of Relation as B’s response is related to A’s question. Consider the following deviant conversation.

A: Have you seen my box of chocolates?
B: Chocolate is bad for your health.

Here, B’s response isn’t helpful to A’s question and violates the Maxim of Relation.

Finally, the maxim of manner states that the speaker should avoid obscurity of expression, avoid ambiguity, be brief (avoid unnecessary prolixity), and be orderly. Here is an example from [Levinson 1983:100].

A: I hear you went to the opera last night; how was the lead singer?
B: The singer produced a series of sounds corresponding closely to the score of an aria from ’Rigoletto’.

It is quite common for speakers to flout Gricean maxims by using implicature. Let’s consider the brief conversation from the beginning of this chapter.
John: Are you going to Fred’s Party tonight?
Barb: I have to work late.

Here, Barb’s answer appears to violate the Maxim of Relation as her answer has nothing to do with John’s question. However, her answer contains an implicature that she will not attend Fred’s party. Thus, the implicature contained in her answer satisfies the Maxim of Relation, while the sentence meaning of her answer flouts it. Consider another example.

A: Do you know where Fred is?
B: He’s in the kitchen or the living room.

Here, B’s answer contains a disjunction that flouts the maxim of manner. Again, there is an implicature in B’s answers that B does not know where John is for sure; however, B knows he’s either in the kitchen or in the living room.

As a final example, consider the following conversation.

A: 철수가 잘 생겼니?
B: 철수는 키가 크지.

(A: Is Cheolsoo attractive?)
(B: Well, he’s tall.)

Here, B flouts the Maxim of Relation to implicate that Cheolsoo is not attractive.
Exercises

**Question 1.** Practice with entailment: For each pair of sentences below determine whether the first sentence entails the second sentence.

A1: Every student has finished the homework.
B1: Every Vietnamese student has finished the homework.

A2: Bob likes the writers or the actors.
B2: Bob likes the writers.

A3: Everyone who got at least three A’s may skip the next assignment.
B3: Everyone who got at least five A’s may skip the next assignment.

A4: John read some of the books.
B4: John read some of the interesting books.

**Question 2.** Practice with presupposition: For each pair of sentences below determine whether the first sentence presupposes the second sentence.

A1: John realizes that Amy is a ghost.
B1: Amy is a ghost.

A2: Fred is getting married tomorrow.
B2: Fred is not married.

A3: It bothers John that the bus arrived late.
B3: The bus arrived late.

A4: John read some of the books.
B4: John read some of the interesting books.

**Question 3.** Practice with implicature: For each pair of sentences below determine whether the first sentence implicates the second sentence.
A1: John has two cats.
B1: John does not have three cats.

A2: Some of the students are reading.
B2: Not all of the students are reading.

B3: Younghui lives in Korea.

A4: Makiko went back to Japan.
B4: Makiko has been in Japan before.

**Question 4.** What is the relation (presupposition, entailment, implicature), if any, between the following pairs of sentences?

A1: Cassidy inherited a farm.
B1: Cassidy owned a farm.

A2: Cassidy inherited a farm.
B2: Cassidy owns a farm.

A3: Max quit smoking.
B3: Max does not smoke now.

A4: Max quit smoking.
B4: Max used to smoke.

A5: It’s sunny today.
B5: It’s warm today.

A6: Maria and Alberto are married.
B6: Maria and Alberto are married to each other.

A7: Maria is an Italian violinist.
B7: Some Italian is a violinist.
A8: Maria is an Italian violinist.
B8: Maria is Italian.

A9: Oscar and Jenny are middle-aged.
B9: Jenny is middle-aged.

A10: Mary used to swim a kilometer daily.
B10: Mary swims less than one kilometer daily.

A11: After Hans painted the walls, Greta installed the cabinets.
B11: Hans painted the walls.

A12: John believes that pigs do not have wings.
B12: Pigs do not have wings.

A13: John knows that pigs do not have wings.
B13: Pigs do not have wings.

A14: If Susan discovers that Bill is in New York he will get angry.
B14: Bill is in New York.

A15: Don is at home or at work.
B15: Don is at home.

A16: Don is at home or at work.
B16: The speaker of A16 does not know whether Don is at home or at work.
Further Reading

- [Huang (2015)](#) - a review article covering current views on Gricean pragmatics
- [Huang (2014)](#) - an up-to-date textbook on pragmatics, the first edition of which has been translated into Korean
- [Potts (2015)](#) - a recent review paper on presupposition and implicature
- online resources:
  - The Stanford Encyclopedia of Philosophy: [https://plato.stanford.edu/](https://plato.stanford.edu/)
Chapter 8

Sociolinguistics

By the end of this chapter you should:

- understand the kinds of social variables that influence language
- understand the basic principles of dialectology, including the concepts merger, isogloss, and isogloss bundle
- be able to explain social dialects, including the concept of prestige dialect
- be familiar with and understand the terms ‘linguistic variable’ and ‘social variable’
- be familiar with the interplay between gender and language
- have a basic understanding of contact languages, including the concepts of pidgin and creole

8.1 Introduction

Sociolinguistics is the study of social aspects of language and the interaction between language and society. Recall from the introduction that all individual languages naturally contain variation. A common misconception about language is that non-standard dialects and languages lack grammatical rules, or are even degenerate or corrupt. All language forms, including sign languages, non-standard dialects and non-standard languages have the same kind of grammar rules as standard languages. They simply have different rules. Indeed, the last several decades of sociolinguistic research have shown conclusively that variation is rule-governed. In this chapter, we will look at how languages vary and change within speech communities, and the ways researchers have analyzed and understood these properties.

Let’s start with a well known example. One of the most studied points of variation in natural language is the pronunciation of the morpheme -ing in English. In standard English, the word sleeping is pronounced [slipɪŋ], but you may be familiar with a different pronunciation, [slipɪn]. The latter pronunciation is often considered more informal. Prescriptivists call this pronunciation ‘sloppy’ or ‘lazy’. Descriptively, however, we say these two pronunciations are simply different. Furthermore, part of linguistic competence is using language appropriately in different contexts. We do not wear a three-piece suit to a soccer game. We dress appropriately
for different contexts. Likewise, we adjust (largely unconsciously) how we speak in different contexts. We speak differently at a job interview than we do at a bar with our friends.

Although people often speak to themselves and think to themselves, people often speak to each other, too. When speakers communicate in a society, they make certain choices, either consciously or subconsciously, about words, pronunciation, and grammatical constructions. Different speech situations require us to make different choices. **Register** refers to the set of lexical choices we make in a given speech situation, and **style** refers to the set of grammatical choices we make.

### 8.2 Dialectology

Dialectology (方言학) is the branch of sociolinguistics that studies dialects. Questions include how dialects form, how dialects differ from each other, what effect contact has on dialects, and to what extent different dialects are mutually intelligible. This is not an exhaustive list of questions dialectologists ask, but it is representative of some of the main questions. Recall from Chapter 1 that a dialect is a mutually intelligible speech form and a dialect continuum is a set of dialects in which each member is mutually intelligible with at least one other member. Remember that two members of a dialect continuum that are located far apart geographically may not be mutually intelligible.

Regional dialects are based on how language change over geographic areas. Differences in dialects have been noticed in Europe since the 12th century. Undoubtedly, people noticed dialect differences earlier than this, but this is the earliest historical evidence of interest in dialectology. The first dialect map of the United States was made in the 1930s.

One common way in which dialects vary is due to **mergers**. In many varieties of English /ɔ/ and /ə/ are separate phonemes. A minimal pair is found with the words cot /kɒt/ and caught /kɔt/. In some varieties of English in North America and in the UK, though, these two sounds have merged so that they both have the form /kɒt/. We say that the vowels /ɔ/ and /ə/ have merged in these varieties of English. This particular merger is known as the **cot-caught merger**. Figure 8.1 shows the distribution of the cot-caught merger in the United States.

Another common merger is the **pin-pen merger**. In this merger, the sounds /ɪ/ and /ɛ/ have merged into [ɪ] before nasal consonants. Thus, while pin and pen have both merged to /pɪn, pit and pet remain distinct. This merger is found almost entirely within the United States. Figure 8.2 shows the distribution of the pin-pen merger in the United States.

In Korean, the เอ/애 merger is complete in virtually all varieties of Korean in South Korea, as discussed in Chapter 4. Thus 개 (‘dog’) and 게 (‘crab’) are identical for all but a small number of older speakers. In the South Kyengsang dialect of Korean the vowels 오 and 어 have merged to 오 [ɑ]. A major dialect boundary in Korean runs along the Taebaek mountain range. It separates the western dialects from the eastern dialects. The eastern dialects are typically pitch-accent dialects, meaning lexical differences can depend on pitch. Here is an
Figure 8.1: Cot-Caught Merger

Map 1. The Merger of /ɔt/ and /ɔt/: Contrast in production of /ɔt/ and /ɔt/ before /n/ in COT vs. CAUGHT.

Figure 8.2: Pin-Pen Merger

Map 2. The PIN/PEN merger: Contrast in speech production of /ʌn/ and /ən/ before nasals in PIN and PEN, HIM and HEM.
example from the Kyengsang dialect. Recall that an acute accent over a vowel (á) refers to high pitch and a grave accent over a vowel (à) refers to low pitch.

(1) Pitch-Accent in Kyengsang

a. 가지 - kácí ‘branch’

b. 가지 - kácì ‘type’

c. 가지 - kàcí ‘eggplant’

Another major distinction in Korean dialects concerns length. As mentioned in Chapter 3, the length distinction has disappeared in the Seoul dialect of Korean. Very few speakers, if any, distinguish 눈 ‘snow’ from 눈 ‘eye’. In the Jeolla (Chonnam) dialect, however, the distinction is still found. In the following examples the orthographic accents refer to the pitch of the vowel. Notice that the long vowel contains two pitch marks.

(2) Length Contrast in Jeolla Korean

a. 강도 ‘thief’ [k`a´aNd`o]

b. 간낭콩 ‘bean’ [k`aNn´aNk`oN]

In Seoul Korean, the underlying pattern for each syllable in a phrase is either LHLH or HHLH, depending on the initial consonant in the first syllable (Jun, 1998). Thus, 감사합니다 [kàmsámnidá] is realised with low pitch on the first syllable, high pitch on the second syllable, and so forth. In Jeolla dialect, the pattern is LHL or HHL, depending on the initial consonant in the first syllable. In Jeolla Korean, however, a short syllable gets one tone and a long syllable gets two tones. Compare the examples from Jeolla Korean in (2).

Finally, in Cantonese, /n/ and /l/ in onset position have merged to /l/ for most speakers. This merger started in the 1940s and is still ongoing. Thus, 男 (naam4, 남, ‘man’), 藍 (laam4, 람/남, ‘blue’) are both laam4 for speakers who have undergone this merger. Some older speakers still distinguish between /n/ and /l/, however (Zee and 徐云揚, 1999).

An isogloss is a line dividing a particular linguistic feature on a map. We can say, then, that the Taebaek mountain range serves as a rough isogloss separating pitch accent varieties of Korean (east of the Taebaek mountain range) from non-pitch accent varieties. Consider the following example from Ojibwe (an Algonquian language spoken in Canada and the northern US). The red isogloss in Figure 8.4 divides the dialects that use migi for ‘bark’ (the sound of a dog) from those that use migine.

When several isoglosses all fall roughly along the same path it is said to be an isogloss bundle and marks a major dialect boundary. Consider Figure 8.5, which shows an isogloss bundle for several features of Ojibwe. Do not worry about the morphological terminology. Observe that the bundle is tight in the eastern extremity, but fans out somewhat in the western extremity. Fans in an isogloss bundle are common.

As seen in the Ojibwe data above, dialects can also be identified on the basis of lexical differences. There are various names for sweet carbonated soft drinks that correlate with regional variation: pop, soda, soda pop, coke, etc. Casual shoes one uses for walking or light sports are called either runners, running shoes, or sneakers.
Figure 8.3: Map of Korean Dialects

Figure 8.4: Isogloss for the Ojibwe word for ‘bark’
In Korean a drawer can be 서랍 or 뜬깡. Chives are either 부추 or 정구드.

Regional dialect variation has often studied phonological and lexical differences among dialects. There are syntactic and morphological differences among dialects, too. Consider the following two questions in the plain style in Seoul Korean and their equivalents in Busan Korean (Lee 2017).

(3) Interrogative Morphology in Korean Dialects
a. pap mek-ess-e?  (밥 먹었어요?)  ‘Have you eaten?’  [Seoul Korean]
b. mwe mek-ess-e?  (뭐 먹었어요?)  ‘What did you eat?’
c. pap mwuk-ess-na?  (밥 묵었노?)  ‘Have you eaten?’  [Busan Korean]
d. mwe mwuk-ess-no  (뭐 묵었노?)  ‘What did you eat?’

Here is an example of a syntactic difference between North American English and British English. The next example involves ellipsis of parts of the verb phrase. In chapter 5 we saw that do so acted as a VP proform. In the following examples, parts of the VP have been elided; however, the pattern differs between the two dialects of English. Consider the following data.
(4) Ellipsis in North American English
   a. I could have done it.  [North American English]
   b. *I could have done
   c. I could have

(5) Ellipsis in British English
   a. I could have done it.  [British English]
   b. I could have done
   c. *I could have

In North American English the entire VP *done it* can be elided. In British English, however, *done* is used as a proform to replace the VP.

Finally, we mention some syntactic properties of North American English dialects. The first construction we examine is the *multiple modal construction* (Battistella [1995]), which is found in a number of varieties of English in the southern United States. Multiple modal constructions are sentences with more than one modal verb. Consider the following examples (Di Paolo [1989]).

(6) a. I don’t think I have any grants you might could apply for.
   b. We might can go up there next Saturday.
   c. You know, if you drank a half a drink, you might oughta go home and sleep it off.

Another construction found in North American dialects include a-prefixing (Wolfram [1988, 1976]). This construction is also found widely in the southern United States.

(7) a. He was a-huntin’ (= ‘He was hunting.’)
   b. He’s a-gonna try it.

This form can be used only in the progressive. It cannot be used with gerunds.

(8) a. *He likes a-huntin’.
   b. *A-huntin’ is fun.

Finally, the so-called *done my homework* construction is found widely in Canada, and in some pockets of the northern United States. The construction involves an inflected form of *be* followed *done, finished,* or sometimes *started* and a noun phrase. Here are some examples (Fruehwald and Myler [2013, 2015]).

(9) a. I am done my homework.
   b. I’m finished my homework.
   c. Don’t talk to me until I’m done my coffee.
Social Dialects are not associated with a particular region, but instead are associated with socio-economic status, race, caste, gender, or sexual orientation. When such a particular speech form forms a recognizable dialect, it is referred to as a social dialect. Within discussions of dialectology the notion of a prestige dialect plays a role in how various dialects are perceived. The dialect which is perceived by speakers as dominant or superior – usually correlated with higher socioeconomic status, higher caste, or dialect of large, important cities.

We will consider one well studied social dialect in North America - African American English, often abbreviated as AAE. Historically AAE exhibited a great deal of negative prestige due to rampant racism. The origin of AAE is still under dispute, but it is thought to have evolved from older varieties of English and various African languages.

Phonological properties of AAE include r-deletion (thus, caught and court have same phonetic form), consonant cluster reduction: passed /pæst/ becomes [pæs]), and the pin/pen merger. There are, of course, many other phonological characteristics of AAE, but we mention just a few here. Consider the following examples of consonant cluster reduction.

(10) Consonant Cluster Reduction in AAE (Green, 2002, 109)

<table>
<thead>
<tr>
<th>orthographic form</th>
<th>underlying form</th>
<th>surface form</th>
</tr>
</thead>
<tbody>
<tr>
<td>test</td>
<td>/trst/</td>
<td>[tr̪s]</td>
</tr>
<tr>
<td>wasp</td>
<td>/wɔsp/</td>
<td>[wɔs]</td>
</tr>
<tr>
<td>mask</td>
<td>/mæsk/</td>
<td>[mæs]</td>
</tr>
<tr>
<td>gift</td>
<td>/gift/</td>
<td>[gif]</td>
</tr>
<tr>
<td>adopt</td>
<td>/ɔdapt/</td>
<td>[ədapt]</td>
</tr>
<tr>
<td>conduct</td>
<td>/kand2kt/</td>
<td>[kandək]</td>
</tr>
<tr>
<td>band</td>
<td>/bænd/</td>
<td>[bæn]</td>
</tr>
<tr>
<td>bold</td>
<td>/bold/</td>
<td>[bol]</td>
</tr>
</tbody>
</table>

Evidence that the underlying form of these words contains two consonants in coda position is found in how they interact with the following word. If the following word begins with a vowel, then both consonants survive. For example test out surfaces as [tr̪stəʊt].

Syntactic properties of AAE include negative concord (which is found in Old-Middle English and in contemporary Romance), the lack of the copula, habitual be, and negative existentials. Let’s consider negative concord first. Standard English only has logical double negatives. If there are two negative items in a sentence, they cancel each other. Korean is the same. Consider the following examples.

(11) Logical Double Negatives
a. You’re not not talking to Bart. (= You are talking to Bart.)
b. Mary didn’t read NO book. (= Mary read at least some books. - Heavy stress on ‘NO’)
c. 사과를 안 먹지 않았어요.
Note that negative concord is found in varieties of English other than AAE, but we present AAE data here. Observe that the multiple instances of negation here do not logically cancel each other out.

(12) Negative Concord in AAE \cite{Green2002, 77}

a. If you don’t do nothing but farmwork, your social security don’t be nothing. (= ‘If you only do farm work, then your social security isn’t usually very much.’)

b. I ain’t never seen nobody preach under announcements. (= ‘I have never seen anyone preach while they’re giving announcements.’)

Predicate adjectives and nouns typically require a \textit{copula} to form a full sentence in Standard English and in Korean. \textit{I am a student}. 나는 \textit{학생}이에요. A copula links the subject to a predicate noun or adjective. Not all languages have a copula, however. Consider the following Russian example.

(13) \begin{align*}
ja & \text{ student.} \\
\text{I.NOM student} & \text{[Russian]} \\
\text{‘I am a student.’}
\end{align*}

AAE typically has no overt copula, either. Consider the following examples \cite{Green2002, 52}.

(14) a. Some of them big and some of them small. (= ‘Some of them are big and some of them are small.’)

b. Some iMacs tangerine. (= ‘Some iMacs are tangerine.’)

Bare, uninflected \textit{be} indicates habitual aspect. Consider the following examples \cite{Green2002, 48}. Note the last example also contains perfective \textit{done}, which indicates a completed event.

(15) a. She be telling people she eight. (= ‘She is always telling people she’s eight.’)

b. Your phone bill be high, don’t it. (= ‘Your phone bill is usually high, isn’t it?’)

c. I be in my office by 7:30. (= ‘I am usually in my office by 7:30.’)

d. She gotta be there for 9, so they be done gone to school. (= ‘She has to be there at 9, so they have husally already gone to school.’)

Negative existentials are formed by inverting a negative auxiliary over the subject. Usually the subject is a negative quantifier such as \textit{nobody} or \textit{nothing}. Existentials in Standard English typically use \textit{there} to introduce the sentence. AAE uses inversion instead. Consider the following examples \cite{SellsEtAl1996, 592}.

(16) a. Can’t nobody say nothin’ to dem peoples! (= ‘There isn’t anyone who can say anything to those people.’)

b. Ain’t nobody never told me what to do. (= ‘There isn’t anyone who ever told me what to do.’)
8.3 Social and Linguistic Variation

Labov (1963) argued that variation is a natural part of human language. The field of variationist studies and sociolinguistics in general started with Labov’s seminal paper. Crucially, Labov’s observation was that speakers of a given dialect or language exhibit variation in their speech. A speaker of English may pronounce the word *running* either as [rʌŋ] or as [rəŋ]. The same speaker may vary among these options depending on a number of factors. This section will consider the main issues of variationist studies within sociolinguistics.

In a given study, the linguistic property under analysis is referred to as the *variable*. Not all properties vary, of course. A property that does not vary is referred to as a *categorical property*. Variationists study properties that do vary - a *variable property*. Factors that affect a variable can either be linguistic or social. Here are some examples of linguistic and social factors. Neither of these lists is comprehensive.

(17) Examples of Linguistic Factors
   a. position in word
   b. position in sentence
   c. phonological properties of nearby segments
   d. tense
   e. number (singular or plural)

(18) Examples of Social Factors
   a. age
   b. gender
   c. socioeconomic status
   d. race
   e. sexual orientation
   f. level of education

An early variationist study of language use involved r-deletion in New York City English (Labov 1972). Labov was interested in the social factors that affected r-deletion. He did this by going to a department store and asking a sales clerk for an item he knew to be on the forth floor. The sales clerk would answer, "fourth floor". Labov took note of whether the clerk kept or dropped the /r/ in each word. In this investigation the variable was r-deletion, and the social factors included the socio-economic class of the shop, gender of the sales clerk, and the rough age of the sales clerk (Labov simply had to guess). Additionally, Labov pretended not to hear the first response and asked the clerk to repeat the answer. Thus, Labov also kept track of r-deletion in the first and second response. In the second response it is assumed that the clerk responds more clearly. Finally, Labov also kept track of both words, ‘fourth’ and ‘floor’ to see if the rate of r-deletion differed between them. To get an idea of the process, the interaction may have gone as follows.
Figure 8.6: Results of the "Fourth Floor" Experiment

experimenter: Excuse me. Where would I find toasters?
clerk: Fourth floor.
experimenter: Sorry? [cupping hand around ear]
clerk: [in a clearer and louder voice] Forth floor.

The results of the experiment are in Figure 8.6; however, only of of the factors are reported here. The y-axis shows the rate at which /u/ was found in the relevant word. Saks is an upscale department store; Macy’s is a mid-level department store; and S. Klein was described as a working class department store. In the bar graph we see that r-deletion correlates with the socio-economic class of the department store. The higher the class the less likely r-deletion takes place. Note also that r-deletion is more frequent on the word ‘fourth’ than on the word ‘floor’.

A well-known phenomenon in Korean that is subject to variation is /w/-deletion. For example, the word 봐 /pwa/ [pa] ‘look!’ typically undergoes /w/-deletion in casual speech. The word 주권 /[i]jukwan/ [i]juk*wan] ‘sovereignty’, however, does not. It turns out that both linguistic and social factors play a role in /w/ deletion in Korean (Silva, 1991). The linguistic factors include the place of preceding consonant (labial, alveolar, postalveolar, velar, glottal), the manner of preceding consonant (plain, aspirated, fortis), and the following vowel (front, back). The social factors include socioeconomic status (upper, middle, or lower) and age (16-25, 26-45, 46+). The following charts show the rate of /w/-deletion for these various factors.
<table>
<thead>
<tr>
<th>Place</th>
<th>rate of /w/-deletion (%)</th>
<th>Manner</th>
<th>rate of /w/-deletion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>labial</td>
<td>81</td>
<td>plain</td>
<td>30</td>
</tr>
<tr>
<td>alveolar</td>
<td>23</td>
<td>aspirated</td>
<td>14</td>
</tr>
<tr>
<td>postalveolar</td>
<td>11</td>
<td>fortis</td>
<td>19</td>
</tr>
<tr>
<td>velar</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>glottal</td>
<td>12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vowel</th>
<th>rate of /w/-deletion (%)</th>
<th>Socioeconomic status</th>
<th>rate of /w/-deletion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>front</td>
<td>22</td>
<td>upper</td>
<td>22</td>
</tr>
<tr>
<td>back</td>
<td>31</td>
<td>middle</td>
<td>25</td>
</tr>
<tr>
<td>age</td>
<td></td>
<td>lower</td>
<td>32</td>
</tr>
<tr>
<td>16-25</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26-45</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46+</td>
<td>21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All three linguistic factors are statistically significant. Many languages have constraints against co-occurring labial segments, so the high rate of deletion after labial consonants is not surprising. Socioeconomic status and age are both statistically significant. The experimenters also tested gender, but it was found not to be statistically significant. Thus, variable /w/-deletion is both a social phenomenon as well as a linguistic phenomenon.

### 8.4 Language and Gender

The notion of a *genderlect*, a social dialect based on gender, can be traced back to Jespersen (1922), where he noted differences in the speech of men and women in a number of languages around the world. Many of Jespersen’s remarks on women and men’s speech can be seen as disparaging at the very least; however, Thomas (2013) presents an interesting re-interpretation of Jespersen’s original work. Regardless, Jespersen was probably the first modern scholar to draw attention to the differences between men’s and women’s speech, despite the lack of scientific rigour in his observations.

The field of language and gender took off in the 1970s with the observation that the study of language and gender is intimately tied with the study of power differences in society (Lakoff, 1975). Although Lakoff’s observations are based mainly on introspection, her work served as a starting point for subsequent research on language and gender.

Early studies such as Lakoff’s concentrated on lexical differences such as colour terms. Lakoff reported a much wider variety of colour terms in women’s speech than in men’s speech. Lakoff contrasts the following

---

1Some of Jespersen’s comments include “...the vocabulary of a woman as a rule is much less extensive than that of a man...” (Jespersen 1922: 248) and “The explanation of this characteristic feminine usage is, I think, that women much more often than men break off without finishing their sentences, because they start talking without having thought out what they are going to say.” (Jespersen 1922: 250).
example as a typical difference between men’s and women’s speech. I leave it to the reader to guess which is more typical of which gender’s speech.

(19)  a. Oh dear, you’ve put the peanut butter in the refrigerator again.
     b. Shit, you’ve put the peanut butter in the refrigerator again.

Returning to the issue of the relationship between gender and power, Lakoff observed that women are more likely to use tag questions in conversation in order eliciting feedback on their opinions. Consider the following examples.

(20)  a. It’s hot in here.
     b. It’s hot in here, isn’t it?
     c. I have a headache, don’t I?

The first sentence merely offers the speaker’s opinion on the temperature in the room. The second sentence invites the listener’s feedback on the proposition. The reason the last sentence sounds odd is because the listener cannot have feedback on the internal state of the speaker. Thus, tag questions often test to see of the listener is in agreement with the speaker. Lakoff argued this is representative of a power difference between speakers.

There is a common myth that women talk more than men. Many proverbs to this effect are found in many languages. Research shows, however, that men typically dominate conversations in mixed gender environments (Onnela et al., 2014). See Kaplan (2016) and Holmes (1998) for more details.

We saw above that many differences in men’s and women’s speech in English is variable. Women tend to use a wider array of colour terms. Women tend to use tag questions more frequently. Some languages encode gender distinctions grammatically. Thai, for example, has two sentence-final politeness markers: khrap is used by men; kha is used by women. The difference here is not variable, but is categorical.

Later studies in language and gender considered the role of sexual orientation in language. Careful phonetic experiments revealed that the differences between gay and straight men’s speech is not as pronounced as stereotypes suggest (Rogers, 2014). Similar results have also been shown for Cantonese (Chi-Kuk, 2007), Puerto Rican Spanish (Mack, 2010), and for lesbian speech in Japanese (Camp, 2009) and Cantonese (Wong and Tu, 2019). The role of sexual orientation has also been studied in second language acquisition in Korean (King, 2008).

8.5 Contact Languages

Languages do not exist in a vacuum. With extremely few exceptions virtually every language around the world has come into contact with other languages. Korean contains numerous words borrowed from Chinese. English contains numerous words borrowed from French. The grammar of language can be influenced by neighbouring languages, too. Thai and Vietnamese have tonal systems similar to those found in Chinese languages. This
similarity is due to contact. Recall the discussion on *sprachbunds* from chapter 1. This section deals with new languages that arise from contact situations. In a language contact situation a *lingua franca* may be used. A lingua franca is a common language used among people of diverse linguistic backgrounds. In many tourist sites English is a lingua franca. In many parts of Africa French is a lingua franca. In Bantu speaking countries in Africa Swahili is often used as a lingua franca.

Throughout human history migrations (forced or voluntary) have created situations in which no common language exists among a group of people. In many cases such situations have arisen in the context of slave trades, where people who speak a variety of languages are forced into an environment where communication is necessary. Invasions have also given rise to environments where people with diverse linguistic backgrounds must communicate. In such environments there is no lingua franca to rely on. People living in such conditions first end up acquiring a *jargon*—a specialized set of vocabulary needed to perform work and other required functions. A *pidgin* is a makeshift speech variety that arises from continued use of jargon. Pidgins are highly variable and ephemeral. They are used only as needed. A *creole* is a language that arises in the environment of a pidgin. Children raised in the environment of a pidgin begin to formulate a consistent grammar. After a few generations, a stable creole is formed. This basic life cycle was formulated by Hall (1962).

In such instances of contact, we distinguish the *superstrate language*, the language of dominant group, from the *substrate language*, the language of non-dominant group. Typically the superstrate language is also the lexifier language, the language that serves as the source of the words in the pidgin and subsequent creole. The grammatical patterns of the creole typically originate from the substrate languages.

Hawaiian Pidgin, despite its name, is a creole. The superstrate language is English, and the substrate languages include Japanese, Tagalog, Cantonese, Korean, Portuguese, and Hawaiian. Here is an example.

(21)  Da watah stay cold. (= ‘The water is cold.’)

Tok Pisin is a creole spoken in Papua New Guinea, where it is an official language. It, along with Haitian Creole, are the only two creoles that have the status of an official language in the world. The superstrate language is mostly English, and the substrate languages are Kuanua and other Austronesian languages. The following chart shows the pronoun system of Tok Pisin. Try to discern the English sources of these pronouns.

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Dual</th>
<th>Trial</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.EXCL</td>
<td>mi</td>
<td>mitupela</td>
<td>mitripela</td>
<td>mipela</td>
</tr>
<tr>
<td>1.INCL</td>
<td>yumitupela</td>
<td>yumitripela</td>
<td>yumpela or yumi</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>yu</td>
<td>yutupela</td>
<td>yutripela</td>
<td>yumpela</td>
</tr>
<tr>
<td>3</td>
<td>em</td>
<td>tupela</td>
<td>tripela</td>
<td>ol</td>
</tr>
</tbody>
</table>
Exercises

**Question 1.** Think of ten examples of English or Korean lexical items that are representative of regional dialects.

**Question 2.** Consider the following dialectal data from the Normandy region of France (adapted partly from Hall 2005). Draw an isogloss for the /k/ /ʃ/ distinction. (map courtesy of d-maps.com)

<table>
<thead>
<tr>
<th>gloss</th>
<th>Rouen (1)</th>
<th>Caen (2)</th>
<th>L’Aigle (3)</th>
<th>Vire (4)</th>
<th>Domfront (5)</th>
<th>Cherbourg-Octeville (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>ka</td>
<td>ka</td>
<td>ŋa</td>
<td>ŋa</td>
<td>ŋa</td>
<td>ŋa</td>
</tr>
<tr>
<td>cauldron</td>
<td>kodrō</td>
<td>kodrō</td>
<td>jodrō</td>
<td>jodrō</td>
<td>jodrō</td>
<td>kodrō</td>
</tr>
<tr>
<td>dog</td>
<td>kiē</td>
<td>kiē</td>
<td>jie</td>
<td>jiē</td>
<td>jiē</td>
<td>kiē</td>
</tr>
<tr>
<td>field</td>
<td>kā</td>
<td>kā</td>
<td>jā</td>
<td>kā</td>
<td>jā</td>
<td>kā</td>
</tr>
</tbody>
</table>

**Question 3.** In Cantonese, there is variation in the pronunciation of /n/ and /l/ in onset position. For some
speakers. 藍 (blue) is pronounced [lam4] and 男 (man) is pronounced [nam4] (numbers indicate tones). For other speakers, both items are pronounced [lam4]. The following data indicate the percentage of speakers who pronounce both items [lam4]. Looking at gender, women are more likely to pronounce both items as [lam4].

Consider now the data on age. Draw a chart that shows the incidence of merger of /n/ and /l/ in Cantonese with respect to age and gender.

<table>
<thead>
<tr>
<th>gender</th>
<th>age</th>
<th>1-18</th>
<th>19-30</th>
<th>31-45</th>
<th>46+</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>61%</td>
<td>85%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>72%</td>
<td>70%</td>
<td>67%</td>
<td></td>
<td>28%</td>
</tr>
</tbody>
</table>
Further Reading

- Shinn (2009) outlines several main investigations on several aspects of Korean sociolinguistics.
- Tagliamonte (2006) is an excellent, in depth resource on variationist sociolinguistics.
- Online Resources:
  - Yale Grammatical Diversity Project: English in North America - https://ygdp.yale.edu/
Chapter 9

Historical Linguistics

By the end of this chapter you should:

- be familiar with the underpinnings of historical linguistics
- understand how languages tend to change over time
- understand how languages and dialects are related in terms of historical relationships
- be familiar with Comparative Reconstruction techniques
- be familiar with the major language families of the world

9.1 Introduction

Historical linguistics tries to understand how languages change over time. A major branch of historical linguistics, comparative linguistics, tries to understand how languages and dialects are related historically. In this chapter we will examine how languages can change over time and how we can use this understanding of language change to determine how languages are related. We will end with a discussion on some of the major language families of the world. One aspect of historical linguistics that is of particular interest to lexicographers is etymology (어원학). Etymology is the study of the origins of words, including how the meaning of a word changes over time.

It is a well known property of languages that they change over time. Recall McWhorter’s lava lamp analogy from the introduction. McWhorter emphasized the slowly but constantly changing patterns of language. Each change introduces a small difference in the language. Over time, enough changes accumulate such that sufficiently older forms of a given language are less recognizable to the modern speaker. If we go back far enough in time, the language is completely unrecognizable.

Languages and dialects change independently over time. Imagine a hypothetical society that speaks a common language. This is depicted as stage A in Figure 9.1. If the society separates and starts to lose contact, over time the speech varieties will start to diverge into two separate dialects, D₁ and D₂ in stage B. At this stage, the two dialects are still mutually intelligible. The two dialects will continue to diverge until they are now no
longer mutually intelligible and are two separate languages, L₁ and L₂ in stage C. This process continues over centuries forming a large family of related languages. A few such families are discussed at the end of this chapter.

![Diagram showing the formation of distinct languages from one source language.](image)

Figure 9.1: Formation of Distinct Languages from one Source Language

English and Chinese, for example, have long written traditions allowing us to directly observe older forms of the languages. To see an illustration of how language changes over time consider the following English data from various time periods.

**Old English: 450CE – 1100CE:**

*And ic cyðe eow, þæt ic wylle beon hold hlaford and unswicende to godes gerihtum and to rihtre woroldlage.* [royal proclamation]

[And I make known to you, that I will be a civilized lord faithful to God’s rights and to the worldly laws.]

**Middle English: 1100 – 1500:**

*Whan that Aprille with his shoures soote...*

[When April with its sweet showers...]

**Early Modern English: 1500 – 1700:**

*Wherefore art thou Romeo? Deny thy father and refuse thy name...*

[Why are you (named) Romeo? Forget your father and refuse your name...]

It is important to keep in mind the difference between *synchronic change* and *diachronic change*. Synchronic sound changes are those that are active in the language, while diachronic sound changes are those that have taken place over a longer period of time, resulting in a change in the language. In chapter 4 we saw many examples of synchronic sound changes. Recall the synchronic sound change in Korean in Figure 9.2. In this chapter we will deal with changes in language that happen over time. That is, we will deal with diachronic changes.
Modern historical linguistics started with Sir William Jones and Gaston-Laurent Coeurdoux in the late 18th century. Both scholars proposed that the languages of India (including Hindi, Gujarati, Marathi, Punjabi, etc.) were related to the languages of Europe. Note that the languages of southern India, the Dravidian languages, are not related to the former. We will discuss these below. Careful study of Sanskrit, Latin and Greek on the part of Jones and Coeurdoux led both scholars to propose the Indo-European family. Before this time, the label “European Languages” was simply used. Sir William Jones later hypothesized that Persian was also related to the Indo-European languages, a hypothesis which turned out to be correct.

A diachronic sound change takes place over time. One of the most well studied diachronic changes is the Great Vowel Shift (GVS) in English. It took place in the 15th and 16th centuries and affected long vowels only. Long vowels in Middle English correspond quite closely to tense vowels in Modern English. The Great Vowel Shift is shown in Figure 9.3.

Trisyllabic Laxing was once a productive process in the history of English that still has effects in Modern English. Trisyllabic laxing took place before the Great Vowel Shift. It involves a long vowel or a diphthong becoming short (or a monophthong) in words of three syllables or more if the first syllable is unstressed. By analogy it has been found on shorter words, too. Trisyllabic laxing interacts with the GVS to produce substantial differences in the vowel qualities of related words in Modern English. Consider serene /særən/ and serenity /særənəti/. The second vowel in each word differs by a number of features. The difference is even more striking in the words divine /dɪvɛɪn/ and divinity /dɪvɪnəti/. In Figure 9.4 the double line refers to the change due to trisyllabic laxing. The straight arrow refers to the change due to the GVS. The dashed line refers to the lack of change since the GVS affected only long vowels. Observing the first pair of words, serene and serenity, we see that before the GVS the second vowel of serene is /ɛ:/ and the second vowel of serenity is /ɛ/. This pair of
vowels differs by only one feature. Specifically /ɛ/ is the lax version of /e:/.

The effects of the GVS cloud this correspondence.

```
<table>
<thead>
<tr>
<th>Before GVS</th>
<th>After GVS</th>
</tr>
</thead>
<tbody>
<tr>
<td>serene</td>
<td>serene</td>
</tr>
<tr>
<td>[ɛː] → [iː]</td>
<td></td>
</tr>
<tr>
<td>serenity</td>
<td>serenity</td>
</tr>
<tr>
<td>[ɛ] → [ɛ]</td>
<td></td>
</tr>
<tr>
<td>divine</td>
<td>divine</td>
</tr>
<tr>
<td>[iː] → [aI]</td>
<td></td>
</tr>
<tr>
<td>divinity</td>
<td>divinity</td>
</tr>
<tr>
<td>[ɪ] → [ɪ]</td>
<td></td>
</tr>
</tbody>
</table>
```

Figure 9.4: Trisyllabic Laxing in English

Other pairs of words affected by trisyllabic laxing include the following. Note that the lax form of Middle English /uː/, /oː/, became /ʌ/ in Modern English. Likewise the lax form of /ɔː/, /ɑː/, became /ʌ/ or /ɑ/, depending on the dialect.

(1) a. profane-profanity
    b. derive-derivative
    c. profound-profundity
    d. school-scholarly

Note that obese-obesity is an interesting exception. The word obese was back-formed from obesity in the 19th century.

Practice 9.1: Transcribe the words in (1). Give the hypothetical transcriptions for the pronunciation before the Great Vowel Shift as depicted in Figure 9.3.

Very little is known of Old Korean (before 1000 CE). The Korean peninsula was not as linguistically homogenous as it is today. Old Korean is the name given to the language spoken in the Shilla Kingdom; however, other related languages were spoken in the areas occupied by Koreans at the time, which reached into modern day China (Song, 2005).

Let’s consider some linguistic properties of Old Korean (Sohn, 1999). The following word list provides some words that are know from Old Korean.
Old Korean word list

a. woraka (used by ruling class)/kici (used by commoners) – ‘king’
b. kuti – ‘falcon’
c. tohel – ‘field’

The phonology of Old Korean was quite different from that of Modern Korean. The precise details of Old Korean are still not well understood, but [Whitman (2015)] gives a clear overview. Here are a few details we do know. The symbol ‘>’ indicates that the form to the right is historically derived from the form to the left. It is thought that consonant clusters were possible (Old Korean [niskOm] ‘king’ > Modern Korean [(n)imkum] (임금). Is it also thought that there was a distinction between /l/ and /r/ in Old Korean and that /r/ was sometimes dropped word-internally (Old Korean [k@ru] > [k@ju] ‘goose’ > Modern Korean [k@qi] ( 거위 ) (Sohn, 1999, 41-43). It is also thought that aspirated consonants may not have existed in Old Korean. Modern Korean [khW] (크 ‘big’) was transcribed as disyllabic huku- in Early Middle Korean. It is thought that the aspirated consonants may have arisen from the sequence *hVC followed by vowel deletion.

The following word list from Middle Korean (approx 1000 – 1500 CE) shows some similarity between Middle Korean and Modern Korean.

Middle Korean word list

<table>
<thead>
<tr>
<th>Middle</th>
<th>Modern</th>
<th>한글</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>namo</td>
<td>namu</td>
<td>나무</td>
<td>‘tree’</td>
</tr>
<tr>
<td>kamakoj</td>
<td>k*amakqi</td>
<td>까마귀</td>
<td>‘crow’</td>
</tr>
<tr>
<td>talA</td>
<td>tali</td>
<td>다르</td>
<td>different</td>
</tr>
<tr>
<td>skum-il</td>
<td>k*um-il</td>
<td>꿈을</td>
<td>dream-ACC</td>
</tr>
<tr>
<td>muzwu</td>
<td>mu</td>
<td>무</td>
<td>turnip</td>
</tr>
</tbody>
</table>

The fortis consonants (ㄲㄸㅃㅆㅉ) appeared during the time of Middle Korean. It was also at this time that the phonemes /z/ and /β/ appeared in Middle Korean (which subsequently disappeared). ‘Turnip’, for example, changed from Middle Korean [muzwu] to the following forms in various current dialects of Korean: musu (전라, 충청, 강원도, 함경남도) musi (전라, 경상남도) mu (충청남도, 강원도, 경기도, 황해, 평안북도, 함경남도) mui (강원도, 황해) mu.yu (경기도, 황해). Here are some other examples (Sohn [1999] 44).
(4) Middle Korean voiced fricatives

<table>
<thead>
<tr>
<th>Middle Korean</th>
<th>Modern Korean</th>
<th>( \text{‘autumn’} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>kọζol</td>
<td>kautl 가을</td>
<td>( \text{‘autumn’} )</td>
</tr>
<tr>
<td>mọζom</td>
<td>maum 마음</td>
<td>( \text{‘mind’} )</td>
</tr>
<tr>
<td>azọ</td>
<td>au 아우</td>
<td>( \text{‘younger brother’} )</td>
</tr>
<tr>
<td>nuβi</td>
<td>nui 누이</td>
<td>( \text{‘(younger) sister’} )</td>
</tr>
<tr>
<td>suβọl</td>
<td>sul 술</td>
<td>( \text{‘alcohol’} )</td>
</tr>
<tr>
<td>tuβọl</td>
<td>tul 두</td>
<td>( \text{‘two’} )</td>
</tr>
</tbody>
</table>

9.2 Words over Time

New words enter languages and other words fall into disuse and eventually disappear. This section discusses historical changes in word structure over time. It is important to keep the diachronic changes here distinct from the synchronic changes discussed in Chapter 2. Here are some examples of synchronic word formation in English.

(5) Synchronic nominalization in English

a. mechanize \( \rightarrow \) mechanization
b. actualize \( \rightarrow \) acutalization
c. sanitize \( \rightarrow \) sanitization
d. colourize \( \rightarrow \) colourization

Similar synchronic changes are found in Korean: 문화 \( \rightarrow \) 문화적 \( \rightarrow \) 문화적으로; 역사 \( \rightarrow \) 역사적 \( \rightarrow \) 역사적으로 (roughly: culture \( \rightarrow \) cultural \( \rightarrow \) culturally; history \( \rightarrow \) historic \( \rightarrow \) historically). These words are formed as they’re needed by virtue of the speaker’s linguistic competence. Recall the wug-test from Chapter 2.

Moving on to the current discussion, we will cover the kinds of changes that take place in word structure over time. The historical changes that led to the formation of these new words are not part of the speaker’s competence. Sometimes, the historical root of the new word is no longer attested in the modern language (that is, modern speakers do not use the word and may not have ever heard of it). Sometimes the new word and its historical origin exist side by side in the modern language.

The first process we discuss is back-formation. The English verb babysit was historically derived from babysitter. Given the wide spread use of the suffix -er in English it is easy to see how babysit was back-formed from babysitter. Babysitter first appeared in English 1937, and babysit appeared in 1947. Here are some more back-formations in English and their sources:\(^1\)

---

\(^1\)The form pease is still heard in the nursery rhyme Pease Porridge Hot but is otherwise no longer attested.
(6) English back-formations

<table>
<thead>
<tr>
<th>back-formed word</th>
<th>Original word</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>cherry</td>
<td>cherries</td>
<td>Anglo-Norman <em>cherise</em></td>
</tr>
<tr>
<td>difficult</td>
<td>difficulty</td>
<td>Old French <em>difficulté</em></td>
</tr>
<tr>
<td>flab</td>
<td>flabby</td>
<td>variant of <em>flappy</em> (no longer in use)</td>
</tr>
<tr>
<td>edit</td>
<td>editor</td>
<td>from Latin, but influenced by French <em>éditer</em></td>
</tr>
<tr>
<td>pea</td>
<td>pease</td>
<td>-s reinterpreted as plural</td>
</tr>
</tbody>
</table>

Back-formations typically act on derivational morphology, so are very common in English. Back-formations also typically give rise to a new word, often with a different lexical category. For example, the verb *edit* is back-formed from the noun *editor*.

**Clipping** is the removal of part of a word to create a shortened form of the word. Clipping could result in the end of a word being removed (*apocope*), the beginning of a word being removed (*apheresis*), or the middle of the word (*syncope*). Examples of clipping in Korean include 모과 < 목과 (*quince*) and 가난 < 간난 (*hardship*). Here are some English examples. In some cases, but the original word and the clipped word are still present in the modern form of the language, such as *advertisement* and *ad*, and Korean 기저귀 and 기저기 (*diaper/nappy*). Sometimes, only the clipped form survives, and the original form has either disappeared altogether, or is rare. English examples include *mutt* from *muttonhead* and *pub* from *public house* and the two Korean examples above.

(7) Clipping in English

<table>
<thead>
<tr>
<th>apocope</th>
<th>apheresis</th>
<th>syncope</th>
</tr>
</thead>
<tbody>
<tr>
<td>examination</td>
<td>parachutne</td>
<td>spectacles</td>
</tr>
<tr>
<td>gasoline</td>
<td>gas</td>
<td>cockroach</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fantasy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>specs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fancy</td>
</tr>
</tbody>
</table>

Clipping and compounding can be combined as in *sci-fi* (from *science-fiction*), *Benelux* (from *Belgium, Netherlands, and Luxembourg*), and *op art* (from *optical art*). Clipped compounds are robust in Korean, especially in casual conversation.

(8) Clipped Compounds in Korean

<table>
<thead>
<tr>
<th>clipped compound</th>
<th>source</th>
<th>meaning of parts</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>남친</td>
<td>남자 친구</td>
<td>man friend</td>
<td>‘boyfriend’</td>
</tr>
<tr>
<td>교카충</td>
<td>교통 카드 충전</td>
<td>transportation card recharge</td>
<td>Recharging one’s transportation card</td>
</tr>
<tr>
<td>셀카</td>
<td>셀프 카메라</td>
<td>self camera</td>
<td>‘selfie’</td>
</tr>
</tbody>
</table>

Korean, Chinese (Mandarin shown), and Japanese all use clipped compounds in the names of many university. 홍익대학교 (*Hongik University*) is clipped to 홍대 (*Hongdae*). 北京大學 (*Běijīng Dàxué, Peking University*) is clipped to 北大 (*BěiDà*). 東京大學 (*Tōkyō daigaku, Tokyo University*) is clipped to 東大 (*Tōdai*).
Clipping typically forms a shortened version of the same word, rather than forming a new word with a different meaning. Recall that back-formation typically forms a new word. Also, clipping often operates on phonological constituents. A sound or a syllable is deleted. This contrasts with back-formation above, which typically operates on morphological units. Compare the following examples.

(9) Comparison of back-formation and clipping
   a. resurrection → resurrect (back-formation upon analogy with other -ion suffixed words)
   b. doctor → doc (clipping by removal of final syllable tor. 목과 → 모과 (clipping by removal of /k/)

In most clipped compounds in Korean it is unclear whether the segment removed is phonological or morphological. This is because many such examples involve words borrowed from Chinese, in which there is often a one-to-one relationship between syllables and morphemes. Consider the following example.

(10) 남자 친구 (namca chinkwu) → 남친 (namchin) ‘boyfriend’

In example (10) each syllable is a morpheme (male, offspring, relative, old, respectively). Thus, the clipped compound could be the result of deletion of syllables (phonological clipping) or morphemes (morphological clipping). The next example, however, can be analyzed only as phonological clipping.

(11) 교통 카드 충전 (kyothong khatu chwungcen) → 교카충 (kyokhachwung)

In (11) the second word 카드 (khatu) is borrowed from English card and is a single morpheme. Thus, the clipped form in (11) arises by phonologically clipping of the second syllable of each word.

Portmanteaus are similar to clipped compounds. The difference is that a portmanteau gives rise to a new word with a new meaning, whereas a clipped compound is merely an abbreviation for a longer expression. Examples in English include motel (from motor + hotel), intercom (from internal + communication), and murse (from man + purse). The Japanese word karaoke is a portmanteau of kara (Japanese for ‘empty’) and ōkesutora (borrowed from Greek, meaning ‘orchestra’).

Borrowing is a familiar and widespread phenomenon in many languages. The earliest English borrowings can be traced back to the time of trade between the Romans and the Angles in the 6th and 7th centuries. Words such as wine, butter, fork, kitchen, and inch were borrowed at this time. Because of the world-wide contact of English-speaking cultures English has borrowings from numerous languages. Here is a small sample.

(12) Words borrowed into English
Source Language Examples
Algonquian languages persimmon, moose, caribou
Arabic admiral, alchemy, alcohol, algebra
Cantonese bok choy, ketchup, kumquat
French art, music, beige, mustard
Hebrew brouhaha, cherub, cider
Indonesian gecko, bamboo, durian
Korean taekwondo, kimchi, soju
Mandarin qi, feng shui, tai chi
Philippine languages cooties, yo-yo, boondocks
Swahili jumbo, yoyo, boondocks
Turkic languages baklava, kebab, mammoth

Korean, of course, has numerous borrowings from Chinese due to extensive trade and contact between Korea and China over several hundred years. Korean has also borrowed numerous words from American English in recent times (see [Cheon 1999](#) for example). The following table lists some borrowings into Korean from other languages.

(13) Words borrowed into Korean

<table>
<thead>
<tr>
<th>Source Language</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mongolian</td>
<td>악대 (aktay ‘gelding’), 난추니 (nanchwuni ‘falcon’), 업진 (epcin ‘brisket’)</td>
</tr>
<tr>
<td>Jeju</td>
<td>지슬 (cisul ‘potato’, in Jeolla dialect)</td>
</tr>
<tr>
<td>Portuguese</td>
<td>빵 (ppang ‘bread’), 마르멜로 (malumeyllo ‘quince’)</td>
</tr>
<tr>
<td>Japanese</td>
<td>스시 (susi ‘sushi’), 고구마 (kokwuma ‘sweet potato’)</td>
</tr>
<tr>
<td>German</td>
<td>아르바이트 (alupaithu ‘part time job’), 호프 (hophu ‘bar/pub’)</td>
</tr>
</tbody>
</table>

*Reanalysis* is a process of language change in which speakers create a novel representation for a linguistic form. Here, we discuss morphological reanalysis. A well known example in English is *alcoholic*. This word is morphologically structured as *alcohol-ic*. However, the string *-aholic* has been reanalyzed as a morpheme giving rise to new words such as *workaholic* and *chocoholic*. Two examples of back-formation from above involve reanalysis. The final consonant in *cherries* and *peas* has been reanalyzed as a plural suffix to give rise to the back-formed forms *cherry* and *pea*. The word *apron* was historically *napron*. The change resulted from a reanalysis with the indefinite article: *a napron* → *an apron*.

The Arabic word *kitaab* (‘book’) was borrowed into Swahili as *kitabu*. Many Swahili nouns, however, begin with the prefix *ki-*, and this word was reanalyzed as *ki-tabu*. Swahili nouns that begin with *ki-* take the prefix *vi-* in the plural. Because of the reanalysis, the plural of *ki-tabu* is *vi-tabu*.

Reanalysis often gives rise to back-formation, as we some in some of the examples above. For example, the final [-z] sound in *pease* was reanalyzed as a plural morpheme, and the word *pea* was back-formed from *pease*. 234
**Practice 9.2:** What kind of word formation process is involved in the following words. Please choose among back-formation, clipping, portmanteau, borrowing, and re-analysis. In some cases more than one process is involved. Note that this exercise might require you to investigate the etymology of these words.

adder (kind of snake), bartend, Bollywood, K-pop, jazzersize, escargots, orientate, igloo, ’cause (meaning ‘because’), teen

9.3 Historical Sound Changes

Certain kinds of change in language are common and pervasive. In this section we will cover some of the more common sounds changes that take place in languages around the world. We will end with a brief discussion on some syntactic changes in human language. An important principle in discussing historical sound changes is the **Regularity of Sound Change.** Recall the cot/caught merger from Chapter 8. This sound change affects all words with [ɔ] and [ɑ], not just some words. In reality, there may be a few exceptions, but the sound change affects almost all of the available cases.

Sound changes take place to make adjacent sounds more similar to each other, **assimilation** or to make them more different from each other, **dissimilation. Total assimilation** is the process of one segment changing to match another segment entirely. **Partial assimilation** is the process of one segment changing some of its features to match another segment. In Chapter 4 we saw examples of intervocalic voicing. This is an example of partial assimilation. Let’s look at an example. Consider the development of the Proto-Indo-European (인도유럽조어, PIE) word for ‘eight’ in Figure 9.5. See yellow box for the use of "Proto-" in historical linguistics. The /c/ in PIE underwent backing to /k/ in Latin and then underwent total assimilation in Italian due to the neighbouring /t/. In Proto- Indo-Iranian /c/ underwent partial assimilation to match the approximate place feature of the /t/. In Proto-Germanic /c/ underwent lenition to /x/ and eventually disappeared in Modern English.

**Proto-language:** The hypothesized source language that gave rise to extant related daughter languages.

**Proto-Indo-European (PIE)** is the hypothesized source of all the Indo-European languages. One branch of Indo-European includes Proto-Germanic, the hypothesized source of the Germanic languages.

Partial assimilation involves copying some of the features of the source sound onto the target sound. Consider the following examples (Vaan 2011). Observe that the /p/ and the /gʰ/ keep their place of articulation but become nasal. (Note that this happens synchronically in Korean. Compare 악당 and 악마.)
Figure 9.5: Development of Proto-Indo-European *(h)octó:w, ‘eight’

(14) Nasal Assimilation from Proto-Indo-European to Latin

<table>
<thead>
<tr>
<th>PIE</th>
<th>Latin</th>
</tr>
</thead>
<tbody>
<tr>
<td>swep-nos</td>
<td>somnus ‘sleep’</td>
</tr>
<tr>
<td>hegnos</td>
<td>anus ‘lamb’</td>
</tr>
</tbody>
</table>

Dissimilation is less common, but is sometimes found in languages around the world. An example in English is found with the -al suffix on Latin bases to makes adjectives. Consider the following examples. Observe that -al appears if the rightmost liquid in the base is /r/ and that -ar appears if the rightmost liquid in the base is /l/. If there are no liquids in the word, then -al appears.

(15) Dissimilation in the English -al suffix

<table>
<thead>
<tr>
<th>base</th>
<th>adjective</th>
<th>base</th>
<th>adjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>orbit</td>
<td>orbital</td>
<td>alveola-</td>
<td>alveolar</td>
</tr>
<tr>
<td>person</td>
<td>personal</td>
<td>single</td>
<td>singular</td>
</tr>
<tr>
<td>verb</td>
<td>verbal</td>
<td>luna</td>
<td>lunar</td>
</tr>
<tr>
<td>practic-</td>
<td>practical</td>
<td>module</td>
<td>modular</td>
</tr>
<tr>
<td>electric</td>
<td>electrical</td>
<td>molecule</td>
<td>molecular</td>
</tr>
</tbody>
</table>

A well-known example of dissimilation in Ancient Greek and Sanskrit is Grassmann’s Law, named after Hermann Grassmann, who first analyzed the pattern in the 19th century ([Collinge](#) [1985], [Grassmann](#) [1863]). The pattern was first described for Sanskrit by Pāṇini in approximately the 4th century BCE, however. Grassmann’s Law states that if an aspirated stop is followed by another aspirated stop in the next syllable, then the first stop loses its aspiration. In Sanskrit this dissimilation applied to all stops; however, in Ancient Greek, it applied only to /t/ ([Langendoen](#) [1966]). This law is illustrated in the following Ancient Greek data ([Gamkrelidze and Ivanov](#) [2010], 22).
Grassmann’s Law in Ancient Greek

<table>
<thead>
<tr>
<th>base form</th>
<th>dissimilated form</th>
</tr>
</thead>
<tbody>
<tr>
<td>/tʰ-u-oː/</td>
<td>‘I sacrifice’</td>
</tr>
<tr>
<td>/e-tu-tʰeː:/</td>
<td>‘It was sacrificed’</td>
</tr>
<tr>
<td>/tʰrik-s/</td>
<td>‘hair’</td>
</tr>
<tr>
<td>/trikʰ-es/</td>
<td>‘hairs’</td>
</tr>
<tr>
<td>/tʰapt-ein/</td>
<td>‘to bury’</td>
</tr>
<tr>
<td>/tapʰ-os/</td>
<td>‘a grave’</td>
</tr>
</tbody>
</table>

Practice 9.3: Here are some examples of assimilation. Predict the resultant form after assimilation has taken place for the missing forms in the blank spaces.

Total Assimilation

<table>
<thead>
<tr>
<th>Caribbean Spanish</th>
<th>Old Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>/asta/</td>
<td>/atta/</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>/mismo/</td>
<td>______</td>
</tr>
</tbody>
</table>

Voicing Assimilation

<table>
<thead>
<tr>
<th>Proto-Indo-Aryan</th>
</tr>
</thead>
<tbody>
<tr>
<td>/⁺bʰudʰtis/</td>
</tr>
<tr>
<td>/⁺Hewgʰ-to/</td>
</tr>
</tbody>
</table>

Palatalization (구개음화) is the change in the quality of a consonant as a result of a (usually) following front vowel or palatal glide. Alveolar consonants typically become postalveolar before a front vowel. Likewise velar consonants often become postalveolar before a front vowel. Here is an example of the historical changes that led to the modern English words *speak* [spɪk] and *speech* [spɪʧ]. Of particular interest here are the Proto-Germanic forms. Both forms have a /k/. In the historic precursor for ‘speak’ we see the sequence /ka/, and in the historic precursor for ‘speech’ we see the sequence /ki/. The front vowel /i/ triggered palatalization of the /k/ resulting in the [ʧ] sound in Modern English.

![Figure 9.6: Development of *speak* and *speech*](image)

In Modern English palatalization can be observed in various dialects. Consider the word *assume*. In many British varieties of English it is pronounced [əsjuːm]. In some North American dialects the /j/ has disappeared: [əsuma]. In some Australian dialects, however, the /j/ has triggered palatalization: [əsum].

Palatalization took place in Mandarin in the 18th century (Chen [1976]). Compare the Mandarin pronunciation of *Beijing* with that of other Chinese languages. Observe that in Mandarin the affricate before /i/ is a
palatal affricate, while in the other languages /k/ persists. Note that the Korean pronunciation 북경 was borrowed before palatalization took place, so the /k/ pronunciation persists. Note also that although the English pronunciation has a palatal affricate, reflecting the palatalized form in Mandarin, the French name Pékín [pekê] and the Italian name Pechino [pekino] preserve the original /k/ from Middle Chinese.

Korean has synchronic palatalization before affixes that begin with /i/ or /j/, as shown in Figure 9.2. Compare the following forms.

(17) Synchronic Palatalization in Korean

<table>
<thead>
<tr>
<th>Non-palatalized</th>
<th>IPA</th>
<th>English</th>
<th>Palatalized</th>
<th>IPA</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>같이</td>
<td>[katʰa]</td>
<td>‘is similar’</td>
<td>같이</td>
<td>[katʃi]</td>
<td>‘together’</td>
</tr>
<tr>
<td>붙여</td>
<td>[putʰa]</td>
<td>‘is stuck to’</td>
<td>붙여</td>
<td>[putʃa]</td>
<td>‘sticks to’</td>
</tr>
<tr>
<td>굳어</td>
<td>[kud(Duration)]</td>
<td>‘is firm’</td>
<td>굳어</td>
<td>[kuʃi]</td>
<td>‘firmly’</td>
</tr>
</tbody>
</table>

Practice 9.4: Given the following forms showing the change from PIE to Proto-Albanian, predict the forms left blank.

PIE                           Proto-Albanian
/*pénkʰwə/ > /*penʃe/          (‘five’)
/*kwær-/ > /*iʃerkā/          (‘build’)
/*kwel(-ti)/ > ____            (‘turn, twist’)
/*mekʰw/ > ____               (‘Don’t!’)

Affrication is the process of a stop becoming an affricate. Here’s an example from Latin to Old French (Pope 1934: 125). The sounds /k/ and /g/ became affricates before the vowels /i/ and /e/.

Latin          Old French
hundred         [kentum]       [sfENT]
people          [gentum]       [dʒENT]
root            [radikina]     [raʃinÉ]
silver          [argentum]     [arDʒENT]

Umlaut is the change in the quality of a vowel in response to another vowel. For instance, a high front vowel such as /i/ can trigger nearby vowels to either raise or move forward. The vowel /a/ can trigger other vowels to lower. The vowel /u/ can trigger other vowels to become more back. The following examples of umlaut in Old English to Modern English shows how some of the English irregular plurals arose. In an earlier stage of Old English the plural morpheme was /-i/. This triggered eventual vowel fronting in a later state before Old English (labelled ‘Pre-Old English II’ in the chart below. Eventually in Old English, the /-i/ plural suffix was lost, but the change on the vowels remained. The Great Vowel Shift is responsible for the vowel qualities in Modern English.
Metathesis is the transposition of two sounds or syllables. In child English the word *spaghetti* is often said as ‘*pasghetti*’. Metathesis took place in the development of English. Consider the following examples (Campbell, 2004).

(18) Metathesis in Old English

<table>
<thead>
<tr>
<th>Old English</th>
<th>Modern English</th>
</tr>
</thead>
<tbody>
<tr>
<td>wæps</td>
<td>wasp</td>
</tr>
<tr>
<td>æks</td>
<td>ask (still heard as [æks] in some dialects)</td>
</tr>
<tr>
<td>bridd</td>
<td>bird</td>
</tr>
<tr>
<td>beorht</td>
<td>bright</td>
</tr>
</tbody>
</table>

Spanish underwent a series of metathesis changes from Latin. Here are some examples (Campbell, 2004).

(19) Metathesis in Spanish

<table>
<thead>
<tr>
<th>Latin</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>parabola</td>
<td>palabra</td>
</tr>
<tr>
<td>miraculum</td>
<td>milagro</td>
</tr>
<tr>
<td>periculum</td>
<td>peligro</td>
</tr>
<tr>
<td>crocodilus</td>
<td>cocodrilo</td>
</tr>
</tbody>
</table>

Finally, Old Finnish had stem final /h/ on some words. In Standard Modern Finnish, the /h/ was lost, but in some western dialects of Finnish it underwent metathesis. Consider the following examples.

(20) Metathesis in Finnish

<table>
<thead>
<tr>
<th>Old Finnish</th>
<th>Standard Finnish</th>
<th>Western Finnish</th>
</tr>
</thead>
<tbody>
<tr>
<td>*orih</td>
<td>ori</td>
<td>orhi</td>
</tr>
<tr>
<td>savuh</td>
<td>savu</td>
<td>sauhu</td>
</tr>
<tr>
<td>valeh</td>
<td>vale</td>
<td>valhe</td>
</tr>
<tr>
<td>veneh</td>
<td>vene</td>
<td>venhe</td>
</tr>
</tbody>
</table>

Lenition (연음화) refers to the weakening of a sound. Consonants typically progress through the stages shown in Figure 9.7 (Vennemann Theo, 2011). Not shown in the chart below is a form of lenition called debuccalization, in which a consonant loses its place of articulation and becomes /h/ or /ʃ/.
Grimm’s Law and Verner’s Law present typical cases of consonant lenition (Collinge, 1985). Grimm’s Law is a three stage process; however, only one stage deals with lenition, which we discuss here. Grimm’s Law states that voiceless stops in Proto-Indo-European became voiceless fricatives in Proto-Germanic. Here are some examples.

(21) Reflexes of Grimm’s Law in Indo-European.

<table>
<thead>
<tr>
<th>Proto-Indo-European non-Germanic examples</th>
<th>Proto-Germanic example</th>
<th>Germanic example</th>
</tr>
</thead>
<tbody>
<tr>
<td>*póxs</td>
<td>Latin: peːs, pedis</td>
<td>English: foot</td>
</tr>
<tr>
<td>Russian: pod</td>
<td>German: Fuss</td>
<td></td>
</tr>
<tr>
<td>*fóts</td>
<td>English: foot</td>
<td></td>
</tr>
<tr>
<td>Russian: pa</td>
<td>Swedish: fot</td>
<td></td>
</tr>
<tr>
<td>*tróts</td>
<td>Latin: tertius</td>
<td>English: third</td>
</tr>
<tr>
<td>Russian: tretij</td>
<td>Old Saxon: thriddio</td>
<td></td>
</tr>
<tr>
<td>*cun</td>
<td>Latin: canis</td>
<td>English: hound</td>
</tr>
<tr>
<td>Welsh: cwn</td>
<td>German: Hund</td>
<td></td>
</tr>
</tbody>
</table>

Various scholars noticed some exceptions to Grimm’s Law. It was Verner who discovered the pattern to these exceptions. He noticed that if the voiceless consonant in Proto-Indo-European was word-internal and followed a stressless syllable, then the voiceless stop became a voiced stop rather than a voiceless fricative. This is known as Verner’s Law. Here are some examples. Observe in the first example that Proto-Indo-European /p/ becomes /f/, following Grimm’s Law; however /t/ becomes /d/, following Verner’s Law. In the second example Proto-Indo-European /p/ became /b/, following Verner’s Law.

(22) Reflexes of Verner’s Law in Indo-European.

<table>
<thead>
<tr>
<th>Proto-Indo-European non-Germanic example</th>
<th>Proto-Germanic example</th>
<th>Germanic example</th>
</tr>
</thead>
<tbody>
<tr>
<td>*phútːr</td>
<td>Latin: pateːr</td>
<td>English: father</td>
</tr>
<tr>
<td>Hindi: pitaː</td>
<td>Old Saxon: fadar</td>
<td></td>
</tr>
<tr>
<td>*septːn</td>
<td>Latin: septem</td>
<td>English: seven</td>
</tr>
<tr>
<td>Sanskrit: saptán</td>
<td>German: sieben</td>
<td></td>
</tr>
</tbody>
</table>
**Practice 9.5:** Consider the following examples of lenition in Nyaheun, an Austroasiatic language, (adapted from Kosaka [2007]). Predict the resulting forms for lenition based on the evidence here.

<table>
<thead>
<tr>
<th>Old Nyaheun</th>
<th>Middle Nyaheun</th>
</tr>
</thead>
<tbody>
<tr>
<td>/*ťapal/</td>
<td>&lt; /*ťapal/</td>
</tr>
<tr>
<td>/*p̌ate/</td>
<td>&lt; /*p̌ate/</td>
</tr>
<tr>
<td>/*ǩačeet/</td>
<td>&lt; _____</td>
</tr>
<tr>
<td>/*ťakuaj/</td>
<td>&lt; _____</td>
</tr>
</tbody>
</table>

*Middle Korean was thought to have consonant clusters of /s/ followed by an oral stop. The /s/ underwent elision and the remaining consonant became fortis. Evidence for this claim comes from contemporary dialects of Korean. For example 폭 is pronounced as [p̣oṭku] in some northern dialects, and 뱐 is pronounced as [p̣iḍoṇ] in some southern dialects.*

Finally, **epenthesis** is the addition of a sound. The most common types of epenthesis are the addition of a vowel to break up a consonant cluster and the addition of a consonant to break up a vowel cluster (also called *vowel hiatus*). For example, the Proto-Slavic *gorđu̯* became *goroḍu̯* in East Slavic and eventually *gorod* in modern Russian, with elision of the final vowel.

(25) *gorđu̯* > *goroḍu̯* > *gorod*
9.4 Comparative Historical Reconstruction

Genetic lineage among languages is determined by examining cognates. Consider the following example of Germanic languages and the historical Proto-Germanic forms from which they are derived. Note that the words are written in their orthographic forms.

<table>
<thead>
<tr>
<th>English</th>
<th>German</th>
<th>Dutch</th>
<th>Proto-Germanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>mother</td>
<td>Mutter</td>
<td>moeder</td>
<td>*módør</td>
</tr>
<tr>
<td>brother</td>
<td>Bruder</td>
<td>broer</td>
<td>*bróðar</td>
</tr>
<tr>
<td>house</td>
<td>Haus</td>
<td>huis</td>
<td>*húsam</td>
</tr>
</tbody>
</table>

It is not difficult to see that English, German, and Dutch are related. As we have discussed, they come from Proto-Germanic. Note that unattested and reconstructed forms are marked with an asterisk. In this section we will learn about the traditional techniques involved in determining historical relationships among languages.

To go back in time linguistically, we must “undo” these changes. For example the change from /apa/ to /aba/ is phonetically plausible as it is a case of weakening as described above. The change from /aba/ to /apa/, however, is phonetically implausible.

Let’s begin by looking at a relatively simple example. We will construct the Proto-Romance forms from the following contemporary Romance data. Although stress is not marked in these examples, it is important to note that the first syllable is stressed in the first three words.

<table>
<thead>
<tr>
<th>Proto-Romance</th>
<th>Spanish</th>
<th>Portuguese</th>
<th>Italian</th>
<th>Sardinian</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>hilo</td>
<td>fiu</td>
<td>filo</td>
<td>filu</td>
<td>fir</td>
<td>‘string’</td>
</tr>
<tr>
<td>vida</td>
<td>vida</td>
<td>vita</td>
<td>bita</td>
<td>vitɔ</td>
<td>‘life’</td>
</tr>
<tr>
<td>vino</td>
<td>viju</td>
<td>viŋo</td>
<td>binu</td>
<td>vin</td>
<td>‘wine’</td>
</tr>
<tr>
<td>pino</td>
<td>pinu</td>
<td>pino</td>
<td>pinu</td>
<td>pin</td>
<td>‘pine tree’</td>
</tr>
<tr>
<td>haθer</td>
<td>fazer</td>
<td>fare</td>
<td>fakere</td>
<td>far</td>
<td>‘to do’</td>
</tr>
</tbody>
</table>

In the first and last example we see that /l/ in Proto-Romance underwent debuccalization in Spanish to become /h/. Furthermore, many of the vowels are the same in these examples, so they can be reconstructed to the same vowel. Also, in the first example there was a liquid in the middle of the word that underwent elision in Portuguese. In the other languages it is either an /l/ or an /r/. Since either change is equally likely (/l/ to /r/ or /r/ to /l/), we use the majority rules principle and assume that the proto form had /l/. This gives us the following so far.

<table>
<thead>
<tr>
<th>Proto-Romance</th>
<th>Spanish</th>
<th>Portuguese</th>
<th>Italian</th>
<th>Sardinian</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>*fil</td>
<td>hilo</td>
<td>fiu</td>
<td>filo</td>
<td>filu</td>
<td>fir</td>
</tr>
<tr>
<td>?i</td>
<td>vida</td>
<td>vija</td>
<td>bita</td>
<td>vitɔ</td>
<td>‘life’</td>
</tr>
<tr>
<td>?i</td>
<td>vino</td>
<td>viju</td>
<td>binu</td>
<td>vin</td>
<td>‘wine’</td>
</tr>
<tr>
<td>?i</td>
<td>pino</td>
<td>pinu</td>
<td>pinu</td>
<td>pinu</td>
<td>pin</td>
</tr>
<tr>
<td>fa??r</td>
<td>haθer</td>
<td>fazer</td>
<td>fare</td>
<td>fakere</td>
<td>far</td>
</tr>
</tbody>
</table>

We notice in all the examples that there is some variation between /o/ and /u/. Vowels tend to become more
central over time, so we surmise that the proto forms contained /u/. We can now reconstruct the following forms.

<table>
<thead>
<tr>
<th>Proto-Romance</th>
<th>Spanish</th>
<th>Portuguese</th>
<th>Italian</th>
<th>Sardinian</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>*filu</td>
<td>hilo</td>
<td>fiu</td>
<td>filo</td>
<td>filu</td>
<td>fir</td>
</tr>
<tr>
<td>?i</td>
<td>vida</td>
<td>vida</td>
<td>vita</td>
<td>bita</td>
<td>vitọ</td>
</tr>
<tr>
<td>?inu</td>
<td>vino</td>
<td>viñu</td>
<td>viņo</td>
<td>binu</td>
<td>vin</td>
</tr>
<tr>
<td>pinu</td>
<td>pino</td>
<td>pinu</td>
<td>pino</td>
<td>pinu</td>
<td>pin</td>
</tr>
<tr>
<td>faʔr</td>
<td>haʔer</td>
<td>fazer</td>
<td>fare</td>
<td>fakere</td>
<td>far</td>
</tr>
</tbody>
</table>

The change from /v/ to /b/ is an example of fortition, which is common in stressed syllables. So we can assume that /v/ underwent fortition in Sardinian to /b/. Finally, we can assume that the proto form in the last example contained /k/, which underwent lenition or elision in the various languages. Note also that some vowels underwent elision in Italian and Romanian. This gives us the following Proto-Romance forms.

9.5 Historical Syntax

The techniques described above have been applied to numerous language families around the world to show genetic affiliation among languages (which we will see in the next section) and to reconstruct proto-languages. However, they have most frequently been applied to phonological and morphological systems of languages. More recently, some scholars have developed techniques for tracking changes above the word level. Harris and Campbell (1995) identify the following questions that historical syntax seeks to address.

1. What kinds of syntactic changes happen in individual languages?
2. What are the general properties of word order changes?
3. What are the implications of syntactic change to our understanding of syntactic theory?

In this section we will only discuss some examples of syntactic changes.

9.6 Genetic Classification

The comparative method has been quite fruitful in providing cladograms (or tree diagrams) of several language families around the world (Schleicher 1853, 1873). A strict interpretation of the cladogrammatic representation of the historical relationships of languages assumes that as societies split into sub-groups, they no longer have
contact, and the two forms of the language evolve independently of each other, first into distinct dialects and eventually into distinct languages, as shown in Figure 9.1. Historically, of course, societies have contact with one another, giving rise to contact-induced change. The cladograms commonly used to represent historical relationships among languages and dialects represent one aspect of how language change happens (Bloomfield 1933; Southworth 1964).

The largest known grouping of a set of languages is called a phylum or more commonly a language family. We have mentioned the Indo-European family above and will look at some other language families in this section. A super-family (or super-phylum) is a proposed larger grouping that links together two or more families. Super-families usually only have weak support among historical linguists. An isolate is a language that is not known to be related to any other language in the world. Some well known isolates include Basque, Haida, and Zuni. See Figure 9.8 for some more examples.

The origins of language are unknown and the question of when and how language began is sometimes called the hardest problem in science. Theories of the origin of human language are classified either as polygenetic or monogenetic. Linguists overwhelmingly support the polygenetic theory of human language development, which states that languages evolved more than once and at different locations in the world. The monogentic theory holds that all human languages can be traced back to a single human language called Proto-World or Proto-Sapiens about 100,000 years ago. The monogentic theory is supported only by a very small number of linguists.

9.6.1 The Altaic Superfamily

The Altaic Hypothesis holds that the languages of the Turkic, Tungusic, and Mongolic families are all related to a hypothesized Proto-Altaic family. A stronger version of the hypothesis holds that Korean and Japonic
languages can also be traced back to Proto-Altaic. Much less commonly, Ainu is also considered part of the Altaic super-family. An excellent history into the study of the Altaic Hypothesis can be found in Poppe (1965).

![Supra-Macro-Altaic](Supra-Macro-Altaic)  
![Macro-Altaic](Macro-Altaic)  
![Ainu](Ainu)  
![Micro-Altaic](Micro-Altaic)  
![Koreanic](Koreanic)  
![Japonic](Japonic)  
![Turkic](Turkic)  
![Tungusic](Tungusic)  
![Mongolic](Mongolic)

Figure 9.9: Grouping of the proposed Altaic superfamily

The **Turkic** family contains approximately 35 languages spoken from Turkey to northeastern Siberia. They are SOV and agglutinative and typically display vowel harmony. Turkic languages include Turkish, Uzbek, Kazakh, and Uyghur.

**Tungusic** languages are spoken in Eastern Russian and Northwestern China (Whaley and Malchukov, 2012). Most Tungusic languages are endangered leaving the future of the this family uncertain. Tungusic languages are SOV and agglutinative. They typically have vowel harmony. Manchu is well known as the language of the Manchurian people who ruled China during the Qing Dynasty. The language is critically endangered now with only about ten speakers (as of 2011). There are efforts in the People’s Republic of China to revitalize Manchu because of its historical value. Xibe, Evenke and Even are the three Tungusic languages with the greatest number of speakers.

The **Mongolic** family comprises a set of languages spoken primarily in Mongolia and northern China (Inner Mongolia) (Janhunen, 2006, 2012). Proto-Mongolic was spoken by Genghis Khan and citizens of the Mongol Empire. The family is partially represented in Figure 9.10.

![Mongolic](Mongolic)  
![Moghol](Moghol)  
![Common Mongolic](Common Mongolic)  
![Shirongolic](Shirongolic)  
![Ordos](Ordos)  
![Buryat](Buryat)  
![Monguor](Monguor)  
![Santa](Santa)  
![Dagur](Dagur)

Figure 9.10: Mongolic Languages

Mongolic languages are SOV and have agglutinative morphology. Many Mongolic languages have an inclusive/exclusive distinction in the first person plural. Moghol is spoken in Afghanistan by descendants of one of Genghis Khan’s armies stationed there. It is unknown if there are still native speakers of this language.

The **Koreanic** family contains two members: Korean and Jeju. Korean was once thought to be an isolate until the recognition of Jeju as a distinct language from the mainland dialects.

The **Japonic** languages consist of a small family as shown below. Like Korean, Japanese was once considered an isolate until it was shown that the other Japonic languages are not mutually intelligible with Japanese.
Ainu is a highly endangered language. It is believed by most historical linguistics to be an isolate; however, as discussed above, it has been suggested that is is an Altaic language; however, this position is not widely held. It is spoken in Hokkaido in northern Japan. Earlier it was spoken on the northern part of Honshu Island. Place names in northern Japan reflect the former Ainu presence in the area. Many towns end in -betsu, which is borrowed from the Ainu word *pet, which means ‘river’. Examples include Noboribetsu and Imabetsu.

9.6.2 Indo-European Languages

The Indo-European family contains hundreds of languages and dialects and is spoken by over three billion people worldwide. Figure 9.13 lists just a small portion of the Indo-European languages found around the world.

Indo-European languages cover most of Europe, western and central Asia, and southern Asia (except for the southern part of India). In modern times Indo-European languages have spread to North and South America. Proto-Indo-European was spoken up to about 2500 BCE, though this is a rough estimate. It is also thought to have originated in western Asia near the Caspian and Black Seas. The large number of cognates across Indo-European for land animals, and the near absence of cognates for sea animals strongly suggests the Proto-Indo-Europeans originated inland (Clackson, 2007).

Of interest to East Asia is the Tocharian branch of Indo-European, which is thought to have died in the 9th century. The Korean word 밀 as in 청밀 (‘honey’) is borrowed from Chinese (mi). Mandarin 蜜 mi derives from Old Chinese *mri or *mli, which in turn was borrowed from Tocharian *m@l̥ɾ. The Tocharian word ultimately derives from Proto-Indo-European *médʰu. The English word mead derives from the same Proto-Indo-European word (Lubotsky, 1998).
Figure 9.13: Cladogram for Indo-European Languages
9.6.3 Sino-Tibetan Family

The Sino-Tibetan languages are spoken throughout East Asia and South-East Asia. They are characterized by largely being tonal. Most Sino-Tibetan languages are SOV; however some (most notably the Chinese languages) are SVO. The divisions of Sino-Tibetan are an ongoing topic of research, so we present a simplified diagram of Sino-Tibetan languages here. Sino-Tibetan is second only to Indo-European in terms of number of speakers. Indeed, Mandarin has the most native speakers of any language in the world.

We begin with a discussion of the Sinitic languages. These are commonly referred to as Chinese. Sinitic languages are generally tonal, SVO and isolating. They have fairly similar syntax and morphology, with some notable differences. They differ greatly in terms of phonology.

9.6.4 Austroasiatic Family

The Austroasiatic languages stretch from south-east Asia to north-western India. The precise classification of Austroasiatic languages is still a matter of debate. We just discuss some general properties of Austroasiatic and mention some well known members of this family.

A general property of Austroasiatic languages is that lexical words often consist of a short syllable followed by a long syllable. Voicing contrasts are common in most Austroasiatic languages. It is common to have a three- or four-way voicing contrast. (See chapter 3 again for descriptions of different laryngeal states: voiceless, voiced, breathy, creaky.)

Vietnamese is probably the most well-known member of the Austroasiatic family. Like Korean, Vietnamese has a three-way contrast in its obstruents. Unlike the rest of the Austroasiatic family, however, vowel contrasts are tonal rather than laryngeal (Pham 2003). Other Austroasiatic languages include Khmer, Mon, and Santali.
9.6.5 Kra-Dai Family

The Kra-Dai language family used to be called Tai-Kadai based on a previous understanding of the structure of the family. The two most well known members of this family are Thai and Lao. There are 93 Kra-Dai languages spoken throughout Thailand, Laos, Vietnam and China.

9.6.6 Austronesian Family

The Austronesian family stretches over an extremely large area as the Austronesian people were prolific sailors.
Answers to Practice Questions

Practice Question 1.
/profeI/, /profænti/, /*profanə/
/dəədEv/, /dəəvətv/, /*deivə/
/profænd/, /profændti/, /*profund/ Before the GVS, /u/ underwent trisyllabic laxing to become /u/. As noted in the text, the vowel /u/ became /ʌ/ in Modern English.
/skəl/, /skələ/, /*sko:l/ Before the GVS, /oː/ underwent trisyllabic laxing to become /ɔ/. As noted in the text, the vowel /ɔ/ became /æ/ or /ə/ in Modern English.

Practice Question 2.
Caribbean Spanish: /mimmo/
Old Swedish: /θakka/
Proto-Indo-Aryan: /*Hewgʰdo/

Practice Question 3.
Proto-Albanian forms:
/*tʃela/ (Note: the change from /a/ to /a/ is not predictable from the data given.)
/*metʃe/ (Note: the addition of /e/ at the end of the word is not predictable from the data given.)

Practice Question 4.
Middle Nyaheun forms:
/*kɔjeet/
/*tɔjuaj/
Exercises

Question 1. Identify the historical processes taking place in the following data.

1. Russian /gorod/, Serbo-Croatian /grad/
2. Proto-Algonquian */penkwi/ > Ojibwe /bingwi/
3. Middle Persian /stūn/ > Early New Persian /sutūn/ (‘column’)
4. Proto-Slavic */kork/ > Bulgarian /krak/ (‘foot’)

Question 2. Reconstruct the Proto-Polynesian forms in the following data.

<table>
<thead>
<tr>
<th>Maori</th>
<th>Hawaiian</th>
<th>Samoan</th>
<th>Fijian</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>pou</td>
<td>pou</td>
<td>pou</td>
<td>bou</td>
<td>post</td>
</tr>
<tr>
<td>tapu</td>
<td>kapu</td>
<td>tapu</td>
<td>tabu</td>
<td>forbidden</td>
</tr>
<tr>
<td>taŋi</td>
<td>kani</td>
<td>taŋi</td>
<td>taŋi</td>
<td>cry</td>
</tr>
<tr>
<td>takere</td>
<td>kaʔele</td>
<td>taʔele</td>
<td>takele</td>
<td>keel</td>
</tr>
<tr>
<td>hono</td>
<td>hono</td>
<td>fono</td>
<td>vono</td>
<td>stay, sit</td>
</tr>
<tr>
<td>marama</td>
<td>malama</td>
<td>malama</td>
<td>malama</td>
<td>light, moon</td>
</tr>
<tr>
<td>kaho</td>
<td>?aho</td>
<td>?aso</td>
<td>kaso</td>
<td>thatch</td>
</tr>
</tbody>
</table>

Question 3. Identify the sound changes from Proto-Austronesian to Bahasa Aceh in the following data (adapted from Nuzwaty [2016]).

<table>
<thead>
<tr>
<th>Proto-Austronesian</th>
<th>Bahasa Aceh</th>
</tr>
</thead>
<tbody>
<tr>
<td>*/bonig/</td>
<td>/bоних/</td>
</tr>
<tr>
<td>*/tanag/</td>
<td>/танаг/</td>
</tr>
<tr>
<td>*/buag/</td>
<td>/boh/</td>
</tr>
<tr>
<td>*/goasig/</td>
<td>/гасих/</td>
</tr>
<tr>
<td>*/pɔnug/</td>
<td>/пɔног/</td>
</tr>
</tbody>
</table>
Further Reading

- [Campbell (2004)] - a good next step for those interested in historical linguistics and the comparative reconstruction technique.

- [Clackson (2007)] - provides an in depth but introductory discussion of issues in the study of Indo European from a pedagogical perspective.

- [Goddard (2005)] - provides an overview over the languages of east and southeast Asia.

- Online Resources:
  
  - WALS Genealogical Language List - [https://wals.info/languoid/genealogy](https://wals.info/languoid/genealogy)
  
  - The Sounds of Middle Korean - [https://www.youtube.com/watch?v=bsDvgXdJwOK](https://www.youtube.com/watch?v=bsDvgXdJwOK)
Chapter 10

Language Acquisition

By the end of this chapter you should:

• understand the stages of language development
• be familiar with the competing theories for language acquisition
• be familiar with the Poverty of the Stimulus argument
• understand the properties of second language acquisition

10.1 Introduction

Have you ever pulled your hair out in frustration over memorizing lists of irregular verbs or trying to internalize complex conjugation paradigms learning a second language as an adult? Learning English as a second language requires you to sit and memorize lists of irregular past tense forms (drive, drove, driven; sing, sang, sung; etc.). How is it that children so effortlessly acquire their first languages? Before starting, let us remind ourselves that we are talking about spoken and signed languages, and not the written representation of language. Unlike spoken and signed language, literacy is a skill which must be consciously developed. And unlike spoken and signed language, learning to write comes with hardships, whether it is remembering the silent ‘k’ in knife or knee, or remembering which Korean words are spelled with ㅏ and which with ㅐ (both of these vowels sound the same for most speakers of Korean), or memorizing complex Chinese characters such as 龙 (dragon). This chapter will not be concerned with the acquisition of writing.

Language acquisition is the study of how humans acquire language. First language acquisition (L1 acquisition) takes place at a remarkable speed. A four-year-old child has mastered a complex system of language, but does not know how to tie their shoes. Second language acquisition (L2 acquisition) in adulthood, on the contrary, takes place much more slowly, and involves difficulties not faced by children acquiring language. The study of language acquisition involves both L1 and L2 acquisition, as well as bilingual acquisition in childhood. The errors that children make during language acquisition provide crucial insight not only into the structure of language, but into how the child’s developing mind acquires language. Consider the following anonymous excerpt. Observe that the child is oblivious to the adult’s attempts to correct the child’s grammar.
Child: My teacher holded the baby rabbits and we patted them.
Adult: Did you say your teacher held the baby rabbits?
Child: Yes.
Adult: What did you say she did?
Child: She holded the baby rabbits and we patted them.
Adult: Did you say she held them tightly?
Child: No, she holded them loosely.

This chapter discusses the stages of L1 acquisition and introduces various theoretical models to account for the observes properties of language acquisition. The chapter ends with a brief discussion of L2 acquisition. In discussing the properties of language acquisition, the age of children is presented as x;y, where x is years and y is months. A child who is 2 years and 7 months old is given as 2;7.

10.2 L1 Acquisition

An ongoing debate in linguistics concerns how children acquire language. Emergentist views language acquisition as the result of environmental factors combined with general cognitive processes (MacWhinney 2013). Crucially, emergentist views hold that there is no separate language acquisition device or specialized language module. That is, children are born tabular rasa—with no preconceived concept of what shape language should take. Empiricist views of language acquisition hold that there is a specific language acquisition device (LAD) that is primarily responsible for language acquisition. The concept of a LAD is similar to that of UG discussed in chapter 1. The details of the LAD have changed considerably over time, but can be traced back to Chomsky (1965). During the Government and Binding era of the 1980s linguists proposed a large number of innate binary parameters to explain acquisition. More recently, however, Chomsky (2013) has proposed that the only universal aspect of human language is perhaps merge, as discussed in chapter 5.

How do children acquire language? Various theories and ideas have been passed back and forth since antiquity addressing this question. We will run through some of commonly held misconceptions here before we address this question more seriously. One common misbelief is that children acquire language by imitation. Recall the conversation on page 254 where the child continues to use the non-adult for holded for the past tense of hold. It is thought that imitation does play a small role in acquisition of vocabulary, but not grammar. Imitation also does not account for regularization errors that children make around the ages 3-5.

(1) *eated, *goed, *안었다, *안 밥 먹어
In psychology, behaviourism played a major role in human development in the 1950s. B. F. Skinner’s *Verbal Behaviour* offered an in-depth explanation for language acquisition based on behavioural approaches to language acquisition. That is, Skinner tried to argue that children acquire language from correction and reinforcement. The following short dialogue illustrates how this approach is misguided.

Child: Nobody don’t like me.
Mother: No, say, “Nobody likes me.”
Child: Nobody don’t like me.
[repeated 8 times]
Mother: Now, listen carefully. Say, “Nobody likes me.”
Child: Oh! Nobody don’t likes me.

In some societies, children are not spoken to until the age of five or even older, as such these children never receive reinforcement or correction (Ochs and Schieffelin 1984). English speaking parent’s actually don’t correct their children very often. Furthermore, negative reinforcement is found with grammatical sentences, and positive reinforcement with ungrammatical sentences is often found.

It is sometimes thought that children acquire language by analogy. Recall the Gleitman data from the chapter one, repeated here.

\[
(2) \, \begin{align*}
\text{a.} \quad & \text{Susan painted the red barn.} \\
\text{b.} \quad & \text{Susan painted the barn red.} \\
\text{c.} \quad & \text{Susan saw the red barn.} \\
\text{d.} \quad & *\text{Susan saw the barn red.}
\end{align*}
\]

Children never make errors of the kind in the last sentence. If children learned by analogy, such errors would be expected from time to time, but they are never found.

Structured input is very useful in L2 instruction (Benati 2018), but does not play a role in L1 acquisition in children. Some parents use child-directed speech when speaking to children, but not all do. Recall also the discussion above where in some societies children are typically never directly spoken to. See Lieven (1994) for an extended discussion. See Cruttenden (1994), however, for a phonetic and prosodic description of child-directed speech and how it might contribute to L1 acquisition.

The next few sections discusses the stages of L1 acquisition.

### 10.2.1 Language in Infancy

The early stages of language acquisition are remarkably uniform across languages. At birth, an infant can distinguish sound of its mother’s voice from that of other humans (Kisilevsky et al. 2003). From birth to about six month of age, an infant can distinguish all possible human sounds, while adults cannot (Kuhl et al. 1992). For instance, all infants can distinguish [i] and [ɪ]; however, after approximately age 0:6, infants acquiring
language in an environment where these two sounds are not contrastive lose the ability to distinguish between them. Recall that [ɾ] and [l] are contrastive in English, but not in Korean. Thus, infants acquiring Korean only will start to lose this contrast at about 0;6. However, [kʰ] and [k] are contrastive in Korean, but not in English. Thus, English speakers lose the ability to discriminate aspiration at about 0;6.

Babbling (웅아리) usually starts at 0;6. It uses mostly the world’s most common sounds, and has a simple syllable structure: CV. Deaf babies start ‘manual babbling’ in nearly an identical pattern that hearing babies babble using speech (Petitto and Marentette, 1991). Babbling has been observed and studied for at least 200 years (Oller, 2000); however, has only recently been studied in Korean (Ha and Oller, 2019). Babies slowly transition from babbling to producing speech sounds in the baby’s linguistic environment. At approximately 1;0 babies start producing basic CV syllables and also start to imitate the intonational contours of the language they’re acquiring. The link between babbling and language acquisition is still poorly understood.

### 10.2.2 First Sounds and Words

The first words are usually formed around 1;0. By age 6;0 a child’s vocabulary contains approximately 14,000 words. How does this transition take place? It is important to note that there is much variation from one individual to the next in terms of the time lines presented here. These first ‘words’ often correspond to complex thoughts. That is, they are holophrastic. Here is an example of some of the vocabulary items of a baby at age 1;4 acquiring English.

(3) **English Vocabulary (1;4)**

- [ʔau] not, no, don’t
- [bʌʔ]/[mʌʔ] up
- [da] dog
- [iʔo]/[siʔo] Cheerios (a kind of cereal)
- [sa] sock
- [asæː]/[sæː] What’s that?

It’s important to note that these first words typically correspond to complete utterances. Thus, if this child says [maʔ] it doesn’t simply mean just ‘up’. It could mean that the child wants up in his mothers arms or that a bird is flying up in the air or that a dog is jumping up. Again, there is much variation in the age at which first words are acquired. The following chart shows ages of six children at 10- and 50-word vocabularies (Robb et al., 1994). Note that the ages have been converted to decimal values.
Fluent speech rarely contains pauses between words. So, we must ask ourselves how infants start breaking down a continuous stream of speech into words. Segmenting words takes place around 1;6 to 2;6. Languages use prosodic cues to pick out words (Jusczyk et al., 1999; Johnson and Jusczyk, 2001). In English, stress used to pick out words. Moras are used in Japanese. In English, trochees learned more easily: ‘puppy’, ‘doctor’. Iambs are usually clipped: ‘guitar’ becomes ‘tar’.

Errors in children’s speech are not haphazard, rather, they are rule-governed (Smith, 1973). For instance, spoon often produced as [pun], but not as [sun]. Errors are consistent with rules found in natural language. Children do not make up implausible rules. For example, word final devoicing is extremely common cross-linguistically. German has word-final devoicing, and Korean loses laryngeal features in word-final position. English, however, maintains a voicing contrast in word-final position. Thus, English contrasts bit and bid; however, 오냐 and 앞 sound the same in Korean (unless followed by a vowel). Word-final devoicing is often found in children’s grammar during acquisition. The child eventually rejects the rule (in English) or retains it (in German).

The Fis-phenomenon is related to the difference between perception and production. Children are able to perceive the difference between different sounds before they are able to produce the same two sounds (Berko and Brown, 1960). For example, /s/ and /ʃ/ are contrastive in English. Children acquire the ability to perceive the difference between these two sounds early, but their ability to produce them is acquired later. Thus, fish is produced as [fɪʃ] in children’s speech, even though the distinction can be perceived by children.

10.2.3 The Development of Grammar

The acquisition of syntax starts with the one-word stage. As we saw above the single word utterances are often holophrastic in that they convey an entire thought. After individual words are acquired the child proceeds to the two-word stage. At the two-word stage, the word order typically follows that of the adult form of the language. There is also a syntactic and semantic relationship between the words. Here are some English examples:

(4) Two-word Utterances

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Meaning</th>
<th>Semantic relationship</th>
<th>Syntactic relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doggy bark</td>
<td>‘The dog is barking’</td>
<td>agent-action</td>
<td>Subject-Verb</td>
</tr>
<tr>
<td>Hit doggy</td>
<td>‘I hit the dog.’</td>
<td>action-theme</td>
<td>Verb-Object</td>
</tr>
<tr>
<td>애기먹어</td>
<td>‘The baby is eating.’</td>
<td>agent-action</td>
<td>Subject-Verb</td>
</tr>
<tr>
<td>배타</td>
<td>‘(somebody) boards the boat.’</td>
<td>action-location</td>
<td>Object-Verb</td>
</tr>
</tbody>
</table>

At the two-word stage in Korean children typically produce OV utterances (for transitive verbs) and SV utterances (for intransitive verbs) (Choi, 1999). Here are two examples from Choi.

(5) Two-Word Stage in Korean

a. paci ip-e (바지 입어, JS 1;9)
   pants put.on-THV
   ‘(I am) putting on pants.’

b. yaong epseci-ess-e (야옹 없어졌어, TN 1;9)
   meow disappear-PAST-THV
   ‘The meow-meow disappeared.’ [translation changed by author to correspond to English child-speech]

At the telegraphic stage sentences start becoming more complex; however, many function words are missing. Often, inflectional morphology is absent at this stage. Here are some English examples.

(6) Telegraphic Stage in English

a. Cat stand up table.

b. What that?

c. Cathy build house.

Verbal morphology in Korean is much more complex than in English. Between 1;9 and 2;6 children first acquire -e (어/아) and -ta (다), which they use for requests and statements, respectively (Choi, 1991, 1995). They then use -e for old information and -ta for new information. At this time, they acquire -ci (치), which indicates shared information and -tay (대), which indicates indirect evidence or hearsay. Here are two examples.

(7) Acquisition of Evidential Markers in Korean

a. chaca-ss-ta
   find-PST-DECL
   ‘(I) found (it).’ (찾았다.)
b. say-ka ppai ha-n-tay  
bird=NOM bye do-PRS-EVID  
‘The bird says bye.’ (새가 빠이 한대.)

The acquisition of irregular forms in morphology goes through three stages (Marcus [1994] Marcus et al. [1992]). In the first stage, children produce irregular forms such as broke and ate. In the second stage, children overregularize the rules of morphology and produce forms such as breaked and eated. In Stage I the child imitates irregular forms but does not associate the irregular form with the regular form. Thus, the child does not associate broke with break. In Stage II the child has acquired the regular morphological form and applies it to irregular verbs. At this stage the child produces many overregularization errors. Finally in Stage III the child starts to acquire the irregular forms.

**Order of acquisition of irregular morphology:**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>imitation of irregular forms</td>
</tr>
<tr>
<td>II</td>
<td>overregulation</td>
</tr>
<tr>
<td>III</td>
<td>acquisition of irregular forms</td>
</tr>
</tbody>
</table>

Children also overgeneralize zero-derivational morphology in English. For instance the noun email can be used as a verb to email in English. Not all nouns can do this, however. Children sometimes overregularize and form the verb to broom from the noun broom. Eventually the verb to broom is replaced by to sweep.

Observe also the following acquisition errors in Korean morphology [Kim (1997)].

<table>
<thead>
<tr>
<th>age</th>
<th>produced form</th>
<th>English translation</th>
<th>target form</th>
</tr>
</thead>
<tbody>
<tr>
<td>2;2</td>
<td>배 부른대</td>
<td>‘He says he’s full.’</td>
<td>부르대</td>
</tr>
<tr>
<td>2;4</td>
<td>애티 두 먹은대</td>
<td>‘These children say they also eat.’</td>
<td>먹는대</td>
</tr>
<tr>
<td>2;4</td>
<td>이거 적은 거 어따 놓아</td>
<td>‘Where do I put the think that I dip this in?’</td>
<td>적는</td>
</tr>
<tr>
<td>1;11</td>
<td>없어 입예</td>
<td>‘It doesn’t have a mouth.’</td>
<td>입이</td>
</tr>
<tr>
<td>2;2</td>
<td>여기서 해밖에야 야</td>
<td>‘This place is a hamburger house.’</td>
<td>여기가</td>
</tr>
</tbody>
</table>

Wh-movement and auxiliary movement are acquired around 2;6 to 4;0; however, there is a great deal of variation from one child to the next. Here are some examples of child English showing lack of inversion, auxiliary copying, and double marking of tense (Weinberg [1990] Crain and Nakayama [1987]).

(8)   a. Did you came home?
      b. Did you did came home?
      c. Can he can look?
      d. What the mouse is doing?

### 10.3 Acquisition of More than One Language

**Simultaneous bilingualism** is the acquisition of two (or more) languages during the normal language acquisition period of childhood. This contrasts with adult L2 acquisition. Adults no longer acquire language in the
same way children do. Adults use more conscious cognitive abilities to acquire a second language than do children.

There are two competing theories of child L2 development. The unitary system hypothesis proposes that a single grammar and lexicon is present in the child. The child may start with this hypothesis, but evidence shows that children early on figure out that they’re acquiring two languages. In the separate systems hypothesis, children have distinct grammars and lexicons for each language.
Exercises

Question 1.

Question 2.
Further Reading

- **Yip and Matthews (2007)** - This monograph contains a detailed, longitudinal study of bilingual language acquisition of English and Cantonese.

- **Online sources:**
  - [https://www.youtube.com/watch?v=PU9yUa2roq0](https://www.youtube.com/watch?v=PU9yUa2roq0) - This website discusses regularization errors in English speaking children
  - [https://www.youtube.com/watch?v=OrQ0LfqxABM](https://www.youtube.com/watch?v=OrQ0LfqxABM) - This video is a refutation by Chomsky of behaviourism.
Chapter 11

Language Documentation, Endangerment, and Revitalization

By the end of this chapter you should:

• understand the issues related language death and endangerment
• be familiar with basic principles of language documentation
• understand the importance of language diversity
• be familiar with language revitalization efforts

11.1 Introduction

There are approximately 7000 languages in the world. Of those there are only ten with 100 million speakers or more. Here they are listed in order of number of speakers[1]

1. Mandarin - 873 million
2. Hindi - 370 million
3. Spanish - 350 million
4. English - 340 million
5. Arabic - 206 million
6. Portuguese - 203 million
7. Bengali - 196 million
8. Russian - 145 million
9. Japanese - 126 million

[1] These number include native speakers only. If second language speakers are included, then English rises to second place. Sources: CIA World Fact Book, Ethnologue (c), 2005.
This chapter deals with language documentation, language endangerment, and language revitalization.

### 11.2 Language Endangerment

There are on the order of 7000 languages spoken in the world today. The vast majority of these have between 1000 and 100,000 speakers. Vast numbers of languages face imminent extinction. Approximately one language dies every other week. Language is a vital component of human culture. Think of how your own language and culture are tightly linked. Think of the kinds of knowledge you possess and how it expressed in your native language. Vast sums of human knowledge are contained within the diversity of human languages, including knowledge about science, history, and the biological world. Many linguists and communities are in a race against time to document and even revitalize dying languages. We begin this chapter with a discussion of language endangerment. Some scholars prefer the term *language shift* to language endangerment. Indeed, the phenomenon of language loss is typically due to speakers of a minority language shifting to a culturally dominant language. Languages can also become endangered in response to natural disasters or epidemics. Furthermore, most programs related to language revitalization are developed in response to language shift in particular.

A common misconception about language endangerment is that the number of speakers of the language is the primary indicator of the level of endangerment. It is true, of course, that endangered languages have fewer numbers of speakers in general, but there are other factors to consider. For instance, Breton, spoken in the northwest of France has approximately 250,000 speakers; however, children are not actively acquiring the language and have instead switched to French. On the other hand, Pirahã, a language spoken in the Amazon Basin, has fewer than 500 speakers. It is spoken by all members of the community, though, including children. Many speakers are monolingual. Only some members of the society have acquired Portuguese to conduct business with members outside the community. This language is quite healthy as long as the community continues to exist. The following four properties are commonly used to assess language endangerment.

1. **Intergenerational Language Transmission**
2. **Absolute Number of Speakers**
3. **Proportion of Speakers within the Total Language**
4. **Shifts in Domain Language Use**

The first property, *intergenerational language transmission* has to do with the rate at which the younger generation acquires the language from the older generation. With healthy languages there is a complete transmission from one generation to the next; whereas with languages in danger, younger generations do not fully acquire the language. Navajo, a Dene-Yeniseian language of central United States, is spoken by over 100,000 people. It is considered one of the healthier aboriginal languages in North America. It is still being acquired by
the majority of Navajo children. Nevertheless, it is not acquired by all Navajo children. Tuscarora, an Iroquoian language in central eastern North America lies at the opposite side of the spectrum with only a few speakers, all whom are in their seventies.

As mentioned above, the absolute number of speakers in and of itself is not a reliable indicator of the relative health of a language; however, languages with small numbers of speakers are more vulnerable to natural disasters and epidemics. They are also more prone to cultural encroachment of larger neighbouring societies.

A more reliable indicator of the health of a language is the proportion of the number of speakers of a language to the total number of members of the society. Sami is a Uralic language spoken across the far northern stretches of Europe. Approximately 70% of the population speak Sami, indicating the language is in early stages of shift. Ainu, a language isolate spoken in northern Japan, on the other hand is spoken by only 10 people out of a population of 25,000 or approximately 0.04% of the population. By some estimates, however, the actual population of Ainu is much larger than 25,000. Thus, Ainu is severely endangered, although revitalization efforts are underway.

Finally, shifts in the domain of language use indicate the relative health of a language. The healthiest languages are used universally in all domains, including home, work, education, entertainment, and ceremonies. Incipient kinds of language shift include change to a more dominant language at institutions of higher education. After several successful years of revitalization, Cornish is used at home by many speakers, but the language of higher education is still English. Hawaiian is also used primarily only at home, although it is also the language of instruction in lower grades at school. With highly endangered languages their use is often restricted to religious or ceremonial domains. Manchu, spoken in the north of China has very few speakers, all of whom are elderly. The language is largely restricted to a few ceremonial practices, although there has been recent interest in reviving the language.

UNESCO classifies endangered languages as follows. Note that some scholars prefer the term sleeping language to extinct language as languages can be revived even when there are no living speakers (Leonard, 2011, 2008).

1. Levels of language endangerment
   a. Safe: language is spoken by all generations; intergenerational transmission is uninterrupted
   b. Vulnerable: most children speak the language, but it may be restricted to certain domains (e.g., home)
   c. Definitely endangered: children no longer learn the language as mother tongue in the home
   d. Severely endangered: language is spoken by grandparents and older generations; while the parent generation may understand it, they do not speak it to children or among themselves
   e. Critically endangered: the youngest speakers are grandparents and older, and they speak the language partially and infrequently
   f. Extinct: there are no speakers left

11.2.1 Causes of Endangerment

Thomason (2015) discusses the following causes of language endangerment. We discuss these in turn. The following text is freely adapted from Thomason’s discussion.

1. Conquest

2. Economic Pressures

3. Melting Pots

4. Language Politics

5. Attitudes

6. Standardization

Conquest and colonialism is one of the major factors in language endangerment or language shift in the world. As one society invades or colonizes another, the language of the invading society largely displaces that of the conquered society. Sometimes, specific policies are enacted to enable the demise of other languages. In North and South America, English, French, Spanish and Portuguese have driven out vast numbers of indigenous languages, as has English in Australia. In Northern Asia, Russian has driven out numerous indigenous languages across Siberia.

A common source of language shift is economic pressures. In order to fit into the dominant society, specifically, to get a job and earn a living, people feel pressure to speak the language of the dominant society.

11.3 Language Documentation

Any instantiation of language, be it a spoken language or a signed language is fleeting. Language documentation seeks to create a lasting, multi-purpose record of a language that can be used by future generations. Furthermore, as we saw in the introduction, languages are not monolithic, invariable entities. There is dialect variation from one community to the next. Dialects are typically part of a dialect continuum. Within a given population there is also speaker variation based on age, gender and other sociolinguistic variables. Ideally, a complete language description should include all such variation. At the initial stages of description decisions have to be made as to what to cover. In the early 20th century, a good description consisted of the following three items: (i) a grammar, (ii), a dictionary, and (iii) examples of texts.

Modern technological advancements allow for the documentation of a much broader array of linguistic phenomena. Both audio and video recording allows us to document prosody and gestures. Prosody, in particular, has been implicated in encoding linguistic properties. Some languages have obligatory rising intonation for polarity questions (yes/no questions) for example. Woodbury (2003) proposes a body of documentary work should have the following properties (slightly adapted from the original text).
1. **diversity**: The corpus should contain samples of language use across a range of genres and socio-cultural contexts, including elicited data.

2. **quantity**: Given modern storage and computational technology, a corpus should include as much media and text as possible.

3. **expandability**: The corpus should be ongoing, distributed and opportunistic. New data and sources should be able to be easily added to the corpus.

4. **transparency**: The corpus should be able to be read by other users, including future researchers.

5. **preservability** and **portability**: The corpus should be prepared for long-term preservation and be adaptable to changing technology.

6. **ethical**: The data must be collected, analyzed and stored according to established ethical protocols and with full cooperation and participation from the community from which the data is gathered.

### 11.4 Language Revitalization

Successful language revitalization involves community effort. Language professionals cannot accomplish the task of revitalizing a language alone. As a metaphor, consider a student in a language class. No matter how competent the language teacher is, the student will only acquire a second language if they really want to. Likewise, the community must want to awaken their language if revitalization is to be successful. Most important is the willingness on the part of adults in the community who are willing to learn the language and act as language mentors. This is especially important in communities where the fluent speakers are mostly elderly and may have health and mobility issues.

**Language nests** originated in New Zealand in 1982 with the revitalization of Māori. They involve an elder or elders fluent in the language being revitalized working with young children in the community (King, 2001). Thus, languages nests aid in intergenerational transmission. Figure 11.1 shows a language nest run by the Miriwung people in Australia. Given that exposure to a language during early childhood greatly enhances acquisition of that language, language nests have had a great deal of success. As a result, language nests are now used in communities all over the world to help save highly endangered languages.

### 11.4.1 Case Study: Cornish

Cornish is a Celtic language spoken in southwest England. The last monolingual speaker of Cornish, Dolly Pentreath, died in 1777. Before its revival, the last fluent speaker died in 1891, however reports of other speakers into the 20th century exist. The decline started shortly after Norman conquest as more and more...
English moved west into Cornwall. Language policies dictated use of English for church services, rather than local languages or Latin. This was partly due to switch from Catholicism to Protestantism. Revitalization began as early as 1904, but was slow to start. At first, a “Unified Cornish” was promoted, which was a standardized form of various Cornish dialects. A number of factors gave rise to a slow start in revitalization: (i) a lack of interest on the part of the public to revive Cornish, and (ii) a dissatisfaction with the standardized form. The proposed standarized form sounded too stilted and artificial. A common point of disagreement is the level of standardization required, especially for the written form.

Cornish now has official recognition as minority language of Europe. It is currently taught in schools and in universities (as a second language, though, not as language of instruction). There are currently 3500 speakers with some knowledge of Cornish, including about 100 fully fluent speakers and 500 daily users.

11.4.2 Case Study: Mohawk

Mohawk is an Iroquoian language spoken in central North America. Jacobs (1998) describes the following chronology of events for the revitalization of Mohawk. The Mohawk Immersion Program in Kahnawà:ke provides another example of an ongoing, successful revitalization program. Revitalization efforts started in the 1970s in Quebec, Canada and is one of the most successful revitalization attempts in Canada. Factors of success include the following: (i) dedicated people who enlist community participation, (ii) successfully navigating different attitudes towards schooling (traditional European system versus Iroquoian system), and (iii) the establishment of literacy standards. The impetus for revitalization was the extremely high rates of language loss detected in the 1950s. The revitalization efforts started off small. There was about 15 minutes of language instruction in Mohawk per day, which was later increased to 30 minutes per day. The first teachers were not
trained in pedagogy, but were simply speakers of Mohawk. They received a small salary from Department of Cultural Affairs for their work.

Various road blocks impeded revitalizations, notably Bill 101 in Quebec. Bill 101 was enacted to protect French in Quebec, but was detrimental to the revitalization efforts of Mohawk. Crucially, Bill 101 dictated that all public education be conducted in French. This led to greater determination on the part of the community to protect the Mohawk language (and Mohawk sovereignty). The school system gradually became more autonomous. They worked with Quebec Ministry of Education, but handled most affairs internally. They developed their own sports and arts programs and established a Mohawk Language Centre. Language immersion began in 1979 with nursery school children, in which Mohawk was sole language of communication. Language immersion gradually increased to higher grades. Both Mohawk and French classes are taught, most classes are in Mohawk.

The development of pedagogical materials was slow at first, but slowly improved. Orthography and grammar books were developed in mid 70’s. A permanent office was created to oversee the creation of new pedagogical materials for Mohawk, including the retention of cultural values and practices, but also the preparation of students to live in a modernizing world both within and outside the Mohawk community.

The revitalization efforts have created a middle generation of non-speakers. The grandparents and grandchildren speak Mohawk, but parents speak only English and/or French (mostly English)
Exercises

**Question 1.** Conduct some research on the Jeju language and determine what risk factors play a role in the endangerment of Jeju.

**Question 2.** Do you think minority dialects (of Korean or of any language) face the same issues as other endangered languages?

**Question 3.** Using the UNESCO Endangered Languages Atlas, learn about an endangered language in one of the world’s hot spots for language loss.
Further Reading

- **Crystal (2002)** is an insightful layperson introduction to the issues of language death
- **Kiaer (2014)** and **Yang et al. (2019)** are excellent discussions of the Jeju language
- **O’Grady et al. (2017)** is an introductory textbook to learn Jeju (in Korean).
- **Thomason (2015)** is an excellent textbook style introduction to the study of endangered languages.
- **Austin and Sallabank (2011), Rehg and Campbell (2018), Hinton (2001), Hinton et al. (2018) and Grenoble and Furbee-Losce (2010)** are excellent handbooks dealing with language endangerment, documentation, and revitalization
- **Brenzinger (2008)** is an excellent academic discussion of language loss and the loss of linguistic diversity
- **online resources:**
  - Endangered Languages Documentation Program [https://www.eldp.net/](https://www.eldp.net/)
Chapter 12

Psycholinguistics and Neurolinguistics

By the end of this chapter you should:

- understand basic properties of processing
- be familiar with basic experimental techniques in psycholinguistics
- understand semantic priming

12.1 Introduction

Psycholinguistics is the study of how the mind processes language. This contrasts with neurolinguistics (in the next chapter) which studies how the brain processes language. Language is not a stored array of fully formed sentences. As we have seen, there is an infinitely large number of possible sentences. How does the mind form sentences? How is vocabulary stored in the mind? What goes on in the mind when we try to process a garden path sentence? These are some of the questions psycholinguists ask.

(1) a. The horse raced past the barn fell.
    b. The cotton clothing is made of comes mostly from India.

There is a distinction between competence, the mental capacity to use and understand language, and performance, the execution of competence in any real-world situation. Consider the following quote by Chomsky (1965).

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its (the speech community’s) language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of this language in actual performance.

Performance includes speech errors, ambiguity resolution based on pragmatics, word/vocabulary recall, the
resolution of incomplete sentences among many other properties. It is the use of language in the real world. Some psycholinguists and sociolinguists suggest that the distinction between competence and performance should be collapsed. Whether the mind or brain processes linguistic competence differently from linguistic performance is an ongoing debate.

Neurolinguistics is the study of how language is related to brain function. Neurolinguistics involves interdisciplinary research with cognitive science, medicine, biology, and traditional linguistics. Biolinguistics is the study of the biological aspects of language. This includes the genetic underpinning of language and language correlates in non-human species. Biolinguistics is the topic of the following chapter.

In the 19th century in Europe there were numerous debates on how the brain functions. These debates hinged on whether the brain had discrete components for managing different tasks (functional specialization) or worked as an undifferentiated whole. Paul Broca and Carl Wernicke noticed that individuals with language impairment had lesions in a specific part of the left hemisphere of the brain. Broca’s and Wernicke’s studies were among the first that offered physical evidence for functional specialization. Figure 12.1 shows a lesion in the brain of one of Broca’s patients, Tan.

As these early studies concentrated on speech deficits, namely, aphasia (실어증), let’s take a moment to understand what aphasia is and what kinds of aphasia there are. Aphasia is a condition resulting from brain injury which impairs language use. There are two main classifications for aphasia. Individuals with Broca’s aphasia (aka non-fluent aphasia, 브로카 실어증) have immense difficulty in speaking. They exhibit drastic loss of both lexical items (cat, dog, apple) and functional items (the, is, does, etc.) as well as loss of many grammatical abilities. Here is an example.

---

1 image source: https://doi.org/10.1093/brain/awm042
In Wernicke’s aphasia (aka fluent aphasia, 베르니케 심어증) full sentences are produced with much the same grammatical principles as normal individuals, but are typically meaningless. Consider the following example.

(3) “You know that smoodle pinkered and that I want to get him round and take care of him like you want before.”

We will return to the properties of aphasia later. First, we will introduce the structure of the brain, concentrating on those aspects crucial to language.

12.2 Speech Errors

Spoonerisms (두음 전환) were mentioned in Chapter 4 in the discussion of syllable structure. A spoonerism is the accidental transposition of the initial sounds of two syllables. Here are some examples.

(4) **English Spoonerisms**

   a. You have tasted a whole worm. (from: You have wasted a whole term.)

   b. a well-boiled icicle (from: a well-oiled bicycle)

   c. Is the bean dizzy? (from: Is the Dean busy?)

The **Korean translation of Queen’s Bureau of Investigation** by Ellery Queen contains the following example.

(5) 그리고 ‘고먼’이라는 이름에서 가능한 두음 전환은 ‘모건’뿐이니 까... (And since the only spoonerism possible from the name ‘Gorman’ is ‘Morgan’...)

In Chapter 4 spoonerisms were taken as evidence that the onset and the rhyme of the syllable are two distinct constituents.
12.3 Processing

Processing refers to how the hearer forms a mental representation of a sentence upon hearing a linear string of sounds and intonational contours. **Top-down processing** assumes that the hearer starts with the meaning and syntactic structure, and breaks the sentence down into smaller, manageable pieces. **Bottom-up processing** assumes that the hearer starts with the sounds, and pieces them together into meaningful units. It is believed that both strategies are involved in real life processing. Evidence that hearers use top-down processing is found in a listening experiment. Full sentences heard in the presence of noise are easier to understand than a string individual words that do not form a sentence. Thus we get clues from the structure and not just the sounds.

Models of sentence processing involve building structure mentally, and then going back to re-check when ambiguity arises (Frazier, 1987). Consider the following example.

(6) The warehouse fires...
   a. ... were set by an arsonist.
   b. ... employees over sixty.

Upon hearing "The warehouse fires" the hearer starts to build structure; however, the beginning of this sentence is ambiguous. Eye-tracking experiments have revealed that the reader’s eyes will go back to the part of the sentence where the ambiguity starts. Two principles of sentence processing are minimal attachment and late closure. **Minimal attachment** states that a new word or phrase is attached in such a way that minimizes structure. **Late closure** states that a new phrase is attached to the phrase currently being worked on. Here are some examples.

(7) The horse raced past the barn fell.

When we hear, ‘the horse’, we assume this is a noun phrase. Then we hear ‘raced’. It is easier to parse ‘raced’ as a verb with ‘the horse’ as its subject.

(8) \[
\begin{array}{c}
\text{NP} \\
\text{Det} \\
\text{the} \\
\text{V} \\
\text{raced} \\
\text{S} \\
\text{NP} \\
\text{Det} \\
\text{the} \\
\text{N} \\
\text{horse} \\
\text{V} \\
\text{raced} \\
\text{VP} \\
\end{array}
\]

Minimal attachment leads us to attach ‘raced’ as a verb in a sentence with ‘the horse’ as the subject. Of course this is eventually the wrong parse, giving rise to the garden-path effect. Eventually we form the following parse for (7).
(9) (10) John said that Mary was fired yesterday.

In the following Korean sentences, consider where the indirect object is attached.

(11) a. 어머니가 예쁜 막내딸에게 삼촌이 사준 차를 좋아했다.
   b. 어머니가 예쁜 막내딸에게 삼촌이 사준 차를 물려주었다.

Here are the approximate English translations.

(12) a. Mother likes the car that her pretty, youngest daughter was given by her uncle.
   b. Mother bequeathed the car that her uncle gave her to her pretty, youngest daughter.

The PP 예쁜 막내딸에게 is the indirect object of the main verb in the second sentence but it is the indirect object of the relative clause in the first sentence. The principle of minimal attachment prefers to place the indirect object in the main clause; however, the main verb 좋다 (‘to like’) doesn’t allow for the presence of an indirect object, so the structure must be recalculated.

12.4 Psycholinguistic Experimental Methodology

One of the earliest psycholinguistic experiments was the lexical decision task. An early use of this test is found in [Meyer and Schvaneveldt (1971)]. The test subject simply indicates whether a string of letters shown on a screen is a word or not. The response time is the length of time for the test subject to respond ‘yes’ or ‘no’. The following figure shows various examples of what might appear in such an experiment.

More common words have a faster reaction time. Thus dog has a fast reaction time while canine has a slower reaction time as it is an uncommon word. If a nonce word (a made-up word) is phonotactically possible it has a slower reaction time. If a nonce word is phonotactically impossible it has a faster reaction time. Thus, blick has a slow reaction time while ftork has a faster reaction time. Blick is a possible word in English, but it happens to not exist. Ftork, however, is impossible in English.

If the test subject is shown the word dog and then is later shown the word canine, the reaction time for canine
decreases substantially. We say that the word *dog* semantically primes the word *canine* as their meanings are so similar. The semantic priming potential of ambiguous words can also be investigated. For example, the word *calf* is ambiguous upper leg (종아리) and baby cow (송아지). We may ask if there is there a priming effect with either “leg” or “cow”. To calculate the priming effect we use the following materials:

Present “leg” for 50-100 ms
Present “calf” for 2000 ms
Test subject gives response A

Present “cow” for 50-100 ms
Present “calf” for 2000 ms
Test subject gives response B

The difference in response time is the difference in the priming effect of “leg” and “horse” with respect to *calf*. More recent investigations have shown that priming effects of ambiguous words presented in a disambiguating context continue to have a priming effect for 20 minutes (Rodd et al., 2013).

Psycholinguists also use *electroencephalography* (EEG, 뇌전도) to examine the interaction between brain waves and language, as shown in Figure 12.3. An irregular linguistic stimulus causes an event-related potential 400 ms after the stimulus is perceived. Specifically, it causes a negative spike in brain wave activity, so is called N400.

Consider the following examples. The first example is normal and causes no N400 spike. The second example is pragmatically odd and so causes an N400 spike. The last example is pragmatically and syntactically ill-formed (*cry* is intransitive and cannot appear with a direct object). The N400 spike is more pronounced.
Figure 12.3: Test Subject wearing electrodes for an EEG experiment

\(\text{The pizza was too hot to}\)

\[\text{ay} \quad \text{drink}\]

\[-1 \text{µV}\]

\[+5 \text{µV}\]

\[0 \quad 400 \quad \text{ms}\]

Figure 12.4: N400 spikes of three experimental sentences

(14) a. The pizza was too hot to eat. - no N400 spike  
b. The pizza was too hot to drink. - N400 spike  
c. The pizza was too hot to cry. - strong N400 spike

One last experimental procedure we will examine considers the acquisition of pragmatic competence. Recall that quantifiers induce scalar implicatures as in the following example.

(15) Some of the dogs are eating.

This sentence implicates that not all of the dogs are eating. It turns out children acquire the pragmatic implicature rather late. In other words, children evaluate the sentence *Some of the dogs are eating* logically and assume this is consistent with a situation in which all of the dogs are eating (Noveck, 2001). The experiment is conducted as follows. Test subjects are shown pictures with an accompanying sentence and are asked to evaluate whether the sentence is true or false. The sentence in (15) is the test sentence and the two pictures of interest are shown below.
Figure 12.5: Test Pictures for Scalar Implicature Experiment
Figure 12.6: The two hemispheres of the brain

For the first image both children and adults respond that (15) is true. For the second image, however, adults typically respond that (15) is false, while children respond that it is true.

12.5 Brain Structure and Language

The nervous system consists of two main divisions. These are the **Central Nervous System** (CNS, 중추 신경계), which consists of the brain and the spinal cord, and (ii) the **Peripheral Nervous System** (PNS, 말초 신경계), which interfaces with the CNS and the rest of the body. The PNS is subdivided into the Somatic Nervous System (SNS), which is responsible for volitional control of muscles and sensation, and the Autonomic Nervous System (ANS), which is responsible for involuntary reactions and automatic activities of the body, such as breathing and heart rate. The basic, cellular unit of the nervous system is the nerve cell, or the **neuron**.

The brain is divided into two hemispheres connected by the **corpus callosum** (뇌량). Each hemisphere is divided into four lobes. The frontal lobe (전구엽) is separated by the central sulcus (중심고랑) and the lateral sulcus (가촉고랑). The frontal lobe is responsible for planning, prediction, speech and discrete body movements. The temporal lobe (측두엽) is separated by the lateral sulcus. It contains the auditory cortex and is responsible for audition, memory processing, and sensory integration. The parietal lobe (두정엽) is separated by the central sulcus and part of the lateral sulcus. It is responsible for reading, tactile sensations, and pain. Finally, the occipital lobe (후두엽) is located behind the temporal lobe and the parietal lobe. It is responsible for visual processing.

Many physiological functions are lateralized to one side of the brain or the other. Language is left-lateralized in about 90% of the population. There is an abundance of physical evidence for lateralization. The linguistic evidence for this is Broca’s area and Wernicke’s area. Recall that Broca and Wernicke discovered lesions in specific areas of the left side of the brain in aphasics. Another piece of evidence related to language is found in the the planum temporale. This region of the brain is larger in the left hemisphere than its corresponding area in the right hemisphere. This difference is more pronounced in fetal brains (31 weeks), suggesting a preparation for language acquisition.

Each half of the brain receives input from and sends signals to the opposite side of the body. This is known
as contralateralization. As figure 12.7 shows, information is transmitted from one hemisphere of the brain to the other through the corpus callosum.4

Patients with damage on one side of the brain may suffer seizures on the opposite side of the body. In severe cases of epilepsy, the corpus callosum (the bundle of fibres linking the two halves of the brain) is severed (split-brain individual).

In split-brain individuals, information received on one side of the brain cannot be relayed to the other side of the brain. If an image or word is shown to the left field of vision of a split-brain individual, this information is received only by the right side of the brain (due to contralateralization). This information is not sent to the left side of the brain (where language is processed – lateralization). Thus, split-brain individuals cannot name objects or read words presented to the left field of vision, although they still recognize objects and know what they are used for.

The brain is divided into four lobes as shown in Figure 12.8. The frontal lobe is responsible for planning, prediction, speech, and discrete body movement. It is separated by the central sulcus and the lateral fissure. The temporal lobe is responsible for audition, memory processing, and sensory integration. It is separated by the lateral fissure and contains the superior temporal gyrus, the auditory cortex. It is the left temporal lobe (in right-handed individuals) that contains Wernicke’s area, which is thought to be tightly associated with language. The parietal lobe is responsible for reading, tactile sensations, and pain. It is separated by the central sulcus and the lateral fissure. It contains the angular gyrus, which plays a role in reading. Finally, the occipital lobe is responsible for visual processing. It is located behind the angular gyrus.

Many physiological functions are lateralized to one side of the brain or the other. Language is left-lateralized
in about 90% of the population. The evidence for left-lateralization is the following. First, Broca and Wernicke discovered lesions in specific areas of the left side of the brain in aphasic individuals. Second, the planum temporale is larger in the left hemisphere than its corresponding area in the right hemisphere. This difference is more pronounced in fetal brains (31 weeks), suggesting a preparation for language acquisition.

Dichotic Listening Research has also provided some clues into how the brain processes language. Test subjects are given different stimuli to each ear via headphones. For example, the left ear hears ‘base’ and the right ear hears ‘ball’. It is found that speech stimuli given to the right ear (which sends information to the left side of the brain) has an advantage. This is known as the Right Ear Advantage. It has been found that most language-related stimuli (including such things as Morse code) are susceptible to the Right Ear Advantage, but non-linguistic vocalizations are not (such as laughing and coughing).

12.5.1 Aphasia

Language functions in the left hemisphere of the brain are not uniformly distributed. Damage to any one specific area does not result in loss of all language function. Crucially, localized damage to a small area of the brain may result in the loss of some language function, but not all language function. Aphasia is any language impairment brought on by brain damage; however, there is a lack of uniform classification of types of aphasia. There are three widely accepted aphasia syndromes, however.

(16) Aphasia
   a. Broca’s aphasia
   b. Wernicke’s aphasia
   c. Conduction aphasia

Broca’s aphasia was originally assumed to be the result of a lesion in an area of the frontal lobe now known as Broca’s area. The cluster of symptoms associated with Broca’s aphasia typically results from more extensive
damage, however. Broca’s area is close to the region of the brain that controls muscles involved in speech. Symptoms include the inability to speak clearly and fluently, speech being extremely laboured, and the lack of function words and inflectional morphology. Once a string of words is finally uttered, it usually makes sense in the context of the conversation, however. It is believed that individuals who suffer from Broca’s aphasia can understand language input (speech uttered to them), there is some loss in comprehension. Finally, Broca’s patients are quite aware of their condition and the mistakes they make in speech and writing. This is a crucial difference with Wernicke’s aphasia.

Wernicke’s aphasia results from a lesion in the temporal lobe, specifically in Wernicke’s area. Wernicke’s area is near the auditory cortex. Sufferers typically have severe loss of comprehension, although hearing itself is not affected. Symptoms vary, but usually include fluent speech with some minor problems related to word recall, and lack of clear meaning or any meaning at all in their utterances. Often sufferers will substitute phrases or words for individual lexical items that cannot be recalled. In severe cases, the patient will simply uses a nonce word. Crucially, patients who suffer from Wernicke’s aphasia are typically unaware of their condition. It is thought that Broca’s aphasia affects phonological systems, while Wernicke’s aphasia affects syntactic and semantic systems.

Conduction Aphasia results from lesions between the temporal lobe and the parietal lobe. It is thought that these are areas that are specifically responsible for associating meaning with form (sound). Patients exhibit fluent speech, but circumlocutions are frequent and there are aberrations in syntactic structure. Comprehension of oral and written material is only mildly affected. Conduction aphasia is thought to result from disruptions between the centres that control sound systems (phonology) and the centres that control structure and meaning (syntax and semantics, respectively). Conduction aphasia is much rarer than Broca’s and Wernicke’s aphasia.

| Broca’s Aphasia: | Non-fluent aphasia  |
|                 | speech is laboured  |
|                 | loss of lexical and grammatical vocabulary |
|                 | awareness of condition |
| Wernicke’s Aphasia: | Fluent aphasia  |
|                 | speech is generally free flowing |
|                 | loss of semantic coherence |
|                 | lack of awareness of condition |

12.5.2 Specific Language Impairment and Other Disorders

*Specific language impairment* (SLI) is the name given to any kind of abnormal development of language ability in a child. As this description indicates, it is not a single specific condition, but rather a family of conditions. Approximately 7% of the population experiences some kind of SLI. It presents in a variety of way and is caused by a variety of factors, including genetic and cognitive.
Exercises

Question 1.

Question 2.
Further Reading

- online resources
  - Speech Error Database - https://www.mpi.nl/dbmpi/sedb/sperco_form4.pl
Chapter 13

Biolinguistics

By the end of this chapter you should:

• understand basic biological underpinnings of language
• have an appreciation for language correlates in other species
• be familiar with the FOXP2 gene

13.1 Introduction

Biolinguistics is the study of biological aspects of language. It is an interdisciplinary field of study that combines biology and linguistics.

13.2 A Biological Foundation for Language?

The evolution of language has been dubbed “the hardest problem in science” by some (although this title has often been given to other problems in science such as “What is consciousness?”). How did it evolve? What kind of genetic mutations made it available in humans but not in other species? Are there correlates or precursors to language in other species? These are just some of the questions on the role of biology in language.

Let’s begin by considering how the vocal apparatus in humans differs from that in other species. For instance in most mammals the tongue is long and flat. In humans it is nearly circular in the middle. The shape of our tongue and our ability to control it allows us to produce vowels such as [i] [a] and [u], which primates cannot do. The human larynx is also descended, giving us a larger pharynx (Negus 1929; Ghazanfar and Rendall 2008). This allows us to produce a broader range of sounds with more clarity.
13.3 A Genetic Basis for Language?

The first genetic link to language was established by studying the KE family, a British family of Pakistani origin, several members of whom are afflicted with a particular form of specific language impairment.

Each cell in the body contains a nucleus with a full set of chromosomes (염색체) needed for the life of the organism. In humans, each body cell contains 23 pairs of chromosomes containing all the DNA necessary for our existence, although only a tiny fraction of the DNA is expressed or activated in a given cell. For example, a skin cell activates those parts of the DNA necessary for the production and function of skin, but not of, say, the liver. The DNA on the chromosomes is divided into functional units called genes. Genes, in combination with the environment, are responsible for the expression of the characteristics of the organism. A gene is responsible for the expression of a protein that exerts an influence of the organism. The set of all genes in an organism is called the genome. Each chromosome contains a particular set of genes. For convenience, in humans the first 22 pairs of chromosomes are simply labelled 1 to 22 and are referred to as the autosomes. The 23rd pair is known as the sex chromosomes and usually called X and Y. Typically, human males have the combination XY and females typically have the combination XX, although the correlation is not absolute and there are variations.

Genetic material in humans (and other organisms that reproduce sexually) is inherited from both parents. Recall that each body cell in humans contains 23 pairs of chromosomes. Sex cells contain only 23 single chromosomes. When a spermatocyte (the sex cell produced by males, more commonly known as a sperm cell) fertilizes an oocyte (the sex cell produced by females, more commonly known as an egg cell), the genetic material of the two combine to form 23 pairs of chromosomes. Each member pair of chromosomes carries the same genes, although the specific information encoded on the genes is different. As an analogy, imagine you have 23 pairs of textbooks: one pair for linguistics, one pair for chemistry, one pair for Swahili, etc. Each member of the pair is by a different author, though. Both members of the pair will contain the same kind of information, but the information is encoded differently.

All members of a species essentially have the same set of genes, but the particular versions of the genes differ among individuals. These versions are called alleles. The set of alleles an individual has is called the genotype. Correspondingly, the set of characteristics an individual has is called the phenotype. A well-known example of these concepts is blood type in humans. There are three phenotypes: A, B, AB, and O. (There are actually finer gradations such as A+ and B-, etc., but these do not concern us here.) The genes controlling blood type are found on chromosome 9. Each chromosome has one of three alleles for blood type: A, B or O. Since humans have two copies of each chromosome, there are six possible genotypes, with the corresponding phenotypes listed below. How the genotypes give rise to the observed phenotypes is not important here.

A common misconception about genes is that there is a one-to-one correspondence between genes and characteristics. In fact, a given gene can (of typically does) affect numerous characteristics, and a given characteristic is typically affected by several genes. Moreover, there are many genes that affect the expression of other genes. Thus, the correspondence between genotypes and phenotypes is both many-to-one and one-to-many.

With advanced studies of genetics and the human genome geneticists and linguists have been investigating
possible genetic underpinnings of language. The famous *FOXP2* gene plays a central role in these discussions. Sometimes dubbed “the language gene”, it is actually found in numerous other species – both animals and plants, including yeast. Keep in mind in the following discussions that genes in non-human species are written in lower case letters: *foxp2*. Given that this gene appears in so many species, including those distantly related to humans, it is clearly not related specifically to language.

The first genetic link with language is *FOXP2*, a member of the FOX gene family. *FOXP2* is a ‘regulatory gene’; it regulates the expression of other genes. (In other words, it is responsible for turning other genes off and on.) *FOXP2/foxp2* is highly conserved across many species (from yeast to humans). It is involved in brain and lung development and the development of the part of the brain used for bird songs. The KE family in Great Britain was found to suffer from a specific kind of language impairment that has the inheritance pattern as a single gene ([Gopnik and Crago 1991](#)). Researchers found a specific mutation on the *FOXP2* gene. The description of the impairment has been controversial, but it either affects individuals fully or not at all. fMRI scans indicate various abnormalities in several parts of the brain implicated in speech and language processing, suggesting that *FOXP2* is important in the development of these areas.

In various linguistic studies conducted by Myrna Gopnik, she found specific grammatical errors ([Gopnik and Crago 1991](#)).

(1) **Grammatical Errors in the KE Family**

a. You got a tape recorders.
b. I find a cops.
c. I was make 140 box.
d. He only got two arena.
e. You make one points.

Observe the number agreement errors and the error in use of progressive.

Neurophysiological studies identified general problems with facial muscle use. Problems afflict non-linguistic as well as linguistic phenomena, such as tongue movement. Patients also exhibited problems in reproducing tapped rhythms ([Vargha-Khadem et al. 2005](#)).

Exactly how *FOXP2* affects language development is still unknown. It may be directly related given its
role in bird songs. It may related to motor control of vocal apparatus and not to language *per se*. Recall that FOXP2 is a regulatory gene. One gene it regulates the expression of is CNTNAP2, which encodes neurexin. Neurexin is expressed during the development of the human cortex. Specific mutations in CNTNAP2 are found to be correlated with a specific kind of heritable language impairment, the repetition of nonsense words. It is also been recently discovered that FOXP2 itself is regulated by other factors. Thus, the exact role of FOXP2 in language is far from clear.
Exercises

Question 1.

   Question 2.
Further Reading

- [Jenkins (2013)] - A brief introduction to the history of biolinguistics
- [Moro (2015)] - This general interest book explores generative linguistics from a biolinguistic perspective.
Glossary

Accusative - (ACC) a grammatical case usually assigned to the grammatical subject of a clause. See p. 168.

Adjunct - A phrase that modifies a predicate, but is not required by the predicate to be complete. Common verbal adjuncts are AdvP and PP.

Argument - An item (usually a noun phrase or a clause) required by a predicate. Arguments of a verb include the subject, direct object, and indirect object.

Constituent - A linear string of words that behaves as a unit as diagnosed by various syntactic tests. See p. 129.

Dual - (DU) a grammatical number that refers to two units.

Grammatical argument - The grammatical relationship between an argument and its predicate. For example the grammatical subject is the entity that triggers agreement on the verb. See p. 168.

Illocutionary Force - The way in which the speaker intends the addressee to accept the proposition. For example, the speaker could be telling the addressee something or asking the addressee something. See p. 155.

Infix - An affix that is added inside a root rather than to the right or left of it. See p. 34.

Logical argument - The argument that is semantically understood to hold a grammatical relation to a predicate. For instance the logical subject is the entity understood as the agent or instigator or the event described by the predicate. See p. 168.

Metathesis - the transposition of two sounds or syllables. See p. 239.

Nominative - (NOM) a grammatical case usually assigned to the grammatical subject of a clause. See p. 168.

Phrase Structure - a system for representing structure and constituency in a sentence. See p. 128.

Polarity question - A yes/no question. A question that requires either ’yes’ or ’no’ as an answer.

Proform - a word or morpheme that substitutes for a constituent. See p. 135.

Proposition - An expression that has a truth value. See p. 155.
Answers to Exercises

Chapter 1

Chapter 2

Question 1:

Adj

Pref

un-

Suf

V

relly -able

Chapter 3

Question 1: [kʰæt], [dʊg] or [dʊɡ], [mɑʊs] or [mɑʊs], [wɜr] or [wɜr], [sæn], [lʌv], [ðɪs], [bɜrd], [dʒʊn], [rɪm] or [rɪm], demonstrative (that book) [ðæt], complementizer (she knows that...) [ðæt] or [ðæt], [dʒæmpt], [staʊ],

or [staʊ], [giif], [dɑv], [lʌv], [get], [ski], [dɜrd]

Question 2: [pʰæn], [mul], [ara], [alja], [manuul], [pʰɑmaʃma], [k*ɛмι], [ɪði*wa], [s*odʒɪ], [kɪmp*ap]’

Question 3: Alternative segments shown in parentheses - [aɪma(n)caro(ɪ)s], [dʒæf], [dɜrdli], [bɛldʒæmndən],

[ɑnʃα(λ)kɛpʰidɪa], [ma(λ)kɛsko(ə)up], [tʰɛlaf(o)φn], [tʰɛntf(ɑ)k], [frɛstənɛ₁], [fəo(ə)ɡfæs]

Question 4:

1. k

2. (a) a - It is the only low vowel. The others are all high

(b) y - It is the only round vowel. The others are all unrounded.

3. p - It is the only voiceless sound. The others are voiced

4. k - It is the only plosive (oral stop) the others are fricatives.

5. (a) s - It is voiceless. The others are voiced

(b) d - It is a plosive. The others are fricatives.
(c) v - The primary articulator is the lower lip. For the others it is the tongue.

6. (a) p - It is voiceless. The others are voiced.

(b) m - It is nasal. The others are oral.

7. d - It is a plosive. The rest are nasal.

8. (a) r - It is lax. The rest are tense. (if you are considering tense/lax to be an independent phonetic feature)

(b) u - It is rounded. The others are unrounded

(c) e - It is a mid vowel (close-mid). The rest are high (or near high). This last answer contains a disjunction, so it has not strictly speaking an adequate answer.

Chapter 4

Question 1:

1. [-son, -cont]
2. [+lab]
3. [-cont, +ant]
4. [+voi, +cont]
5. [-voi, +cor, -son]
6. [+dor]
7. [-son, +voi]
8. [-back]
9. [+hi]
10. [+tns]
11. [+rnd]

Question 2:

1. [+cor]
2. [+son]
3. [-son, +cont]
4. [+hi]
5. [+back, -rnd]
6. [-hi]
Question 3:

<table>
<thead>
<tr>
<th></th>
<th>pen</th>
<th>wink</th>
<th>thimble</th>
<th>powder</th>
<th>pouting</th>
</tr>
</thead>
<tbody>
<tr>
<td>UR</td>
<td>/prɛŋ/</td>
<td>/wɪŋk/</td>
<td>/θɪmbl/</td>
<td>/pəʊdɪɾ/</td>
<td>/pəʊtŋ/</td>
</tr>
<tr>
<td>aspiration</td>
<td>pʰɛŋ</td>
<td>pʰaʊdɪɾ</td>
<td>pʰaʊtŋ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nasalization</td>
<td>pʰɛŋ</td>
<td>wɪŋk</td>
<td>θɪmbl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>velarization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canadian Raising (in some dialects)</td>
<td></td>
<td></td>
<td>pʰΛʊtŋ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flapping</td>
<td>[pʰɛŋ]</td>
<td>[wɪŋk]</td>
<td>[θɪmbl]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>[pʰɛŋ]</td>
<td>[wɪŋk]</td>
<td>[θɪmbl]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Question 4: Assume the UR for 닭 is /tak/ rather than the prescribed but seldomly spoken /talk/.

<table>
<thead>
<tr>
<th></th>
<th>/mɑk-ta/</th>
<th>/mɑk-num-ta/</th>
<th>/tak-i/</th>
</tr>
</thead>
<tbody>
<tr>
<td>nasalization</td>
<td>mɑŋ-num-ta</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fortition</td>
<td>mɑk-tʰa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>intersonorant voicing</td>
<td>mɑŋ-num-da</td>
<td>tag-i</td>
<td></td>
</tr>
<tr>
<td>coda C stopping</td>
<td>mɑkʰ-tʰa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>[mɑkʰ-tʰa]</td>
<td>[mɑŋ-num-da]</td>
<td>[tag-i]</td>
</tr>
</tbody>
</table>

Question 5:

```
  ω
 ο   ο
  σ  σ  σ  σ  σ  σ  σ
 / O R O R R O R O R R |
| t N C tʃ N C N C pʰ N s N b N |
| r k s j d i m a æ ı |

  ω
 ο   ο
  σ  σ  σ  σ  σ  σ  σ
 / O R O R R O R O R |
| t ɪ N v N C N C pʰ ɔ ɪ t O C |
| æ ı ı d i m æ n t s |
```

Question 6:
Question 7:

[s] and [ʃ] are separate phonemes in Ajagbe. There is a minimal pair to show that they are distinct phonemes: [asi] and [aʃi]

Question 8:

The minimal pair [kapi] and [kaʃi] shows that /p/ and /ʃ/ are separate phonemes. Further, the minimal [pal] and [pʰal] shows that /p/ and /pʰ/ are separate phonemes. However, [pʰ] and [ʃ] are allophones of the same phoneme. Their distribution can be stated as follows.

\[ \text{[ʃ]} / V \_\_ V \]
\[ /pʰ/ \]
\[ [pʰ] / elsewhere \]

Question 9:

\[ \sigma \]
\[ /i/ \]
\[ [-stress] \]
\[ [i] / _-\_ \]
\[ [i] / elsewhere \]
Question 10:
\[
C \rightarrow [+\text{voi}] / \left[ +\text{voi} \quad -\text{son} \right]
\]

Question 11:

The two allomorphs of the 3rd person singular masculine prefix are [ha-] and [la-]. Their distribution is as follows.

\[
[\text{la-}] / \_
\]

\[
[\text{ha-}] / \_
\]

[ha-] / elsewhere

The two allomorphs of the factual morpheme are [wa-] and [wa?-]. Their distribution is as follows.

\[
[\text{sa-}] / \_ [-\text{son},+\text{cont}]
\]

[wa?-] / elsewhere

Chapter 5

Question 1:

The following trees do not take the advanced discussion in section 5.3.6 into account.

\[
S' \quad \text{COMP} \quad S \quad \text{NP} \quad \text{VP} \\
\quad \text{Det} \quad \text{AdjP} \quad \text{N} \quad \text{V} \quad \text{Det} \quad \text{AdjP} \quad \text{N} \\
\quad \text{the} \quad \text{Deg} \quad \text{Adj} \quad \text{student} \quad \text{ate} \quad \text{the} \quad \text{Deg} \quad \text{Adj} \quad \text{apple} \\
\quad \text{very} \quad \text{hungry} \quad \text{Det} \quad \text{AdjP} \quad \text{N} \quad \text{really} \quad \text{tasty}
\]
paraphrase: Is it with a fork that the children will eat the cake?

paraphrase: Is it the cake with a fork that the children will eat?

Question 2:
Question 3:

Polarity questions are formed by doubling the verb and placing the negative morpheme \( m6 \) between the two instances of the verb.

Chapter 6

Question 1:

A1 entails B1.
A2 does not entail B2.
A3 entails B3.
A4 does not entail B4.

Question 2:

A2 does not presuppose B2.
A3 presupposes B3.
A4 does not presuppose B4.

**Question 4:**

1 $\rightarrow$ 2, 3
2 does not entail any of the sentences here.
3 does not entail any of the sentences here.
4 $\rightarrow$ 2
5 $\rightarrow$ 1, 2, 3, 4
6 $\rightarrow$ 1, 2, 3, 4, 5
7 $\rightarrow$ 4, 5

**Question 5:**

A1 implicates B1.
A2 implicates B2.
A3 does not implicate B3.
A4 does not implicate B4.

**Question 4:**

entailment, implicature, implicature, presupposition, implicature, implicature, entailment, presupposition, entailment, entailment, presupposition, no relation, presupposition, presupposition, no relation, implicature
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>accusative</td>
</tr>
<tr>
<td>ADN</td>
<td>adnominal</td>
</tr>
<tr>
<td>AG</td>
<td>agent/agentive</td>
</tr>
<tr>
<td>AGR</td>
<td>agreement</td>
</tr>
<tr>
<td>ANIM</td>
<td>animate</td>
</tr>
<tr>
<td>ASP</td>
<td>aspect</td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
</tr>
<tr>
<td>CL</td>
<td>classifier</td>
</tr>
<tr>
<td>COMP</td>
<td>complementizer</td>
</tr>
<tr>
<td>COMPL</td>
<td>completive</td>
</tr>
<tr>
<td>COP</td>
<td>copula</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>DECL</td>
<td>declarative</td>
</tr>
<tr>
<td>DEM</td>
<td>demonstrative</td>
</tr>
<tr>
<td>DET</td>
<td>determiner</td>
</tr>
<tr>
<td>DU</td>
<td>dual</td>
</tr>
<tr>
<td>DUC</td>
<td>dualic - refers to the involvement of two objects</td>
</tr>
<tr>
<td>EXCL</td>
<td>exclusive</td>
</tr>
<tr>
<td>FACT</td>
<td>factual</td>
</tr>
<tr>
<td>FEM</td>
<td>feminine</td>
</tr>
<tr>
<td>FORMAL</td>
<td>formal</td>
</tr>
<tr>
<td>FUT</td>
<td>future</td>
</tr>
<tr>
<td>HON</td>
<td>honorific</td>
</tr>
<tr>
<td>HORT</td>
<td>hortative</td>
</tr>
<tr>
<td>IMPER</td>
<td>imperative</td>
</tr>
<tr>
<td>IMPF</td>
<td>imperfective</td>
</tr>
<tr>
<td>INANIM</td>
<td>inanimate</td>
</tr>
<tr>
<td>INCL</td>
<td>inclusive</td>
</tr>
<tr>
<td>INDIC</td>
<td>indicative</td>
</tr>
<tr>
<td>INTER</td>
<td>interrogative</td>
</tr>
<tr>
<td>MASC</td>
<td>masculine</td>
</tr>
<tr>
<td>NEG</td>
<td>negative</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative</td>
</tr>
<tr>
<td>N.PREF</td>
<td>nominal prefix</td>
</tr>
<tr>
<td>N.SUF</td>
<td>nominal suffix</td>
</tr>
<tr>
<td>NZLR</td>
<td>nominalizer</td>
</tr>
<tr>
<td>PASS</td>
<td>passive</td>
</tr>
<tr>
<td>PERF</td>
<td>perfective</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>PRS</td>
<td>present</td>
</tr>
<tr>
<td>PSB</td>
<td>possibility</td>
</tr>
<tr>
<td>PST</td>
<td>past</td>
</tr>
<tr>
<td>PUNC</td>
<td>punctual</td>
</tr>
<tr>
<td>Q</td>
<td>question particle</td>
</tr>
<tr>
<td>RECIP</td>
<td>reciprocal</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SPEC</td>
<td>specific</td>
</tr>
<tr>
<td>THV</td>
<td>thematic vowel</td>
</tr>
<tr>
<td>VLZR</td>
<td>verbalizer</td>
</tr>
</tbody>
</table>
Alphabetical Index

AAE, 216
acquisition
   babbling, 256
   errors, 254, 257
agreement, 10, 28, 29, 51, 168, 288
Ainu, 90, 244, 246, 265
airstream mechanism, 68
Altaic Hypothesis, 245
American Sign Language, see ASL
animacy, 41
aphasia, 273, 282
Apuriña, 162
Arabic, 58
ASL, 56, 76, 88
Austroasiatic, 241, 248
babbling, 256
Balkan Sprachbund, 21
Bengali, 149
Berko, Jean, 27
Blackfoot, 5, 41, 42, 104
Bloomfield, Leonard, 18, 88, 189
bound root, 31
Breton, 264
Cantonese, 34, 52, 99, 103, 133, 168, 175, 212, 222
Chinese, 21, see also Mandarin, Cantonese, etc.
Chomsky, Noam, 3–5, 14, 18, 128
click, 68
compositionality, 32
constituent, 129
coordination, 138
Cornish, 265, 267
Cree, 21, 63, 117
creole, 222
Czech, 30
Dakota, 161
debuccalization, 239, 242
deep structure, 167
degree word, 149
dialect, 12, 20, 21, 71, 72, 75, 81, 209, 210, 212, 213, 227, 239, 266, 270
dialect continuum, 20
dialectology, 210
dvandva, 52, 54
ellipsis, 214
emergentism, 254
empiricism, 254
English, see also AAE
   American, 82
   Australian, 82
   British, 82, 215
   Canadian, 82, 101, 110, 215
   Middle English, 227–229, 241
   North American, 215
   Old English, 227, 238, 241
   US, southern, 215
entailment, 188
etymology, 226

Finnish, 239
Fortis, 70
FOXP2, 288
Frege, Gottlieb, 189
French, 41, 42, 72, 81, 131, 137
Canadian, 97
garden path sentence, 272
Genie, 5, 19
German, 10, 17, 20, 112, 170, 257, 264
Gopnik, Myrna, 288
Greek, 228
Ancient, 237
Grice, Paul, 202
Gricean Maxims, 202
Grimm’s Law, 240
Guugu Yimithirr, 13
Haas, Mary, 18
Hawaiian, 265
Hindi, 17, 228, 240, 263
Hixkaryana, 161, 162
Hokkien, 238
Hopi, 14
hortative, 155
Hungarian, 25

Ibibio, 78
idiom, 132
illocutionary force, 155, 156, 159, 201
implicature, 197, 203
conventional, 200
scalar, 200, 201, 278
Indo-European, 228, 246
Indonesian, 56
Inuktitut, 14, 50
isolate, 244
Italian, 6, 242, 243
Japanese, 78, 92, 149, 233, 245
Jeju Language, 245
Jespersen, Otto, 220
Jones, Sir William, 228
Korean, 6, 7, 13, 24, 30, 42, 54, 69, 76, 79, 133, 147, 156, 168, 196, 219, 222, 245
acquisition of, 258, 259
bound roots, 31
Busan dialect, 214
clippings, 232
fortis consonants, 70
Jeolla dialect, 75, 212, 234
Kyengsang dialect, 212
Middle Korean, 230
Old Korean, 229
onomatopoeia, 4
palatalization in, 91, 238
passive voice, 170
Romanization of, 2
suppletion in, 56
transcription of, 79
vowel length, 75
Kubeo, 25

Lakhota, 68
Lakoff, Robin, 220
language nest, 267
late closure, 275
Latin, 31, 47, 52, 228, 235, 236, 238, 240, 241, 268
lexical decision task, 276
Malagasy, 162
Manchu, 245, 265
Mandarin, 25, 30, 50, 131, 233
Marshallese, 57
merger, 212, see also vowel, merger
minimal attachment, 275
Mohawk, 50, 133, 268
Mongolian, 176
Montague, Richard, 190
Māori, 267
N400, 277
Navajo, 264
Ngiti, 59
Northern Sotho, 133
Nyaheun, 241
Ojibwe, 21, 212
Oneida, 25, 30, 124
Onondaga, 31, 33, 124, 137, 156

passivization, 167
Persian, 133, 228, 240, 251
pidgin, 222
PIE, 235, 240, 246
Pig Latin, 103
Pirahã, 16, 264
pitch accent, 212
polarity question, 163, 166
Portuguese, 2, 32, 88, 113, 133, 135, 222, 242, 243, 264
proform, 130, 134, 135, 153
proposition, 155
Proto-Germanic, 235, 237, 240, 242
Proto-Romance, 242
Pulmonic, 68
Pāṇini, 12, 18, 127

Received Pronunciation, 82
reduplication, 56
register, 210
Romance, 243
Romanian, 21, 242, 243
Russian, 241
Sami, 265
Sanskrit, 228

Sapir-Whorf Hypothesis, 12, 14
Sardianian, 243
Sardinian, 242
scalar implicature, 280
signed languages, 16, 76
Sino-Tibetan, 17, 21
Spanish, 28, 51, 239, 242, 243
    Caribbean, 237
speech act, 201
spoonerism, 103, 274
sprachbund, 21, 222
style, 210
surface structure, 167
Swahili, 25, 56, 162, 234
Swedish, 78, 240
    Old Swedish, 237
Thai, 30
Tocharian, 246, 247
Tok Pisin, 222
tone, 34, 65, 76
transitivity, 147
Turkish, 51, 78, 99, 245
Tuscarora, 265
two-word stage, 19
variable, 218, 220
Verner’s Law, 240
vowel, 79
    merger, 80, 210, 212
word order, 161, 165
Wug Test, 27, 231
Xhosa, 69
Yéli Dnye, 13
Zapotec, 162
Bibliography


Murray, L. (1795). *An English Grammar*. J.B. Baldwin, Bridgeport, CT.


Wilson, T. (1553). *The arte of rhetorique for the vse of all suche as are studious of eloquence, sette forth in English*. London.


