The Linguistic Analysis of Word and Sentence Structures

JULIANNE DONER

The Linguistic Analysis of Word and Sentence Structures

THE LINGUISTIC ANALYSIS OF WORD AND SENTENCE STRUCTURES

JULIANNE DONER

University of Manitoba
Winnipeg

The Linguistic Analysis of Word and Sentence Structures Copyright © by Julianne Doner is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, except where otherwise noted.

CONTENTS

About the cover xiii
Accessibility statement xvi
Becoming a linguist xviii
Contact the author xxi
Revisions history xxii

Part I. Foundations

1. What are morphology and syntax?

1.1. What is morphology? 5
1.2. What is syntax? 11
1.3. Morphosyntax as a subfield of linguistics 17
1.4. Becoming a linguist: Study tips for linguistics classes 23
1.5. Becoming a linguist: Myths and misconceptions about academic integrity 31
Connor Mark
Chapter 1 practice exercises 38

2. The scientific method in morphosyntax

2.1. Linguistics as science 45
2.2. Evidence and hypotheses 50
2.3. Prescriptivism and descriptivism 56
2.4. The lexicon and dictionaries 67

2.5. Becoming a linguist: Identifying academic sources 72
2.6. Becoming a linguist: Scientific writing in morphosyntax 79
2.7. Becoming a linguist: Citation in morphology and syntax 88
Chapter 2 practice exercises 99

3. Theories of grammar and language acquisition

3.1. What is a grammar? 107
3.2. Generative Grammar 113
3.3. The acquisition of morphosyntax 118
3.4. Poverty of the Stimulus and Universal Grammar 125
3.5. Debating Universal Grammar 130
3.6. Computer use of language 135
3.7. Becoming a linguist: Empirical and theoretical arguments 143
Chapter 3 practice exercises 147

4. Linguistic data in morphosyntax

4.1. Language modality 151
4.2. Methods of data collection 157
4.3. Language, dialect, and grammar 162
4.4. What is grammaticality? 166
4.5. The validity of speaker judgments 171
4.6. Becoming a linguist: Discussing data 175
4.7. Becoming a linguist: Glossing spoken language data 180
Chapter 4 practice exercises 185

5. Decolonizing linguistics

5.1. Language and community 191
5.2. Colonialism and decolonization 196
Chapter 5 practice exercises 202

6. Kinds of morphemes and morphological processes

6.1. The building blocks of morphology 207
6.2. Affixes 210
6.3. Inflection and derivation 219
6.4. Reduplication and non-concatenative morphology 222
6.6. Allomorphy 227
6.7. Clitics 233
6.8. Becoming a linguist: Anatomy of an academic article 239
6.9. Becoming a linguist: How to read an academic article 247
6.10. Guided reading: Is *n't* a clitic or an affix? 253

7. Predicates, arguments, and clauses

7.1. Predicates 267
7.2. Arguments 271
7.3. Clauses 276
7.4. Statements, questions, and commands 282
7.5. Relative clauses 287
Chapter 7 practice exercises 290

8. Parts of speech

8.1. Determining part of speech 299
8.2. Nouns 305
8.4. Adjectives and adverbs 327
8.5. Functional parts of speech 333
8.6. Common pitfalls 341
8.7. Case study: French *possessive adjectives* are not adjectives 345
Connor Mark
Chapter 8 practice exercises 353

Part II. Morphology

9. Words

9.1. Word formation processes 363
9.2. Different meanings of *word* 369
9.3. Packaging words and morphemes 373
9.4. Wordhood: An unsolved problem 382
Chapter 9 practice exercises 391

8. Segmenting words into morphemes

10. Compounding

9. Inflection vs. derivation

11. Hierarchical structure in words

11.1 Hierarchical structure in compound words 407

12. Morphophonology

13. Productivity and word formation

14. The lexicon

15. Clitics

Part III. Syntax

16. Constituency and hierarchy

17. Phrase structure rule grammars

18. Syntax trees

19. Binding theory

20. Predicates and arguments

21. Types of clauses

22. Intro to X-bar

23. Intro to movement

24. Different ways of doing syntax

Part IV. Morphology-syntax interface

25. Y model of grammar

26. Morphosyntactic features, selection, and agreement

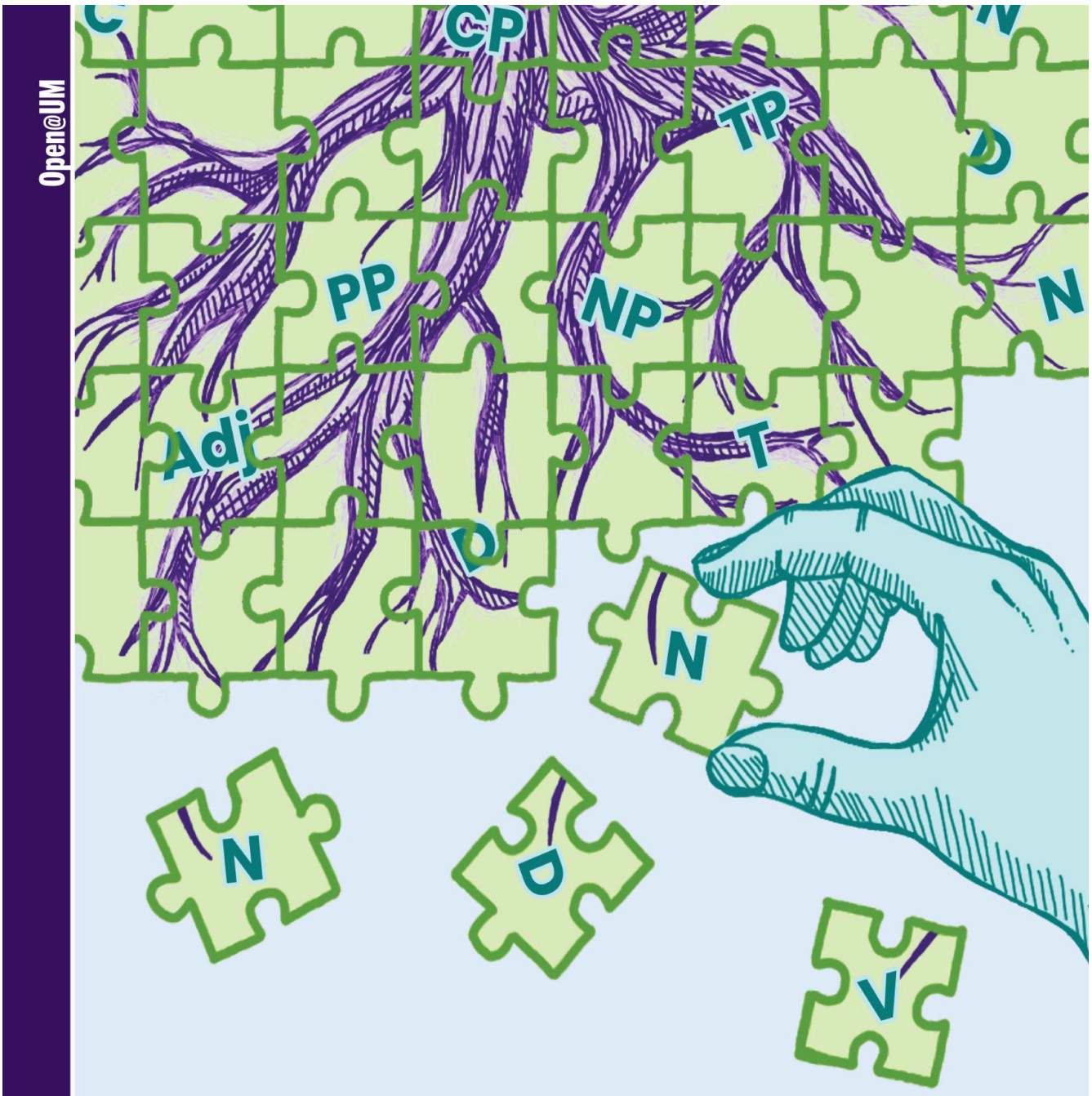
27. Language universals and diversity

28. Distributed morphology

Glossary 451
List of abbreviations 475
Lists of tables and figures 478
Gloss Template HTML and CSS 484

ABOUT THE COVER

In March 2024, we ran a design contest for the cover of *The Linguistics Analysis of Word and Sentence Structures*. After receiving many excellent contributions, we selected the winning design by Ysabelle Tankeko because of its aesthetic appeal and her incorporation of linguistic elements into the design.



The Linguistic Analysis of Word and Sentence Structures

Julianne Doner

University of Manitoba

Figure 1. The cover for this textbook, designed by Ysabelle Tankeko.

About the artist

Ysabelle Tankeko is a second-year University of Manitoba student, majoring in psychology, with a minor in linguistics and biological science. In the future, Ysabelle hopes to become a speech-language pathologist or an occupational therapist. Though she is still unsure of which career path she wants to pursue, Ysabelle has a passion for helping children. During her free time, Ysabelle enjoys crafting, playing badminton, and hanging out with her dog, Akina. Ysabelle also loves picking up new hobbies and trying new things.

Ysabelle has always had a passion for art and design, and with her newfound interest and appreciation for linguistics, she decided to join the textbook design contest.

Artist's note on the cover design

Regarding the textbook design, I tried to illustrate things I remember learning from LING 1000 with Dr. Will Oxford. Both morphology and syntax use tree diagrams to show the structure of words and sentences, leading to the creation of the tree. The tree is displayed upside down to reflect the structure of the tree diagram and its branches. The puzzle represents the building of words or sentences. I always thought that figuring out the structure of a tree diagram was similar to completing a puzzle. I've also incorporated the abbreviations of phrases found in syntax structure. Following the X-bar template, I tried to position the letters similarly to how a complete diagram would look like.

ACCESSIBILITY STATEMENT

We are committed to meeting and exceeding accessibility standards in our design and delivery of this textbook.

This textbook follows the accessibility guidelines found in the *Accessibility Checklist for OER Development* from eCampus Ontario, which was based on Conestoga Library Services' *Adapting or Creating OER* accessibility checklist.

One major development is that interlinear glosses were built using novel code, which is available in the backmatter of this text. Ideally, the use of such code will raise the accessibility standards for interlinear glosses in other OERs and beyond. The new code gives the following advantages.

- Screen readers can read glosses as tables and thus allow more flexible navigation of them (i.e., users can view them in a grid format). It also means that users can quickly hop from gloss to gloss via navigating to the next or previous table more quickly.
- The glosses are formatted consistently in terms of visuals, mostly conforming to traditional glosses, allowing sighted users to read them more easily than if they were formatted using regular tables.
- Numbers are rewritten in transparent text for glosses with more than one sub-example. This means that screen readers will read the number again while sighted users will not be distracted by seeing the number repeated, even if using an external application that changes the background colour (i.e. for a dark mode interface).

Other standard accessibility practices undertaken include but are not limited to the following.

- High colour contrast.
- Screen reader navigability, as tested using NonVisual Desktop Access (NVDA).
- Alt-text to describe images.
- The ability to use the OER in multiple forms, including PDF download as well as use on both computers and mobile devices.

In addition to formatting, we also used simple, approachable language and broke up the text into smaller passages.

There are also future opportunities to improve accessibility, including but again not limited to the following.

- Improve screen readers' ability to accurately read non-English forms in interlinear glosses and within the

text.

- Anything that users requiring improved accessibility tell us. We are open to feedback. See Contact the author for information on how to provide feedback.

BECOMING A LINGUIST

Throughout this textbook, I have included a series of sections under the name *Becoming a linguist*. These sections have three main goals:

- To help students learn the "hidden curriculum" of the field of linguistics; that is, the knowledge and skills needed to succeed in linguistics, but aren't always explicitly taught.
- To incorporate writing throughout the curriculum, as with the "Writing across the curriculum" approach, to help students develop their general writing skills, learn genre-specific linguistics writing skills, and deepen their learning through writing tasks.
- To guide instructors, especially new instructors, in incorporating hidden curriculum and writing instruction and assessment into their courses.

End-of-chapter exercises also include activities that support the *Becoming a linguist* content, especially the *Communication and study skills* exercises.

Becoming a linguist content

Currently, the following *Becoming a linguist* sections are already included in the text, with more on the way!

- Section 1.4: Study tips for linguistics classes
- Section 1.5: Myths and misconceptions about academic integrity (by Connor Mark)
- Section 2.5: Identifying academic sources
- Section 2.6: Scientific writing in morphosyntax
- Section 1.7: Citation in morphology and syntax
- Section 3.7: Empirical and theoretical arguments
- Section 4.6: Discussing data
- Section 4.7: Glossing spoken language data
- Section 6.8: Anatomy of an academic article
- Section 6.9: How to read an academic article

Planned sections:

- Section 4.8: Glossing signed language data

- Finding academic sources
- Careers in linguistics
- The research process
- How to take notes
- Applying to grad school
- Navigating uncertainty and competing theories
- The names of languages
- Getting the most out of your course syllabus

If you have any suggestions, please feel free to contact the author.

References and further resources

More resources for students

Charity Hudley, Anne, Cheryl Dickter, and Hannah Franz. 2017. *The Indispensable guide to undergraduate research: Success in and beyond college*. New York: Teachers College Press.

Macaulay, Monica. 2011. *Surviving linguistics: A guide for graduate students*, 2nd edition. Somerville: Cascadilla Press.

Research on the hidden curriculum in linguistics and beyond

Bergenhenegouwen, George. 1987. Hidden curriculum in the university. *Higher Education* 16: 535-543.

Doner, Julianne, and Connor Mark. 2024. Explicit writing instruction in an OER textbook. Oral presentation at the annual meeting of the Canadian Linguistics Association. Slides.

Doner, Julianne, Lisa Sullivan, Emilia Melara & Heather Yawney. 2024. Why aren't we teaching writing?: The advantages of early explicit writing instruction in linguistics. *Canadian Journal of Linguistics* special issue on Teaching and Learning in Linguistics.

Pappas, Panayiotis A., Maite Taboada & Kathryn Alexander. 2019. Teaching linguistic argumentation through a writing-intensive approach. *Language* 95(3): e339–e363.

Petrucci, Peter R. 2002. A writing-to-learn approach to writing in the discipline in the introductory linguistics classroom. *The WAC Journal* 13: 133–143.

Portelli, John, 1993. Exposing the hidden curriculum. *Journal of Curriculum Studies* 25(4): 343-358.

Smith, Buffy. 2013. *Mentoring at-risk students through the hidden curriculum of higher education.* Lexington Books.

CONTACT THE AUTHOR

Report an issue

If you would like to report an issue with this textbook, please do so using this form.

Report an adoption

If you have adopted this text for use in a course, please let me know by filling out this form.

Collaboration and suggestions

If you would like to give suggestions or contribute to the textbook, please e-mail me at julie dot doner at umanitoba dot ca.

I am especially open to:

- Language profiles, especially from community members
- Problem sets and practice exercises
- Suggestions of hidden curriculum topics to include
- Non-English examples, especially of understudied or minoritized languages
- Applications to current events
- Interesting resources to add to the *References and further resources* section
- Suggestions about how to further promote Equity, Diversity, Inclusion, and Decolonization
- Multimedia resources

I am open to student-written contributions.

REVISIONS HISTORY

As *The Linguistics Analysis of Word and Sentence Structures* is an online textbook, it will be continuously updated. Major revisions, such as the addition of new content, will normally be listed here. Minor revisions, such as the correction of typos and clarification of wording, will not normally be listed.

Version #	Date	Description of changes
1.0	January 5, 2024	Initial release of Chapters 1 and 2
1.1	January 7, 2024	Initial release of Chapter 3
1.2	January 17, 2024	Initial release of Chapter 4 (less 4.8)
1.3	January 24, 2024	Initial release of Chapter 8 (previously chapter 6)

PART I. FOUNDATIONS

In Chapters 1-6, we will learn how to:

- Define the fields of morphology and syntax.
- Apply the scientific method to the study of morphology and syntax.
- Describe the main theoretical assumptions of generative grammar.
- Describe and identify various kinds of morphemes and morphological processes..
- Collect, share, present, and format morphosyntactic data.
- Classify words according to part of speech.
- Develop writing, argumentation, and study skills, especially as they pertain to linguistics.

1. WHAT ARE MORPHOLOGY AND SYNTAX?

This textbook is an introduction to the study of the sub-fields of **morphology** and **syntax**. In this chapter, we begin by defining both these words and some basic concepts in both subfields. We will discuss some of the applications of morphology and syntax, and show how they interact with the larger field of **linguistics**. We will also introduce some study habits that will help you succeed in the study of linguistics, as well as give an overview of some common misconceptions about academic integrity.

Learning objectives

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

- Define the fields of **morphology**, **syntax**, and **linguistics**.
- Describe some applications of the study of morphology and syntax and why they are relevant.
- Define **morphemes** and distinguish between simplex and complex words.
- Apply some study strategies to common tasks in linguistics courses to improve your learning, problem solving, and writing.
- Understand what counts as an academic integrity violation.

1.1. WHAT IS MORPHOLOGY?

This textbook is about the structures of words and sentences. In the field of linguistics, the study of the structure of words is called **morphology**.

> Technically, the word *morphology* means the study of form or shape. In the context of linguistics, this means the form and shape of words, but other fields of study use the word *morphology* as well. For example, in biology, *morphology* means the study of the form or shape of living things. For this reason, when you do Internet searches for sources in morphology, you will likely get better results if you include a word like "linguistics" or "language" in your search terms alongside *morphology*.

To study the structure of words, we look at their components. Let's take the word *untidy*, meaning 'not organized or clean,' as an example. Even though *untidy* is a single word, we can break it up into two pieces, *un-* meaning 'not,' and *tidy* meaning 'organized or clean.' These two components are called **morphemes**. We define morphemes as the smallest meaningful units in language. When we break up a word into its morphemes, we say it has been **segmented**.



Figure 1: An untidy workspace. Photo by SN.CHE. Used under Pexels license.

Some words are **simplex**, meaning they only contain a single morpheme. The word *pumpkin,* a kind of squash, is a simplex word. Other words are **complex**, meaning they contain more than one morpheme. The word *untidy*, as we have already seen, is complex. Another word for *complex* is **polymorphemic**.



Figure 2: A selection of pumpkins. Photo by Allie Reefer. Used under Pexels license.

Note that we can break up the word *untidy* into even smaller pieces. For example, we can break it up into consonants and vowels, which we could write in the **International Phonetic Alphabet (IPA)** as /ʌ n t ɑi ɾ i/. Likewise, we can break up the word *pumpkin* to get /p ʌ m p k ɪ n/. Although individual consonants and vowels do function as units in **phonology**, they are not *meaningful* units. Each of these individual consonants and vowels do not mean anything on their own, which is why they are not morphemes.

We can also break words up into syllables. The word *untidy* has three syllables, /ʌn/, /tɑi/, and /ɾi/. The sequence /ʌn/ happens to be a morpheme in this word, but the other two syllables are not. A morpheme may be one syllable long, like *un-*; a morpheme may be longer than a syllable, like *tidy*; or a morpheme may even be shorter than a syllable, like the plural marker *-s* in the word *cats*! Just like consonants and vowels, syllables are structural units of phonology that do not encode meaning.

Another possible definition of *morphology* is the systematic co-variation of form and meaning. In this definition, *systematic* means a pattern or system that is predictable or rule-governed. *Co-variation* means that when some element varies (in this case form), another element (meaning) varies in a predictable way at the same time. *Form* means the sounds or signs used to produce a word. In other words, morphology is the study of how, if you change the shape of a word, its meaning changes, too; and if you change the meaning of a word, its shape changes, as well.

Let's consider again the morpheme *un-* in the word *untidy*. This morpheme shows up in lots of other words, as shown in Table 1. This is an example of systematic co-variation of form and meaning. In the first column, we have a set of words and their definitions. When we add *un-* to the beginning of each of these words, the meaning changes to mean the opposite. Adding *un-* is a variation in form, while adding 'not' to the definition is a variation in meaning. These two variations are related since the same change in form always leads to the same change in meaning, and vice versa.

Table 1. Words with *un-*

tidy	'clean, organized'	*untidy*	'not clean, not organized'
happy	'in a pleasurable state of mind'	*unhappy*	'not in a pleasurable state of mind'
like	'similar to'	*unlike*	'not similar to'
fair	'equal, just'	*unfair*	'not equal, not just'

On the other hand, the word *pumpkin* is simplex, meaning it does not have any components that exhibit a systematic co-variation in form and meaning. By coincidence, the word pumpkin is made up of two components that do show up in other words, *pump* and *kin*. On its own, *pump* means 'use pressure or suction to force air or liquid to move from one contained area into another', which has nothing to do with squash vegetables like pumpkins. Although the words *pumpkin* and *pump* share form, they do not share meaning. Their similarity is a coincidence. Likewise, the word *kin*, on its own, means 'family'. Again, this has nothing to do with squash vegetables and its similarity in form to *pumpkin* is a coincidence. Neither *pump* nor *kin* are morphemes in the word *pumpkin*.

Some of the things we study in morphology include the following:

- We look at related words to identify the co-variations in meaning and form and identify morphemes.
- We classify morphemes based on their structural, functional, and meaning properties.
- We determine whether a given string of morphemes is a word or a phrase.
- We look at ways words are built or created.
- We classify words according to the words they can appear alongside, their meaning, and their structure.
- We look for common patterns in word structure and word formation across the languages of the world.
- We study how the structure of words interacts with other components of language, such as **phonology**.

Morphology matters because a lot of language play and creativity happens with morphology, when we build new words and modify old ones. (For a fun example, check out this article explaining the grammar of doge!) We have intuitions about the words in our native languages and other languages we speak well. We have opinions about which words are good or come out sounding more awkward, which are funny and creative, which are confusing, and which are plain and boring. But how do we know these things? And which of these patterns are particular to specific languages, and which seem to show up in all languages?

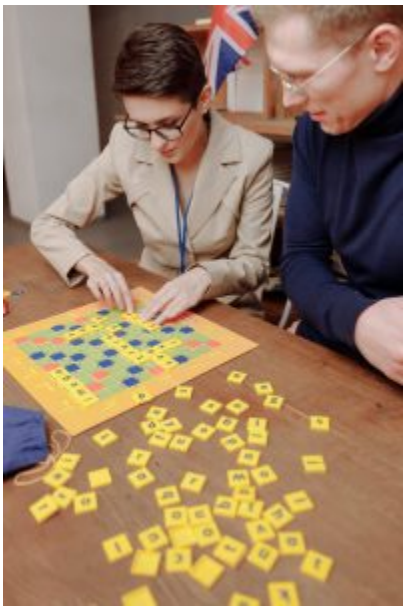


Figure 3: Language play. Photo by Thirdman. Used under Pexels license.

Key takeaways

- Morphology is the study of the structure of words or the systematic co-variation of form and meaning.
- Words consist of morphemes. Morphemes are the smallest unit of meaning.
- A simplex word only contains one morpheme, but a complex word contains multiple morphemes.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=71#h5p-8

References and further resources

Comedy and satirical linguistics

🎉 Wishnetsky, Susan. 2006. Morphemes: A new threat to society. *Lingua Pranca*. https://specgram.com/LP/26.coma.morpheme.html

For a general audience

🧠 CrashCourse. 2020. Morphology: Crash Course Linguistics #2. https://www.youtube.com/watch?v=93sK4jTGrss

🎉 McCulloch, Gretchen. 2014. A linguist explains the grammar of doge. Wow. *The Toast*. https://the-toast.net/2014/02/06/linguist-explains-grammar-doge-wow

For linguistics students

🔍 Anderson, Catherine and Bronwyn Bjorkman. 2020. Why we do morphology. *Word to the Whys*. Teaching in Linguistics Community of Practice. Podcast. https://sites.google.com/view/word-to-the-whys-podcast/home/morphology

1.2. WHAT IS SYNTAX?

Syntax is the study of sentence and phrase structure. In other words, syntax is the study of how we combine words and encode the relationships between them.

A **phrase** is any grammatically well-formed string of words. Some examples of English phrases are shown in (1).

(1) a. a pumpkin pie
b. down the stairs
c. very hungry
d. after the truly serious and unfortunate events of last night
e. rode their bikes to school every day

A **sentence**, on the other hand, is a phrase that forms a complete thought. Some examples of English sentences are shown in (2).

(2) a. A pumpkin pie rested on the windowsill.
b. I rolled a ball down the stairs.
c. Eliud Kipchoge was very hungry after winning the marathon.
d. I decided to talk to my therapist after the truly serious and unfortunate events of last night.
e. My children rode their bikes to school every day until it got too cold.

Both phrases and sentences consist of strings of words, but is any string of words allowed? Consider the strings of words in (3). Compared to (1), they are scrambled and do not make much sense. In linguistics, we put an asterisk (*) at the beginning of an example to indicate that it is not well-formed. From the examples in (3), we can see that the *order* of the words matter.

(3) a. *pumpkin a pie
b. *the stairs down
c. *hungry very
d. *after truly the serious unfortunate and night events last of
e. *their rode every day to bikes school

Interestingly, we are able to identify if a sentence or phrase is well-formed even if we don't understand all of the words used in the sentence. In (4), I give an example with **nonce** (made-up) words. Even though you can't understand the sentence (because I just made up most of the words in it!), you can still tell that the words are in the right order and the sentence is well-formed.

(4) The osathies pishdupped the sudfy daigus.

Language profile: English

English is one of the most spoken languages of the world, with nearly 380 million first-language speakers and over 1 billion second-language speakers (Eberhard et al. 2023). It originated in England when people from northwestern Germany (the Angles, Saxons, and Jutes) invaded England in the fifth century, which had previously been inhabited by Celtic peoples.

English belongs to the Germanic branch of the Indo-European language family. But it is also heavily influenced by some other languages. In 1066, the Norman French invaded England. They spoke a dialect of French and brought many French words into English, including *court*, *crown*, *village*, *diamond*, *feast*, and *cream*. Their influence also caused major changes in the grammatical structure of English. The variety of English spoken before the Norman invasion is known as **Old English**, while the variety of English spoken afterwards is known as **Middle English**. We now speak a variety known as **Modern English**, which is normally considered to have been spoken since around the year 1500 (Denning et al. 2007).

Because the British Empire took control of so many countries worldwide, English is spoken in dozens of countries, each of which has its own dialects and varieties of English—often more than one! All of these dialects have unique linguistic properties and a rich cultural heritage. You can listen

to audio recordings of many different dialects and accents at the International Dialects of English Archive.

Word order patterns vary from language to language. In English, we put **prepositions** (words like *in, at, on,* and *to*) before the noun they go with. But in many of the world's languages, the same kind of word goes after the noun. In those languages, they are called **postpositions** instead. Japanese and Turkish are both examples of languages with postpositions, as shown in example (5).

(5) a. *Japanese*

nihon	ni	
Japan	in	
'in Japan'		

b. *Turkish*

vapur	ile	
boat	with	
'with a boat'		

(Pereltsvaig 2021: 98, 100)

In the Japanese example in (5a), the postposition *ni* meaning 'in' comes after the word *nihon* meaning 'Japan' to form the phrase 'in Japan.' In the Turkish example in (5b), the postposition *ile* meaning 'with' comes after the word *vapur* meaning 'boat' to form the phrase 'with a boat.'

Formatting linguistic data

Linguists use very specific conventions for formatting and discussing linguistic data. We will learn together about how to format and discuss examples in Chapter 4. However, as you read this textbook and other linguistics resources, you should pay attention to how examples are formatted

and discussed, so that you can learn how to format and discuss data in your own work. Look for patterns and conventions, what works well, and what could be clearer.

Did you know? Adpositions

If you do not want to distinguish between postpositions and prepositions, you can use the word **adposition**.

According to Dryer (2013), about half of the world's languages use prepositions and about half of the world's languages use postpositions. A minority of languages do something else, such as not using adpositions at all, putting adpositions in the middle of words, or allowing adpositions to appear on either side of the noun.

Some of the things we study in syntax include the following:

- We look at how words are grouped into sentences and phrases.
- We classify phrases and sentences based on their structural, functional, and meaning properties.
- We determine the possible word orders of different kinds of phrases in particular languages.
- We determine whether words or phrases have relationships with other words or phrases in a sentence and how that affects their placement or their morphological form.
- We look for similarities and differences in phrase and sentence structure across languages.
- We create models to describe how each of the components of syntax interact to form phrases and sentences.
- We study how the structure of phrases and sentences interacts with other components of language, such as **semantics**.

One of the things that sets human language apart from other types of communication is how we are able to create new, specific meanings by combining words in a structured way. We do not merely string words together and guess at or infer how the different words are related. Their relationships are explicitly encoded through syntax—which is why we can say unlikely or even illogical things and have the meaning come through clearly.

This property is rather curious and unique. Emojis do not have it. When scientists have taught primates to sign, there is little evidence for syntactic structure in their communication. But without it, would we be able to have philosophical treatises or science fiction novels?

Key takeaways

- Syntax is the study of phrase and sentence structure.
- A phrase is any well-formed string of words, while a sentence forms a complete thought.
- In linguistics, we put an asterisk (*) in front of phrases and sentences that are not well-formed.
- Different languages have different word order patterns. For example, some languages have prepositions that go before the noun while other languages have postpositions that come after the noun.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=76#h5p-10

Consult the following data in answering the check yourself questions:

(6) *Boumaa Fijian*

au	na	talai	Elia	i	‘Orovou
I	FUTURE	send	Elia	to	‘Orovou

‘I’ll send Elia to ‘Orovou.’

(Dixon 1988: 216, as cited by Dryer: 2013)

(7) *Lezgian*

duxturrin	patariw	fena
doctors	to	went

'She went to doctors.'

(Haspelmath 1993: 218, as cited by Dryer: 2013)

References and further resources

For a general audience

🧠 CrashCourse. 2020. Syntax – Morphosyntax: Crash Course Linguistics #3. https://www.youtube.com/watch?v=B1r1grQiLdk

For linguistics students

Denning, Keith, Brett Kessler, and William R. Leben. 2007. *English Vocabulary Elements.* 2nd edition. Oxford: Oxford University Press.

🔍 Doner, Julie and Tim Mills. 2020. Why we do syntax. *Word to the Whys.* Teaching in Linguistics Community of Practice. Podcast. https://sites.google.com/view/word-to-the-whys-podcast/home/syntax

Pereltsvaig, Asya. 2021. *Languages of the World: An Introduction.* 3rd edition. Cambridge: Cambridge University Press.

Academic sources

Dryer, Matthew. 2013. Order of adposition and noun phrase. In *WALS Online* (v2020.3), ed. Matthew Dryer and Martin Haspelmath. Zenodo. http://wals.info/chapter/85

📑 Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2023. *Ethnologue: Languages of the World*. 26th edition. Dallas, TX: SIL International. http://www.ethnologue.com.uml.idm.oclc.org

🔍 Meier, Cameron (executive editor). 2023. *International Dialects of English Archive*. https://www.dialectsarchive.com

1.3. MORPHOSYNTAX AS A SUBFIELD OF LINGUISTICS

Both **morphology** and **syntax** are related subfields of **linguistics**. Linguistics can be defined as the scientific study of language. We will go deeper into the scientific aspects of **morphosyntax** in Chapter 2.

The study of natural human language

What do we mean when we say *language*? Linguists primarily study natural human language.

A **natural** language is a language that developed within a community without being planned or designed. Instead, a natural language evolves gradually over time based on how the community uses it. In contrast, constructed languages are designed intentionally. **Constructed languages** include languages designed for art, media, or fun, like Klingon for the Star Trek series, or for a communicative purpose, such as Esperanto, which was designed with the intention that it would be an international language that would promote peace.

Constructed languages and non-human communication systems are valid and interesting objects of study. However, depending on the kinds of questions you are asking, they may or may not be able to provide the evidence you need. For example, you may be studying the properties that languages around the world have in common. A constructed language, however, may have been constructed on purpose to sound alien or be unlike human language. Even if it wasn't designed to be different on purpose, we cannot know whether the conscious choices of a language designer will result in the same kinds of patterns as subconscious language acquisition unless we study them separately and then compare. Some linguists hypothesize a genetic or biological component that underlies the human capacity for language. If this is the case, then we would expect there to be differences between human communication and the communicative practices of other species.

The place of morphosyntax within linguistics

Within the field of linguistics as a whole, morphology and syntax are the study of the mediation of form and meaning. The form of language — the possible speech sounds and signs — is determined at the level of phonetics and phonology. The meaning of language is determined at the level of semantics and pragmatics. In morphology and syntax, we study how to encode particular meanings with particular forms.

Let's use the example in (1) to illustrate the relationship between form and meaning.

(1) The toddler kicked the balls across the field.

The form of the sentence in (1), if it is spoken, sounds similar to (2), when represented in an **IPA** transcription, and its meaning describes something like the image shown in Figure 1.

(2) /ðətɑɾlə˞kɪktðəbɑlsəkɹɑsðəfild/



Figure 1: A toddler kicking a ball in a field. Photo by Paul Ney. Used under CC BY-NC-SA license.

But how do we get from the form in (2) to the meaning in Figure 1? Well, we divide up the string of sounds in (2) into morphemes and words, each of which has its own meaning. But that, alone, is not enough. The order of the words and morphemes and how they are combined also matter. Look at the examples in (3). Notice how they all have the same words and morphemes as (1), but combined in different ways. The different combinations produce different meanings, some of which make no sense.

(3) a. The balls kicked the toddler across the field.
b. The field kicked the balls across the toddler.
c. The kicker toddled the balls across the field.
d. The toddlers kicked the ball across the field.

In morphology and syntax, we are assigning a structure to a string of sounds or signs and using that structure to determine its meaning. This structure indicates the relationships between the morphemes and words. For example, the *-s* plural marker on *balls* in (1) indicates that there is more than one ball. On its own, *-s* just means more than one something, but we don't know what that something is until it is part of a structure. Likewise, the morpheme *kick* indicates that an action of kicking is involved, but until *kick* is combined with the past

tense marker *-ed,* we don't know when it happened, and unless it is combined with a subject, object, and other modifiers, we don't know who did the kicking, what was kicked, or where the kicking happened.

The subfields of linguistics

Linguistics is very broad and covers every aspect of language structure and use. The lines are fuzzy, but we could roughly divide linguistics into subfields that focus on the structure of language and subfields that focus on how the structure of language intersects with other aspects of the human experience. All of these areas of research are important for advancing our understanding of how language works. There are also interactions between all of the different subfields. For example, you could study how to train a computer to produce artificial speech with the correct pronunciation (phonetics, phonology, and computational linguistics), how children acquire morphology (morphology and first language acquisition), or how a word gains a special meaning within a particular community and then spreads outside of the community (lexical semantics, historical linguistics, and sociolinguistics).

Morphology and syntax, as well, can be studied in combination with any other subfield. Here are some examples:

Semantics

Semantics is the study of meaning. One important idea in semantics is **compositionality**, which is the observation that the meaning of a sentence is derived by the meaning of its parts and the way that they are combined. This is very closely related to the study of morphosyntax since morphosyntax is what determines how they are combined.

Phonetics and phonology

The subfields of **phonetics** and **phonology** intersect with morphology through the phenomenon of **allomorphy**. A morpheme may be pronounced in different ways depending on its phonological context. We will learn more about this in Chapter 12.

The subfields of **phonetics** and **phonology** also intersect with syntax through **prosody**. Many syntactic structures in various languages affect the prosody of the sentence. For example, in English, questions have a special rising intonation.

Sociolinguistics

In **sociolinguistics**, different dialects and language varieties are studied. Any language variety is a valuable object of study within the subfields of morphology and syntax. Some critics of morphosyntax point out that we do not adequately account for variation in language.

Psycholinguistics

In the subfields of **psycholinguistics** and **acquisition**, linguists study the processing and acquisition of morphosyntactic patterns. Some critics of morphosyntactic theory argue that there is a mismatch between morphosyntactic models of grammar and the ways we seem to process language.

Anthropological linguistics

Anthropological linguistics is the study of the interaction between language and culture. Different languages encode different aspects of the world directly into the grammatical structure of language. For example, some languages have grammatical categories that can be used to indicate formality or respect, while others do not.

Discourse analysis

In **discourse analysis**, linguists study how different sentences or utterances are combined into wider texts or conversations. Sometimes this involves the use of morphological or syntactic **discourse markers**, like *oh*, *well*, *so*, *y'know*, etc... Some syntacticians work on integrating discourse markers into syntactic structure, while morphologists might analyze their morphological structure.

Computational linguistics

A **computational linguist** might study how to program a computer to produce and interpret language in a human-like way. Although the ability of computers to manipulate language has improved by leaps and bounds, there are still some areas of difference, if you look closely.

Crip linguistics

Crip linguistics is not so much a subfield, but more a critical reflection on linguistics as a whole from the

perspective of disability studies, challenging us to not consider any attempts at communicating to be deficient or defective. The term *crip linguistics* was coined by deaf linguists Octavian Robinson and Jon Henner (e.g., Henner and Robinson 2023) based on the reclaimed usage of *crip* in critical disability studies (e.g., McRuer 2006, Kafer 2013). In morphosyntax, we may investigate how various communicative behaviours exhibit morphological or syntactic patterning.

Which subfields of linguistics are missing here? How are they also connected to morphology and syntax?

Key takeaways

- Linguistics can be defined as the scientific study of language.
- The primary object of study in linguistics is natural human language.
- In both morphology and syntax, we study how to encode particular meanings with particular forms.
- Morphology and syntax interact with all of the other subfields of linguistics.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=78#h5p-11

References and further resources

For a general audience

🎉🔍 Okrent, Arika. 2010. *In the land of invented languages: Adventures in linguistic creativity, madness, and genius*. Toronto: Random House.

Academic sources

📚 Henner, Jon, and Octavian Robinson. 2023. Unsettling languages, unruly bodyminds: A Crip Linguistics manifesto. *Journal of Critical Study of Communication and Disability* 1 (1): 7–37. https://criticalstudycommunicationdisability.org/index.php/jcscd/article/view/4

Kafer, Alison. 2013. *Feminist, queer, crip.* Bloomington: Indiana University Press.

McRuer, Robert. 2006. *Crip Theory: Cultural Signs of Queerness and Disability*. New York: New York University Press.

1.4. BECOMING A LINGUIST: STUDY TIPS FOR LINGUISTICS CLASSES

Reading to learn

Reading is often a major part of the course material in university classes. I have two tips for how to best use your course readings to learn the material.

First, read before class. If you read before class, then when your instructor explains the material, you will be hearing it the second time, and it will make more sense. If it feels like class is moving too fast, spending more time with the reading before class might help you keep up. Doing the readings before class will also help you predict where the instructor is going, take better notes, make new connections, and ask better questions. You may find it helpful to *also* revisit the reading after class, but you can do so strategically, revisiting just the tricky bits or parts you may have misunderstood the first time through.

Second, read with questions in mind. You should be asking three kinds of questions:

- When you don't understand something or find something confusing, write out a question that puts your confusion into words. This will help you to better identify what you understand well or need to study more. It will also help you figure out how to fill in the gaps in your learning, whether by listening for it in lecture, by asking about in office hours, or by revisiting the reading.
- Ask questions about problems or inconsistencies you find. This will help you to become a more critical thinker.
- Ask questions about how the course material connects to other units in the same course, to other courses, and to your daily life. This will help you remember the material long-term and make it more meaningful to you.

Problem solving skills

Doing linguistics involves a lot of problem solving and data analysis. The goal we are working towards is that you will be able to apply your problem-solving skills to novel data. That is, that you will have the skills to identify patterns in data that you have never seen before.

When you watch your instructor demonstrate problem solving in class, it can be very easy to trick yourself into thinking that you know how to solve that kind of problem, as well. As you watch your instructor, you will

likely feel that every step makes sense. But that doesn't mean that you will be able to figure out what the next step should be when your instructor is not there.

The best way to improve your problem-solving and analysis skills is to get lots of practice! You do not want your first time attempting any kind of problem to be on a test or an assignment worth lots of points. You want to practice throughout the semester, so that when the test or assignment comes, you will be ready!

When practice exercises are assigned to you as homework, first attempt them on your own. If you get stuck, note where you get stuck and then get help. If your instructor is going to go through the answers to your homework in class, do the homework before that class and pay attention to what the instructor did differently than you. When I teach, I try to explain how to avoid as many false paths as I can think of, but my students are always more creative than me and find more!

Although I recommend first attempting the exercises on your own, it can be very helpful to compare your answers with your classmates. If you're both stuck, figuring out together how to get unstuck will be easier than doing so on your own, and you're more likely to remember it than if I tell you what to do.

Before you work with your classmates, though, be sure to check your course syllabus about what group work is or is not allowed in your class.

Tackling assignments

I recommend breaking up larger assignments into multiple steps. This is especially helpful if the assignment feels scary or overwhelming. If you break up an assignment into several smaller steps, it can feel much more manageable. If it still feels overwhelming, break it into even smaller pieces. On a day when you are really struggling, writing just one word might be a reasonable goal!

Breaking an assignment into several steps is also helpful because it gives your brain more time to make connections and find creative solutions, it allows you to identify potential problems and get help early, and it gives you more time to identify and correct your mistakes. For problem sets especially, it is often very helpful to have the problem stewing in the back of your mind for several days. It helps you make connections that aren't immediately obvious.

You should first look at the assignment soon after it's assigned or posted. Just read it without planning on doing it! That will allow you to identify how much time and which resources you will need and if you have any questions.

You should also separate your thinking and analysis time from the decision process of how to present or write up your answer. By separating these two very different tasks, you will likely have a much better result.

If you have a full or partial draft ready early, before the due date, you can use that opportunity to get feedback from the instructor or your classmates. You will also be able to catch more of your own mistakes if you look at it with a fresh eye on a different day than the day you wrote it up.

When you get feedback or a grade on your assignment, separate receiving the feedback from applying it. First

give yourself time to process your emotions. Later, you can revisit the feedback and decide how to revise your work or how to improve your next assignment.

Reading to write

As well as using them to learn the course material, you should also be using the course readings to learn how to be a better writer and communicator.

As you go through your university degree, you will be exposed to and asked to produce kinds of writing that might be new to you. In fact, this doesn't stop when you graduate! You may be asked to produce writing of different genres, for different communities and audiences, and with different goals. You will be taking many different courses in different fields throughout your degree, which have different genres, cultures, conventions, and more. For example, in some fields, such as English literature or syntactic theory, the convention is to present evidence and analysis together in every paragraph as you go, but in other fields, such as quantitative sociolinguistics or chemistry, the convention is to present your analysis all together in a special discussion section. If you also read and write in multiple language varieties, you may notice that different language communities prefer different strategies and rhetorical structures, as well. English academic writing usually uses explicit and direct argumentation, while other language communities use more inductive and deductive argumentation.

So how should you, as a student, learn to navigate all of these different conventions and genres? First of all, it's okay and normal if you're not perfect the first time you try. Learning to be a better writer is a lifelong skill.

In this textbook, I am trying to explicitly explain some of the conventions we use in morphology and syntax writing. But the best way to learn how to write in different genres is to read in those genres. The good news is that you are *already* reading in these genres through your course-assigned readings. The only thing you need to change is to start paying attention to how your readings are structured.

Here are some questions you can ask as you read the course readings:

- How is the reading structured? Do they tell it like a story? What 'moves' do they make?
- Is it an effective structure? Not all academic writing is well-written and sometimes bad writing can help you learn more! If it is effective, how can you use the same strategies in your own writing? If it is not effective, how would you improve it?
- How does the author group related ideas? How does the author move from one topic to another?
- How does the author guide the reader? What signposts do they give the reader to help the reader navigate the reading?

When you don't get the grades you want

Most students hit a point in their education (or really several points) where things get hard. This does not mean that you have "reached your limit" or that you're not smart enough, or anything like that. Learning does not happen in a straight line, and so these bumps happen. For me, looking back, it happened between 11th and 12th grade, between 3rd and 4th year university, and also in my MA degree. But if you approach the learning bump as part of the journey or as a problem to be solved, instead of a scary thing to run away from, that's what's really going to help you the most in life.

A lot of time, all you need to do to get past the bump is to not give up. Eventually things will click, and after you will look back and say, "Oh I wonder why I thought that was so hard!" Sometimes, it is also worth it to evaluate your study skills, for example changing when or how you're doing the textbook reading. If this is the first time you've struggled in a course, then you might be learning some new study skills for the first time (on top of learning the course material). That's good! But it is also hard. But the good news is that you can do hard things!

Study smarter, not harder

If it's time to re-evaluate your study skills, here are some tips to consider. One mistake a lot of students make is to study *more* but in an inefficient way. Instead of spending equal time on everything, figure out which parts need the most attention and focus on those.

- Identify what needs to be memorized and what needs to just be understood. Very little material needs to be memorized in linguistics. Instead, if you understand the reasoning behind the way we do things, it will be easier to remember, and it will also be easier to determine which tools are needed for a given problem and to apply your knowledge to new data.
- Try to study in active ways (like identifying key points and writing then down) instead of passive ways (listening to the lecture over and over).
- Try to study in little chunks with lots of breaks, maybe spread out over lots of days, instead of all at once. This will result in the material going into your long-term memory instead of your short-term memory, so that you don't have to study it *again* the next time it turns up in your degree.
- Try to identify specifically what you understand well and where you're stuck.
- Use more than one format and more than one sense. Read the material *and* listen to the material. Draw a picture, diagram, or chart that helps you organize the material in your mind. Read the material out loud. Different people may have different preferred learning styles, but *everyone* learns better when you combine multiple styles.
- If you're starting to feel tense or anxious, take a break, maybe go for a walk, have a glass of water, and

come back when you're feeling calm.

What to do if you're falling behind

For a variety of reasons, you might fall behind in one or more of your courses. Don't panic! The first thing you should do, as soon as you realize you've fallen behind, is reach out to your instructor. The earlier you ask for help, the more the instructor can do for you. Your instructor can help you determine if it's realistic for you to catch up, give you extensions on assignments, and help you decide which assignments to prioritize. If you wait until the end of term to ask for help, though, there is often very little your instructor can do at that point.

Depending on your exact circumstances, it is often better to focus on the upcoming coursework rather than trying to complete all of the missed coursework *and* follow the new material at the same time.

Here are some things you should do to help you make a plan on how to catch up:

- Consider why you got behind in the first place. Was it a temporary disruption to your schedule that is no longer an issue, or is it possibly going to continue disrupting your ability to complete your coursework? If the latter, make a plan to mitigate its disruption and/or reduce your commitments.
- Check the syllabus for the late submission policies for your course. If you've fallen behind for exceptional circumstances beyond your control, such as ill physical or mental health, there may also be additional accommodations available to you not listed in the syllabus.
- Look up the drop deadlines for your institution and determine what the penalties are for dropping at different dates. If you've fallen behind on multiple courses, it might make sense for you to drop one so you have more time to focus on the remainder.
- Check the syllabus to see how much each assignment is worth.
- Figure out how much work you need to catch up on and if it's realistic for you to be able to catch up. Calculate the weight of the assignments you missed and whether it's mathematically possible for you to pass and/or get your ideal grade.
- When making your plan for what to do next, prioritize the upcoming assignments that are worth more.

The linguistics department at your university probably has an undergraduate linguistics group or club. These groups may offer social gatherings, study events, and even tutoring. Getting connected with other students studying linguistics can help your studies!

First generation students

Are you the first in your immediate family to attend university? Will you be the first to *finish*? Will you be the first to go to graduate school, if you go? Are you the first in your family to attend university in the country where you're studying?

If you answered yes to any of these questions, you are a first generation student.

Although first generation students make up a sizeable population of the university, many first generation students feel lost and alone. There may be a lot of background knowledge that some of your classmates and instructors take for granted, but you don't know. Maybe you don't even know you don't know it!

There are lots of different kinds of first generation students, with different backgrounds and experiences. Many first generation students simultaneously also face other boundaries to university education, including financial boundaries and identities targeted by discrimination.

If any of this sounds familiar, don't despair. This *Becoming a Linguist* series is written with you in mind!

Although I hope this textbook will help, there are two important things you should do to help overcome these barriers. First, ask questions and be curious! Ask questions in class. Ask questions after class. Email your instructor. Post on discussion boards. Go to office hours. Your instructors' job is to help you learn, so ask them for help! Second, form a community around you. Here are some ways you can build community:

- Make friends with your classmates. Things are less scary when you're not alone.
- Join clubs. Look for undergraduate linguistics clubs and clubs where other people who share your identities meet up.
- Look for mentorship: by peers who are a little bit ahead of you in the program, by recent graduates, by people in the career you're interested in, or by faculty members.
- Join the Linguistic Society of America's First Generation Access and Equity Committee or attend some of their events.

A third thing you can do is explore some resources about how to succeed in university. Some suggestions are listed below in the *References and Further Resources* section, and even more are listed on the LSA's website.



Figure 1. The logo for the LSA's First Generation Access and Equity Committee, designed by Iyad Ghanim.

Key takeaways

- Complete the assigned readings before class.
- Read with questions in mind.
- Get lots of practice with problem-solving and analysis.
- Break up assignments into multiple steps over multiple days.
- Pay attention to the structure of assigned readings to help you learn to write better.
- Don't give up if you aren't getting the grades you want!
- Try to study in active ways, in short sessions spread out over many days.
- If you're falling behind, talk to your instructor and make a plan.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=223#h5p-12

References and further resources

For students

Charity Hudley, Anne, Cheryl Dickter, and Hannah Franz. 2017. *The Indispensable guide to undergraduate research: Success in and beyond college*. New York: Teachers College Press.

First Generation Access and Equity Committee. n.d. Resources for First-Gen Students in Linguistics. *Linguistic Society of America.* https://www.lsadc.org/af_membercommittee.asp?committeeid=2&sp=1

Macaulay, Monica. 2011. *Surviving linguistics: A guide for graduate students*, 2nd edition. Somerville: Cascadilla Press.

🔍 Receiving and giving effective feedback. Centre for Teaching Excellence, University of Waterloo. https://uwaterloo.ca/centre-for-teaching-excellence/catalogs/tip-sheets/receiving-and-giving-effective-feedback

1.5. BECOMING A LINGUIST: MYTHS AND MISCONCEPTIONS ABOUT ACADEMIC INTEGRITY

Connor Mark

While students might have an idea of what falls under academic integrity, some things are not always intuitive and so they may find themselves breaking rules by accident. Some common myths are...

Myths: Using resources

"I only need to cite people that I'm quoting."

You need to cite any source that has information that you are using. This could be statistics, theoretical information, examples, or examples. When in doubt, cite it. If you are unsure how to cite something properly, ask your instructor or a librarian. By citing, not only are we attributing information to people who have contributed to the field, but we are also keeping track of how information has been developed.

"Artificial Intelligence is allowed because it's like a spell checker."

As of now, there is no consensus on the use of AI for assignments. Check the syllabus and ask your instructor before using it. For example, while some instructors may allow you to use AI in specific contexts for specific purposes, you should assume it is not allowed for graded work by default. Writing about a course topic in your own words helps you to recognize what you do or don't understand, remember the material, and begin to think critically about it. Using AI to help you write runs counter to these goals.

"I can re-use a paper I wrote for another course in this course."

Submitting the same work to two different courses is considered self-plagiarism in many cases. It is considered

an academic offence because you would receive credit twice for doing the work once. It is good to see points of connection between your different courses, but if you re-submit a paper you wrote for another course, your paper will not demonstrate the new material you learned in this course. You also lose the opportunity to go deeper into the topic and to continue to improve your writing skills. In some cases, instructors may allow or even encourage you to build on and develop a project you started for another course or assignment; however, you should also check with your instructor first. You need to be transparent about which work is re-used and which is new.

Myths: Working with classmates

"I can work on assignments with classmates."

For the most part, assignments are meant to be done alone. Even collaborating and changing up your answers a bit counts as inappropriate collaboration. While you might encounter courses where collaboration on assignments is permitted, this is not the norm. Although working together can help you improve your learning in some cases, instructors also want to make sure you understand and are learning all of the material, not just some of it. Check your course syllabus for your individual course's policies.

"If my groupmates are not doing the work, I have to carry the group project."

On the contrary, writing the name of someone who has not contributed sufficiently on a piece of work could be a form of personation and thus a violation of academic integrity. Instead of trying to "carry" a group project, contact the instructor. The purpose of group projects is to learn how to collaborate with others on something larger and to learn from each other, which is facilitated through everyone putting in the work. In addition, it is not fair for your groupmate to get credit for an assignment they did not contribute to.

"I am only responsible for my own part of a group project."

If you divide up a group project, you are still responsible for all of the parts. You should read over your groupmates' parts and make sure you understand them and can explain them. If a groupmate plagiarizes, the entire group may be held responsible.

"I can have someone else sign the attendance sheet or

write the test for me."

No. Claiming to be someone else is not permitted, especially when it involves forging a signature. If you have to miss a graded component of the course, contact the instructor as soon as possible to see what can be done. The instructor is likely required to make reasonable accommodations for many kinds of personal or medical emergencies, according to the policies of your university.

The final word

As a rule of thumb, when you are unsure of what is allowed and what is not, refer to the course syllabus, your institution's academic integrity guidelines, and even your instructor if needed. Instructors would rather help you learn how to cite properly than try to bust you for failing to cite accurately by accident.

Things to do

While we often focus on what not to do, here are some things that you should do:

Read the syllabus

Read the syllabus and your institution's academic integrity guidelines. There is a lot of information, but learning about them and understanding the general principles of academic integrity will help you. This is especially important if you are a student from another country because academic integrity rules vary around the world.

Pick appropriate groupmates

Form appropriate groups for projects. Put otherwise, recall that students have differing goals in a course. While some students just want to pass, others want an A+. In the event that A+ students and D students are in the same group, there is a greater chance of students accusing others of not pulling their weight or asking their groupmates to do too much work.

Set internal deadlines

Set internal deadlines for group projects. Similarly, by coming up with your own internal deadlines for a group

project, you can be assured that everyone will have contributed sufficiently before the deadline is closing in, reducing the chance that people will perceive others as not helping.

Organize your notes

Keep your research notes organized. Clearly label your notes with their sources. Make sure your notes also clearly distinguish which parts are copied word-for-word and which are in your own words, for example, by using quotation marks. In some cases, an annotated bibliography might also be helpful. This applies to writing assignments the most, but keeping track of where you got your information and taking notes on it before you have started writing (or during the writing process) makes it easier to accurately cite sources and understand how theories and other ideas have developed.

Consider a citation manager

Consider using a citation manager. These tools can store all of your citations, allowing you to generate references from them. They can also be used as a database for sources that you have read in the past, allowing you to more quickly find the source of particular information if you need it again in the future.

Key takeaways

- There are a variety of reasons to cite things, including your own work.
- You must work alone unless you are given permission to work in a group. You should not put anyone's name on a group assignment if they did not contribute.
- If you're unsure of what's allowed and what's not, read the syllabus and assignment instructions. If necessary, ask your instructor as well.

Check yourself!

–

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=2346#h5p-37

References and further resources



University of Manitoba. Academic integrity. https://umanitoba.ca/student-supports/academic-supports/academic-integrity

◈ University of Manitoba Faculty of Arts. Academic integrity. https://umanitoba.ca/faculties/arts/student_resources/student_responsibilities_integrity.html



University of Manitoba Libraries. Citation management software: Intro to citation managers. https://libguides.lib.umanitoba.ca/citationmanagers

CHAPTER 1 PRACTICE EXERCISES

Data analysis

Exercise 1. Simplex or complex?

Beginner-Intermediate [1.1]

Are the following English words simplex or complex? If they are complex, segment them into morphemes using hyphens.

a. sofa
b. beautiful
c. bookshelf
d. pastries
e. recliner chair
f. reassessment
g. sunny
h. spilled
i. strawberry pie filling
j. distressing

Exercise 2. English analysis

Intermediate [1.1, 1.2]

Consider the following nursery rhyme.

Little Bo-Peep has lost her sheep,
And doesn't know where to find them;
Leave them alone, and they'll come home,
Wagging their tails behind them.

Brainstorm a list of things that someone would need to know about English in order to understand this nursery rhyme. Which of the things you listed pertain to morphology or syntax?

Communication and study skills

Exercise 3. Personal skills

Beginner [1.4]

What background skills, experience, and knowledge do you bring to this course that might help you succeed? For example, do you have experience with multiple languages or dialects? Do you have experience with pattern recognition and problem solving? Are you skilled at creative language play? Do you have good time management and organizational skills?

Exercise 4. Cultural reflection

Intermediate [1.4]

What is a traditional value from your culture? How will you apply this value during this course? For example, you could apply the Inuit value of working together for a common cause by establishing a study group or the Métis value of courage by participating in class in ways that stretch your comfort zone.

Exercise 5. Academic integrity

Intermediate [1.5]

Which of the following would count as an academic offence? Explain what you should do instead.

1. Using lots of quotations in your paper, with in-text citations and a bibliography.
2. Listing the headings or topic sentences of a paper word for word to summarize a text, without quotation marks.
3. Providing a paraphrase without quotation marks.
4. Working together with a classmate when group work is not allowed.
5. Using a word-for-word definition from wikipedia without citing it.
6. Submitting an answer composed by Artificial Intelligence without permission to do so.
7. Having a friend proofread your essay for grammar errors.

8. Listing the same ideas in the same order, but in your own words.
9. Visiting the university writing centre for advice.
10. Copying another student's answer on homework or a quiz.
11. Buying an essay from an essay-writing service.
12. Checking the spelling of a word in a dictionary without citing it.
13. Forgetting to include a references section in your paper.
14. Hiring a tutor who tells you the answers if you can't figure them out on your own.

Research and application

Exercise 6. WALS data analysis

Beginner [1.2]

Look at map 85A on the online language database WALS and answer the following questions.

> To calculate percentages, divide the number of languages with prepositions or postpositions by the total number of languages and then multiply by 100.

a. Based on the sample in this map, what percentage of the languages of the world use **prepositions**?

b. Based on the sample in this map, what percentage of the languages of the world use **postpositions**?

c. Are languages that use prepositions and languages that use postpositions spread randomly throughout the world, or do they seemed to be grouped according to some sort of pattern?

d. Scroll down to the list of languages and click on the name of a language from the list that you've never heard of before.

- What is the language called?
- Does this language use prepositions, postpositions, or something else?
- What language family does it belong to?
  (Hint: Look at the top of the page, just below the menu.)
- Where is it spoken?
  (Hint: Look underneath the map on the right side of the page.)

e. Write up your answer to all of the above questions in paragraph format, using in-text citations as

appropriate.
(Hint: Cite the information you got from this specific map as Dryer (2013) and the information coming from the WALS database as a whole as Dryer and Haspelmath (2013). This is according to the instructions for citing WALS on the WALS homepage.)

Language Journal

For your language journal, you should pick a language that you don't know much about. At the end of each chapter, there will be a few questions for you to research about your language. By the end of the semester, you will have created a profile of the morphosyntax of your language!

In the spirit of decolonization, I encourage you to pick a language that disrupts the status quo in some way, which can begin to provide you with a deeper understanding of language diversity and the unevenness of the language landscape of the world and in the field of linguistics, such as:

- A language that the student has never heard of before (which you can discover by searching a language database like WALS, Ethnologue, or Glottolog).
- A signed or tactile language.
- A language that is negatively affected by colonialism or globalization, especially if the student is or plans to be involved in community-led revitalization or documentation efforts.
- A non-standard variety of a language (e.g., AAVE or Michif French).

Exercise 7. Language profile

Beginner-Intermediate [1.2]

Research the basic background facts of your language. Where is it spoken? Approximately how many speakers are there? What language family does it belong to?

Exercise 8. Prepositional vs. postpositional

Beginner [1.2]

Does your language use prepositions or postpositions? Include an example that supports your answer.

2. THE SCIENTIFIC METHOD IN MORPHOSYNTAX

In this chapter, we will consider how we use the **scientific method** in morphology and syntax. We will look more closely at how the scientific method applies in morphology and syntax and which methods we should use. We will also consider the genre of scientific writing in morphosyntax, how to identify academic sources, and how to cite in linguistics.

Learning objectives

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

- Apply the **scientific method** to simple morphosyntactic questions, including formulating a **falsifiable hypothesis**.
- Describe the difference between **prescriptive** and **descriptive** approaches to grammar and explain why we use the descriptive approach in linguistics.
- Identify which information about a word would be encoded in the mental lexicon.
- Identify academic sources.
- Apply the conventions of scientific writing within the field of morphosyntax.
- Write in-text citations and bibliography entries according to convention in linguistics where appropriate.

2.1. LINGUISTICS AS SCIENCE

In Section 1.3, we defined linguistics as the **scientific** study of language. In this section, we will learn why we consider linguistics to be a science.

Often, when people think of science, they think of the physical sciences, which is the study of the physical properties of our universe. Although some subfields of linguistics, such as phonetics, include the study of physical properties, linguistics generally belongs to the category of social sciences. In the **social sciences**, we study the properties of societies. We can also categorize linguistics as a **cognitive science**, which is the study of the mind. Linguistics is the branch of cognitive science that studies what language can tell us about how the human mind works.

The scientific method

All of the types of science — physical, social, and cognitive sciences — are united by their methodology, which is known as the **scientific method**. The scientific method consists of four parts:

1. Observe data.
2. Formulate a **hypothesis**.
3. Test the hypothesis against more data.
4. Revise the hypothesis.

The scientific method can be treated as a cycle. After you revise your hypothesis in Step 4, you continue to collect more data, and test and revise your hypothesis, getting more and more precise and accurate with each revision. The second time through the cycle, and every subsequent time, you observe the data you collected from testing your previous hypothesis to formulate a new, revised hypothesis.

> In this textbook, we will be focusing on methodologies that use the scientific method. However, there are also other kinds of methodologies used within linguistics that do not use the scientific method and are equally valid. For example, in historical linguistics, the comparative method is used to reconstruct languages from before we have a written record. Although the comparative

method uses rigorous and systematic methods, it is not possible to test the hypotheses developed in this method, because we have no record of the languages being reconstructed. Likewise, some other subfields, like discourse analysis and linguistic anthropology, use qualitative methods akin to textual analysis. The methodologies that you choose to use depend on the kinds of questions you are asking and the resources available. Every methodology has its strengths and weaknesses. Our understanding of language will be strongest if we, as a community of linguists, investigate it from multiple perspectives, using many different methodologies.

An example of the scientific method in morphosyntax

Let's look at an example as applied to morphosyntax.

Step 1: Observe data

The first step is to observe some data. We will start with the data in (1).

(1) a. **We** wrote a story.

b. You gave **us** a gift.

Sentence (1a) includes the pronoun *we*, which means something like 'a group of people including the speaker.' Sentence (1b) includes the pronoun *us*, which also means something like 'a group of people including the speaker.' So an observation we could make about this data is that *we* and *us* are very similar in meaning.

Step 2: Formulate a hypothesis

The next step is to make a hypothesis. Since we observed that *we* and *us* are very similar, one possible hypothesis could be that they are perfect **synonyms**.

Hypothesis 1: *We* and *us* are perfect synonyms.

This hypothesis makes a prediction. If *we* and *us* are synonyms, then we might expect them to be interchangeable. That is, we might expect us to be able to switch one for the other without changing the meaning of the sentence.

Step 3: Test the hypothesis

So let's test out the predictions of our hypothesis! We will take the examples from (1), but switch *we* for *us* and vice versa and see what happens.

(2) a. *__Us__ wrote a story.

b. *You gave __we__ a gift.

When we switch *us* and *we*, the result, shown in (2), doesn't sound quite right, so we mark them with an asterisk symbol (*) to show that the sentences are not well-formed.

What have we figured out? Contrary to our first hypothesis, it doesn't seem like *we* and *us* are exactly the same. They are different in some way, which is why switching one for the other in (2) results in ill-formed sentences.

Step 4: Revise the hypothesis

The difference between the sentences in (1) and those in (2) is not a difference in meaning, but a difference in grammaticality. Perhaps, then, the difference between *we* and *us* is not a difference in meaning, but a difference in grammatical function.

Hypothesis 2: *We* and *us* have the same meaning but different grammatical function.

This hypothesis also makes a prediction. According to this hypothesis, if we change the structure of the sentence but keep the same meaning, it might affect whether we need to use *we* or *us*.

Step 5: Test the revised hypothesis

One way to change the structure of these sentences is to put them in **passive voice**. If we take the sentences in (1) and change them to the passive voice, keeping the same pronoun, we get the sentences shown in (3). If, on the other hand, we make them passive but switch the pronouns, we get the sentences in (4).

(3) a. *A story was written by **we**.

b. ***Us** were given a gift.

(4) a. A story was written by **us**.

b. **We** were given a gift.

This time, the predictions we made in Hypothesis 2 are consistent with our data! Sentences (1a) and (4a) mean the same thing but have different structure, but (1a) must have *we* and cannot have *us*, but sentence (4a) must have *us* and cannot have *we*. Likewise, sentences (1b) and (4b) also mean the same thing but have different structures and also use different pronouns.

And so on…

Even though our Hypothesis 2 ended up passing our test, that doesn't mean we're done. There are two main ways we might continue to refine our hypothesis.

First, notice that our Hypothesis 2 is rather vague. We have proposed that there is a difference in structure, but what exactly is this difference? If we were to continue on with this example, we might try to figure out how to describe the structural difference between *us* and *we*.

Second, notice that we've only looked at eight sentences. You could even say we've only looked at two sentences and variations on those two sentences! That's not very much data! Another way to continue revising and refining our hypothesis would be to test it against even more data, and different kinds of data, to see if our hypothesis continues to stand.

Key takeaways

- Linguistics is often classified as a science. It is sometimes classified as a social science and sometimes a cognitive science.
- In morphosyntax, we often use the scientific method to investigate the properties of language.
- The four steps of the scientific method are:

    1. Observe data.
    2. Formulate a hypothesis.
    3. Test the hypothesis against more data.
    4. Revise the hypothesis.

- The scientific method can be considered an iterative cycle. We repeat it, revising our hypotheses and getting more specific with each cycle.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://pressbooks.openedmb.ca/wordandsentencestructures/?p=85#h5p-16

2.2. EVIDENCE AND HYPOTHESES

As we use the scientific method to study language, there are a few things we need to keep in mind about the process.

Falsifiable hypotheses

When we formulate a hypothesis, it is important that we make a **falsifiable hypothesis**, which is a hypothesis that is logically possible to be proven wrong. It is important that your hypothesis is falsifiable so that you can do the next step in the scientific method, test your hypothesis! If it is impossible to fail your test, then perfoming the test doesn't help us figure anything out.

Let's take an example: our hypothesis is that ghosts exist but are invisible, have no smell or taste, make no sounds, and cannot interact with the physical world whatsoever. How could we test this hypothesis? We cannot see, smell, touch, hear, or taste a ghost. We cannot trap a ghost as evidence. This hypothesis is unfalsifiable. There is no test we could do that would show that it is wrong. Therefore, it is not a good hypothesis.



Figure 1: A decorative ghost. Photo by Juan Vargas. Used under Pexels license.

Confirmation bias

Another thing to watch out for when doing scientific research, or in fact any kind of critical thinking, is **confirmation bias**. Confirmation bias is a human tendency to look for evidence and information that is consistent with what they already believe.

Here is an example. Let's pretend you work at a bakery and you believe that the macarons are the best and most popular item sold at your bakery. Every time someone buys a macaron, you may think, "See! Macarons are the best! Everyone likes them!" But if someone buys something else, you may not notice it or dismiss it as an exception. This would be a simple example of confirmation bias at work. In order to counter against this confirmation bias, you could count how many times people buy each kind of treat. Sometimes, when you do this, you may be surprised at the numbers!



Figure 2: Four macarons. Photo by Arminas Raudys. Used under Pexels license.

Negative evidence

Another thing you could do to counter confirmation bias is to consciously train yourself to watch for **negative evidence**. Negative evidence is evidence that something is not possible, did not occur, or is absent. If you wanted to look for negative evidence that macarons are the most popular treat, you might count every time someone does *not* buy a macaron.

Negative evidence is very important in linguistic analysis. In linguistics, negative evidence often takes the form of ungrammaticality. When we test a hypothesis about the structure of language, we should consider *both* what patterns we predict to be possible and what patterns we predict to be impossible according to our hypothesis, and we should check for both.

In the last section, we compared the use of *we* and *us* by looking at both positive and negative evidence. The positive evidence were the examples of the sentences we could say, such as those in (1).

(1) a. **We** wrote a story.
b. You gave **us** a gift.
c. A story was written by **us**.
d. **We** were given a gift.

The negative evidence were the examples of the sentences that are awkward or unusual, such as those in (2).

(2) a. *__Us__ wrote a story.
b. *You gave **we** a gift.
c. *A story was written by **we**.
d. *__Us__ were given a gift.

If we only had the examples in (1), we wouldn't know for sure whether *we* and *us* were interchangeable or not. It could be just a coincidence that *we* was always used in subject position and *us* was always used in object position. It is the examples in (2) that show us that it matters what position they go in by showing us where they *can't* go.

Necessary and sufficient evidence

When you are making a hypothesis, you should think about what kind of evidence would allow you to draw conclusions about the accuracy of your hypothesis. There are two kinds of evidence to consider: **necessary evidence** and **sufficient evidence**.

Necessary evidence is evidence that must be the case in order for your hypothesis to be true. For example, let's say that you think your sister ate the last cookie from the cookie jar. In order for this hypothesis to be true, the cookie jar must be empty. If there are any cookies left in the cookie jar, then it is impossible that your sister ate the last cookie. An empty cookie jar is necessary evidence that your sister ate the last cookie from the cookie jar.



Figure 3: A cookie jar full of cookies. Photo by Alina Matveycheva. Used under Pexels license.

However, even if the cookie jar were empty, this would not be sufficient evidence that your sister ate the last cookie. Maybe someone else ate it! In order to prove that it was your sister who ate the last cookie, you'd have to catch her in the act, or at least with crumbs around her face.



Figure 4: A girl eating a cookie. Photo by Charles Parker. Used under Pexels license.

But even so, in order to prove she ate the *last* cookie (and not the second-last one), you need to catch her taking the last cookie from the cookie jar and eating it. This would be sufficient evidence that your sister ate the last cookie from the cookie jar.

Let's apply this to a linguistic example. In the last section, our initial hypothesis was that *we* and *us* are perfect **synonyms**. In this case, the fact that *we* and *us* have the same meaning is necessary evidence. It could not be true that they are perfect synonyms if they do not share the same meaning. It is not possible to get sufficient evidence, though. In order to prove that they are perfect synonyms, you would have to show that they are always interchangeable in every possible context. Since there are infinite possible contexts, it is not possible to test them all.

When we make our hypothesis, thinking about what constitutes necessary and sufficient evidence can help us decide what data to look for to test our hypothesis. It can also help us decide whether the hypothesis is falsifiable. If a hypothesis is falsifiable, it should be possible to identify the kind of evidence that would prove its opposite.

Key takeaways

- Hypotheses should be falsifiable. That is, it should be possible to prove them wrong.
- Confirmation bias is when you only notice the evidence that supports what you already believe. Everyone has confirmation bias to some extent. To counter confirmation bias, we should intentionally look for evidence that contradicts what we believe.
- Negative evidence is evidence that something is not possible or does not exist. In linguistics, negative evidence often takes the form of ungrammatical utterances.
- Necessary evidence is evidence that must be true in order for your hypothesis to be true.
- Sufficient evidence is evidence that, on its own, is enough to prove your hypothesis.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it

online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=526#h5p-15

References and further resources

For a general audience

🧠 Veritasium. 2014. The most common cognitive bias. YouTube. https://www.youtube.com/watch?v=vKA4w2O61Xo

Academic sources

Davis, Henry, Carrie Gillon, and Lisa Matthewson. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language* 90(4): e180–e226. https://muse.jhu.edu/article/563094/summary

2.3. PRESCRIPTIVISM AND DESCRIPTIVISM

> How you language is beautiful. Don't let anyone tell you your language is wrong. Your languaging is the story of your life.
>
> -Jon Henner

A lot of people, when they think about grammar, they think about a set of rules that they have to follow. Many people think they are "bad" at grammar. Others like to correct others' grammar. Some people may have grammar pet peeves that get them especially annoyed. This is known as **prescriptive grammar**—creating and enforcing rules that tell people how they should use language.

Linguists think about grammar in a very different way. Because we approach grammar scientifically, we do not make value judgments about the ways people use language. Instead, we observe it and try to describe it. This is known as **descriptive grammar**.

Language variation and identity

Language is always changing, and language is not uniform. Everybody uses language in slightly different ways, and sometimes in very different ways. We use these differences in the ways we use language to perform our identity—to let other people know how we fit into our society. The ways we use language can indicate to others where we're from, how old we are, whether we are cool or smart, our gender identity, and every other aspect of our identities that you can think of. The study of how our language varies is a major part of the field of **sociolinguistics**.

Unfortunately, not all identities are equally valued by society. The value society places on certain kinds of people often gets transferred to the language use associated with those people. Thus, the varieties of language used most often by the most powerful and prestigious people get viewed as "correct" or "standard" while the varieties of language used by the less powerful and prestigious people get viewed more negatively.

However, when studied objectively, all varieties of language follow systematic rules. All varieties allow for clear and creative communication. They are all equally valid, and none of them are better or worse than the others.

These different varieties do communicate different things, though. They communicate your identity, sometimes intentionally and sometimes unintentionally! They communicate whether you are rich or poor, whether you are well-educated or not, whether you fit in with a particular crowd or not. Likewise, the people you interact with perceive you in certain ways because of your language, sometimes intentionally and sometimes unintentionally. The most proficient language users aren't the ones who always use the most prestigious variety of language; they're the ones who are able to use many different varieties well and are able to choose which varieties are most effective for different contexts.

It is important to remember, though, that not everyone has equal access to the prestigious varieties of language. Children do not get to choose where they are born and grow up, the language variety used by their peers or parents, or the quality of their education. The varieties of language you use has less to do with your intelligence or attention to detail, and more to do with the communities you grew up in. It is possible to learn another variety of language, of course, just as it is possible to learn another language. But children who grow up being exposed to the prescriptively correct and prestigious forms will not have to put in as much effort to produce the prescriptively correct forms in formal situations. Not exposing children to culturally significant varieties of language can also cause harm, though, as children need the tools to be able to explore and express their identities.

Negative language attitudes are pervasive in our schools and in the media, and so it is natural to have internalized some of these attitudes yourself. However, it's worth thinking about where those attitudes came from and the power structures they uphold.

Does descriptivism mean I can say whatever I want?

When we say that all varieties of language are valid, what we mean is that all varieties are systematic and rule-governed. It does not mean that our language has no social consequences. For example, a slur may be constructed in a way that follows morphological rules, but it is still rude to say and can cause harm to you, your community, and the people you say it to.

Grammar checking computer programs

It is likely that you have seen ads for computer programs that check your writing. Many word processors have basic spelling and grammar checkers built in. You can also buy subscriptions to more powerful and sophisticated ones, some of which are even powered by artificial intelligence. I encourage you, as a linguist and as a student, to critically engage with these grammar checking programs. Most likely, these programs blindly enforce prescriptivist rules without a deeper understanding of the patterns of language, including prescriptivist rules with classist and racist origins.

If you do use a grammar checking program, think about every suggestion it gives you before adopting it. You might even consider looking at the settings of your program to see if you can customize some of your settings. For example, I have turned off the setting that marks passive voice as an error.

Case study on *ain't*

Let's look more closely at one non-prestigious grammatical construction, *ain't*. *Ain't* is widely used in several dialects of the United States, including Appalachian English and African American Vernacular English. The following quote describes its use in Appalachia. According to this description, using *ain't* is an important part of showing that you belong to the community.

> *Isn't, aren't, hasn't, haven't* are strictly avoided. *Ain't* or *hain't*, depending on the need for emphasis, is used universally in the place of these "furrin" contractions. The bright mountain boy eager to improve his English but determined to avoid the use of *ain't*, is constrained to use *he's not* and *they've not* in lieu of *he isn't* and *they haven't* if he wishes to escape the damnatory label of being "stuck up."
> -Cratis D. Williams (1962: 16), as cited in Montgomery (2014: 31).

Outside of Appalachia, however, *ain't* is considered prescriptively incorrect. But if you look more closely, you'll realize that it actually fills a gap in standardized English.

Ain't is an example of a contraction, which is a shortened form of a common phrase into a single word.

We use several contractions in standardized English, as shown in Table 1, but there is a gap! We don't have a contracted form of *am not*.

Table 1. Some contractions of standardized English

Full phrase	Contracted form
am not	?
are not	*aren't*
is not	*isn't*

It is not surprising that some dialects of English have developed a contraction that fills that gap. Some dialects use *ain't*, while others, such as Irish English, use *amn't*.

Where do prescriptive rules come from?

You may have been taught many prescriptivist rules about "correct" language in school. If you studied English formally, you may have been taught some of the following rules. In this section, we will discuss where these rules came from.

Table 2. Examples of prescriptive rules in English

Rule	Prescriptively incorrect example	Prescriptively correct example
Do not split infinitives.	*I want **to quickly eat** so I'm not late.*	*I want **to eat quickly** so I'm not late.*
Do not end a sentence with a preposition.	*That's the person I sat **next to**.*	*That's the person **next to whom** I sat.*
Avoid the passive voice.	*This book **has been read by many students**.*	***Many students have read** this book.*
Say *and I* not *and me*.	*You walked past my brother **and me**.*	*You walked past my brother **and I**.*
Do not use double negatives.	*I did**n't** say **nothing**!*	*I did**n't** say **anything**!*

Imposing rules from other languages

Some prescriptivist rules are imported from other languages. For example, through the Enlightenment and Romantic periods, Greek and Latin-speaking scholars imposed Greek and Latin spelling and grammar rules on English. Scholars (sometimes correctly, and sometimes incorrectly) inserted silent letters into words to reflect

their Greek or Latin origins. This is where the silent <b> in words like *debt* and *doubt* come from. But they didn't stop at spelling reform!

In Greek and Latin, infinitives are one word, and so a word cannot intervene in the middle of an infinitive construction. But in English, an infinitive is two words (*to* + a verb). These scholars decided that, if a word cannot go in the middle of an infinitive construction in Greek and Latin, we shouldn't be able to do so in English, either.

Another rule that has been imported is the prohibition against ending a sentence with a preposition. This is actually a pretty unique property of English grammar that is not possible in many other languages. Again, scholars began teaching that you shouldn't end a sentence with a preposition to make English more like Latin and Greek.

Arguably, these scholars imposed Latin and Greek rules on English because they wanted to show off their higher education. At the very least, these rules imply that Latin and Greek are better than English. The prestige of Latin and Greek is not as strong as it used to be, but these rules get repeated by teachers and editors who may or may not be aware of their origins.

Language change

Language is always changing, and young people of every generation coin new words and even develop new grammatical patterns. However, sometimes the new grammatical forms aren't accepted right away, and the new form is labelled as "incorrect."

One example of this is the comparative form of the word *fun,* as shown in Table 3. Younger speakers produce words such as *funner* and *funnest*, especially in informal contexts, whereas older speakers only produce *more fun* and *most fun*.

Table 3. Change in the comparative and superlative forms of fun

	Older variant	Newer variant
Comparative form	*more fun*	*funner*
Superlative form	*most fun*	*funnest*

Stylistics and rhetoric

Another source of prescriptivist rules is from stylistics and rhetoric, or the study of how texts are structured in order to be the most effective and persuasive. In stylistics and rhetoric, literary scholars may give you advice about what kinds of sentences and words to use because of the effect they have on the reader. For example, one piece of advice that is often given is to avoid the passive voice. One justification for avoiding the passive voice is

because it removes responsibility and accountability from the person who performed the action. This critique is often directed at the news, for example, for statements like those in (1). In (1b), the actions of the police officer are downplayed.

(1) a. A police officer shot a bystander.

b. A bystander was shot by a police officer.

However, sometimes the actor is not known, or if it is known, is not important. In these cases, the passive voice is definitely appropriate. Furthermore, it is possible to downplay the actor without using the passive voice. Consider this quote from NASA administrator Bill Nelson, shown in (2).

(2) We have had mistakes done in the past.

(Watts 2024)

The sentence in (2) makes it sound like mistakes just happened, and does not lay the blame on anyone, but it doesn't actually use the passive voice. Therefore, instead of a blanket rule to avoid the passive voice, a more appropriate recommendation is to be aware of whether you are downplaying the actor of a sentence and whether doing so is appropriate for the context.

Another justification given for avoiding the passive voice is the argument that the active voice gives a more vibrant and active feel to your writing. While this might sometimes be the case, it is possible to write dull sentences in active voice, as in (3a), and vibrant sentences in passive voice, as in (3b). Furthermore, a vibrant and active feel is not always the one you will be aiming for!

(3) a. The wind bent the tree.

b. The tree was bent and ravaged by the constant windstorms.

Although rules coming from stylistics or rhetoric might be reasonable advice in some cases, these are just guidelines given for a particular purpose rather than rules about how language is structured. In syntax, we are more concerned about the possible structures than about which structure is the most persuasive or clear. Although some styles of writing may be clearer or more appropriate in a given context, that doesn't make the other ways of writing "wrong." Unclear, ambiguous, or misleading sentences that are nonetheless grammatical still provide us with valuable information about how language is structured and processed.

Hypercorrection

Another source of prescriptivist rules is from something called **hypercorrection**, which is when a rule is over-applied and extended to new contexts. The rule about *and I* originated as a hypercorrection.

The difference between *I* and *me,* just like the difference between *we* and *us*, is about whether the pronoun appears in subject or object position. *I* generally appears in subject position, while *me* appears in object position.

(4) a. I ate dinner.
b. *Me ate dinner.
c. Dinner was eaten by me.
d. *Dinner was eaten by I.

Originally, the pattern was the same whether or not the pronoun came after *and*. But, in English, *me* is the default form that shows up when it is neither subject nor object, and for some speakers it started showing up after *and* in subject position, like in (5b). Other people started correcting that usage, and then people started saying *and I* in every context, even in object position, like (5d). In these questions, the % indicates that the sentence is accepted by some speakers but not all.

(5) a. My friend and I ate dinner.
b. %My friend and me ate dinner.
c. Dinner was eaten by my friend and me.
d. %Dinner was eaten by my friend and I.

This is an example of a hypercorrection. This hypercorrection has been so pervasive that it has arguably changed some people's grammars.

Linguistic imperialism and discrimination

Finally, some prescriptivist rules arise due to contact between two different language communities with a power imbalance. In these situations, the grammar of the language community with less power is often labelled as incorrect or even illogical. Often, this happens alongside other forms of racism and/or colonialism. For example, because the British Empire colonized so many countries around the world, there are many different global dialects of English. Many of these dialects can be considered mixed languages, which combine properties of two different languages. However, although these dialects are expressive and creative, follow systematic rules,

and are valid language varieties, they are often described as "broken English." For example, Nigerian Pidgin English is spoken by over half of the population of Nigeria (Faraclas 2008), including university graduates, professors, lawyers, and journalists, but it still has the reputation of being the language of the non-literate (Akande and Salami 2010).

These language attitudes can have significant impact on society. For example, in the infamous acquittal of George Zimmerman which launched the Black Lives Matter movement, the six hours of testimony by the star witness Rachel Jeantel were disregarded by the jury. Public remarks by a juror indicate that they disregarded her testimony because of her speech (Rickford and King 2016).

A published academic article written in non-standard English? Is it even possible? Check out Young (2010) for an overview of how descriptivism can work in everyday life, not just linguistics—written *in* non-standard English.

What should I do when I notice discriminatory bias in myself?

First of all, recognize that the emotions you are feeling are real and are the result of a number of complex factors, many of which you did not choose, such as the environment you grew up in. We may be tempted to deny or ignore unpleasant thoughts, including our biases, but doing so does not get rid of them. Instead, it just keeps them in our subconscious, where we can't do anything about them.

Instead, when you notice a discriminatory bias in yourself, do two things:

1. Acknowledge that your feelings and/or thoughts have a discriminatory bias.

2. Intentionally choose actions, reactions, and words that counteract the bias.

Later, when you are in a safe space with the time and emotional energy to do so, spend some time thinking about where these thoughts and feelings came from. You should do so in private, either alone or with trusted friends. However, if you do so with friends, keep in mind that such discussions have a high emotional cost for your friends who face discrimination on a regular basis,

which you may or may not be aware of. Always check that your friends are able and willing to participate in such discussions.

Key takeaways

- We use language to perform our identity. Just like there are not good and bad identities, there are not good and bad forms of language.
- Not all identities are equally valued by society. The value society places on certain kinds of people gets transferred to the language use associated with those people. The language used by more powerful and prestigious people is often considered more "correct" than other varieties, even though all varieties are systematic and rule-governed.
- In linguistics, we want to describe how language is used, called the descriptive approach, rather than judging how language should be used, called the prescriptive approach.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=88#h5p-17

References and further resources

Comedy and satirical linguistics

🎉 Phlogiston, Phineas Q. 2008. Cartoon theories of linguistics part XI – Prescriptivism vs. descriptivisim. *Speculative Grammarian* CLIV(3). https://specgram.com/CLIV.3/04.phlogiston.cartoon.xi.html

🎉 Starkey, Ryan. 2019. Linguistic prescriptivists make terrible zoologists. *Starkey Comics*. https://starkeycomics.com/2019/04/18/linguistic-prescriptivists-make-terrible-zoologists

For a general audience

🔍 Bee, Zoe. 2021. Grammarly is garbage, and here's why. https://www.youtube.com/watch?v=Q5rB9jDbTPU

For linguistics students

🔍 Charity Hudley, Anne, Christine Mallinson, and Mary Bucholtz. 2022. *Talking college: Making space for Black language practices in higher education.* New York: Teachers College Press.

🔍 Curzan, Anne. 2014. *Fixing English: Prescriptivism and language history.* Cambridge: Cambridge University Press.

Sources for examples

Watts, Marina. August 25, 2024. Nasa makes decision to bring home astronauts who have been stuck in space for 80 days. *People*. https://people.com/nasa-makes-decision-to-bring-home-astronauts-stuck-in-space-for-80-days-8701062

Academic sources

Akande, Akinmade T. and L. Oladipo Salami. 2010. Use and attitudes towards Nigerian Pidgin English among Nigerian university students. In *Marginal dialects: Scotland, Ireland and beyond,* ed. Robert McColl Millar, 70–89. Aberdeen: Forum for Research on the Languages of Scotland and Ireland.

Faraclas, Nicholas. 2008. Nigerian Pidgin English: Morphology and syntax. In *Varieties of English 4: Africa, South and Southeast* Asia, ed. Rajend Mesthrie, 340–367. Berlin: Mouton de Gruyter.

Montgomery, Michael. 2014. Hain't we got a right to use *ain't* and auxiliary contraction?: Toward a history of negation variants in Appalachian English. *Southern Journal of Linguistics* 38 (1): 31–68.

🔍 Rickford, John, and Sharese King. 2016. Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language* 92 (4): 948–988. https://www.linguisticsociety.org/sites/default/files/Rickford_92_4.pdf

Young, Vershawn Ashanti. 2010. Should writers use they own English? *Iowa Journal of Cultural Studies* 12 (1): 110-117.

2.4. THE LEXICON AND DICTIONARIES

When we learn a language, there are two types of things we know. First, we know the descriptive grammar rules of the language, which we will learn more about in Chapter 3, and we know vocabulary from the language.

We store our vocabulary in our **mental lexicon**. The lexicon is the part of our memory where we store our knowledge of all the words we know. This includes how to pronounce the word, what the word means, and anything else about the word that is not predictable. That is, anything about the word that we have to memorize rather than that we can predict based on the grammar rules of the language. For example, for a noun with an irregular plural like *child ~ children* or a verb with an irregular past tense like *go ~ went,* we would have to store the irregular form in our lexicon. But for nouns with regular plurals like *cat ~ cats* and verbs with regular past tenses like *walk ~ walked*, we would not store it, instead we would build it from the rule. Likewise, some complex words have unpredictable or **non-compositional meanings**. For example, the word *edible* doesn't simply mean "can be eaten"; it means "can be eaten without ill effect to your health." The idiosyncratic component of meaning must also be stored in the lexicon.

Psycholinguistic studies show that productive complex words that are frequent may also have their own entries in the lexicon and may be accessed as unanalyzed wholes (Sims-Williams 2021). This means that in future language change, the stem and the complex word created from it might take separate paths — one can pick up idiosyncratic meanings or take on changes in pronunciation that the other does not. Some examples of this include *brother*, which formerly had an irregular plural *brethren*. The irregular plural picked up an idiosyncratic meaning (religious co-affiliation) and a new, regular, plural form, *brothers* was coined. Likewise, the plural of *cloth* used to be *clothes*. The irregular plural picked up the idiosyncratic meaning of *garment* and a new plural, *cloths*, was coined.

Deciding exactly which parts of language are productive, and can be built from our grammar rules, versus which parts of language are memorized in our mental lexicon is an important question in the study of morphosyntax. Another related question is *when* we access the lexicon as we build our sentences and words. There are three main options:

1. We access the lexicon only at the beginning. We pick out which words or morphemes we are going to use at the beginning of the process, and then we move them to a workspace where we can combine them to build a structure.
2. We access the lexicon only at the end. We build a structure using abstract pieces, and then we plug in the appropriate words at the end.
3. We access the lexicon multiple times throughout the process, inter-weaving the selection of words and the building of structure.

Dictionaries

A dictionary is kind of similar to a mental lexicon, because it also records the vocabulary of a language. But there are some important differences between a mental lexicon and a dictionary, as summarized in Table 1.

Table 1. The differences between mental lexicons and dictionaries

Our mental lexicons	**Dictionaries**
Created instinctively and subconsciously based on daily use.	Created for specific purposes by **lexicographers**.
Includes slang and informal language.	Less likely to include slang and informal language.
Instantly updates with exposure to new language innovations.	Lags a few years behind language innovations.
Only includes terms known by the language user.	May include technical and archaic terms.
Includes **idioms** and longer pieces of non-compositional meaning.	Rarely includes idioms.
Highly individual.	Tries to describe standard or common usage.
Linked to detailed information about the cultural background and real-world knowledge.	Usually minimal information about the cultural background and real-world knowledge included.
Includes all meanings and uses of a word.	Includes only the most common meanings and uses of a word.
May not include spelling and does not include **etymology**.	Usually includes spelling and sometimes includes etymology.

Some of the things we will learn together are things that you might be able to look up in a dictionary, like the part of speech (noun, verb, etc.) of a word. But in this course (and most linguistics courses), we want to learn how to figure it out ourselves, not how to look it up. There are a few reasons for this. First of all, dictionaries aren't always accurate. As mentioned in Table 1, dictionaries often lag behind language change and also only include the most common or standard patterns of language. In our linguistic analyses, we want to consider all the possible structures, not just the most common ones. Second, as a linguistics student, you are training to become the kind of person who will one day be qualified to *write* dictionaries. You can look things up in a dictionary to confirm your analysis or get a second opinion, but if you depend only on the dictionary, then you are skipping part of the learning process!

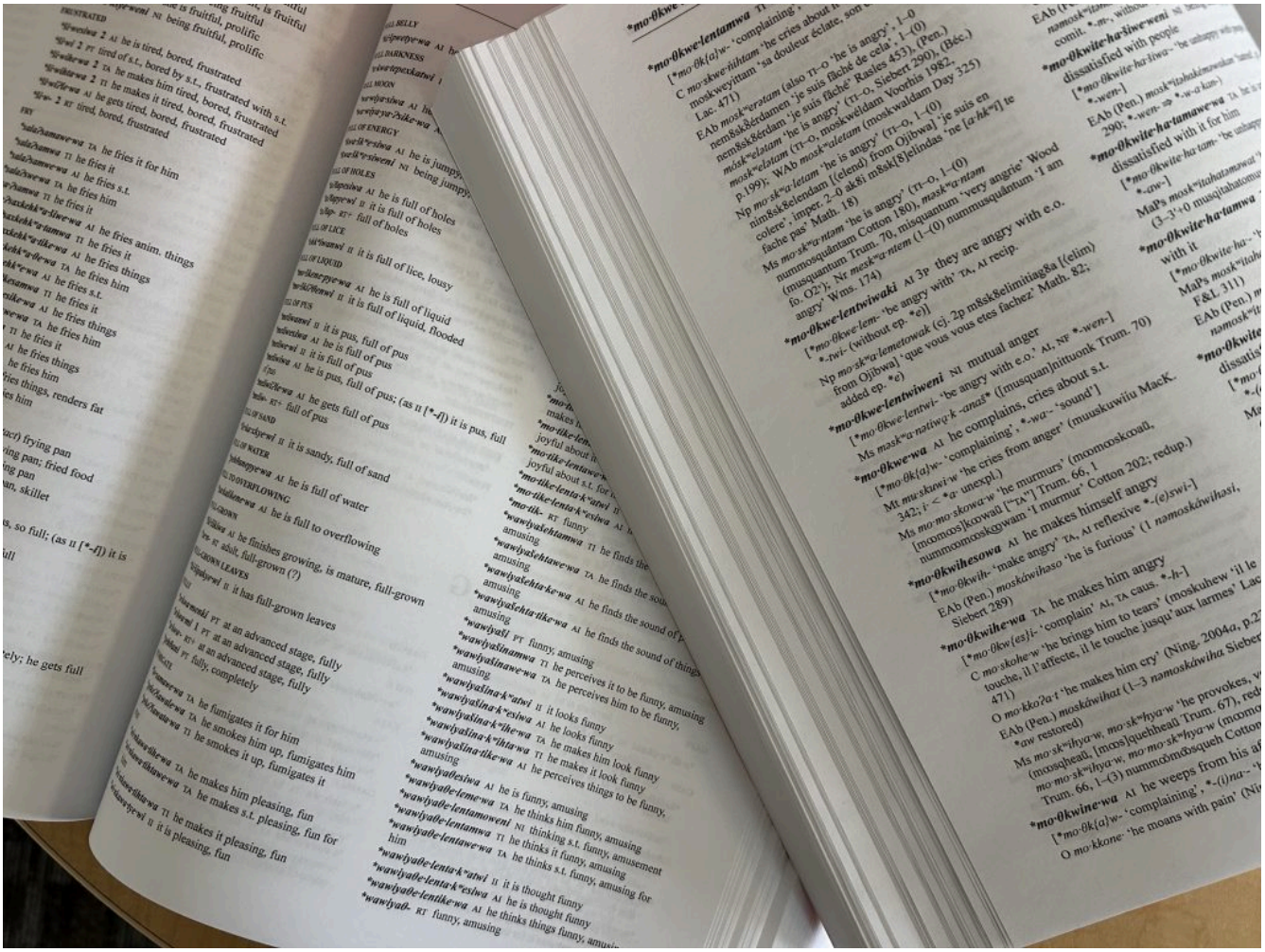


Figure 1: *The Dictionary of Proto-Algonquian* by David Pentland. This dictionary includes a lot of etymological information about numerous Algonquian languages, which would not be included in the mental lexicons of speakers. Photo by Julie Doner. Reproduced with permission.

Key takeaways

- We store our vocabulary and all of the parts of language that are not predictable from rules in our mental lexicon.
- Whereas dictionaries are curated intentionally for specific purposes, mental lexicons are created subconsciously and includes all of the forms known by the language user.
- You should avoid looking up the answers for the exercises in this course in the dictionary.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1310#h5p-21

References and further resources

For a general audience

🧠🔍 Curzan, Anne. 2014. What makes a word "real"? TedxUofM. https://www.ted.com/talks/anne_curzan_what_makes_a_word_real

For linguistics students

🧠 Anderson, Catherine, Bronwyn Bjorkman, Derek Denis, Julianne Doner, Margaret Grant, Nathan Sanders, and Ai Taniguchi. 2022. Section 7.4: The mental lexicon. *Essentials of Linguistics, 2nd edition*. https://ecampusontario.pressbooks.pub/essentialsoflinguistics2/chapter/7-4-the-mental-lexicon

🧠 Anderson, Catherine, Bronwyn Bjorkman, Derek Denis, Julianne Doner, Margaret Grant, Nathan Sanders, and Ai Taniguchi. 2022. Section 7.10: Why not the dictionary? *Essentials of Linguistics, 2nd edition*. https://ecampusontario.pressbooks.pub/essentialsoflinguistics2/chapter/7-10-why-not-the-dictionary

Sources for examples

Pentland, David. 2023. *Proto-Algonquian Dictionary: A Historical and Comparative Dictionary of the Algonquian Languages*. Winnipeg: Algonquian and Iroquoian Linguistics.

Academic sources

Sims-Williams, Helen. 2021. Token frequency as a determinant of morphological change. *Journal of Linguistics* 58 (3): 571–607.

2.5. BECOMING A LINGUIST: IDENTIFYING ACADEMIC SOURCES

In many of your university-level assignments, you will be asked to engage with academic or scholarly sources. In fact, unless you are told otherwise, you should probably assume that you will be expected to use academic sources by default. In this section, we will learn how to identify academic sources and what the review process is like for academic sources.

What is an academic source?

One of the main ways that an academic source is different from other kinds of publications is that an academic source is usually focused on **original research**; that is, the author of the source is not only reporting information that is already known about the topic, but is coming up with something new. For example, they could have gathered new data, developed a new theory, or extended an old theory to a new context. This means that academic sources are often very narrow in topic and quite specialized.

Another way academic sources are quite different from other kinds of resources is the **peer review** process. When an article or another piece of scholarship is peer reviewed, the editor will contact other experts on the topic and ask them to read and critique the manuscript. There will usually be two or three reviewers, who will provide detailed comments on the manuscript and give a recommendation of whether it is suitable for publication. Often, an article will go through multiple rounds of review before it gets published. Different journals and publishers have slightly different processes, but most often the process will be either blind, which means the author doesn't know who the reviewers are, or double-blind, which means the author doesn't know who the reviewers are *and* the reviewers don't know who the author is.

This process is most typical of an academic journal, but there are other kinds of sources worth mentioning, as well. Most books at a typical bookstore do not describe original research or go through a peer review process, but academic books do. Dissertations include original research but don't technically go through a peer review process; however, they are defended by the author and approved by a panel of experts, and so they have roughly equivalent status. On the other hand, they are often not as polished as published research.

In contrast, other kinds of publications do not present original research and are not peer-reviewed. Most books in a typical bookstore and the magazines at the check-out counter are what would be called "popular publications." They are written by professional writers—not necessarily an expert in the topic—for a general audience. Although editors fact-check the content, they are not peer reviewed by experts and have few citations. Trade publications fall in between scholarly and popular publications. They are written for a specific

group of people, and so are more specialized, but they still do not report original research, nor are they peer reviewed. An example might be a magazine for dentists summarizing the most recent dentistry tools and techniques.

Table 1. Kinds of publications

	Academic source	**Popular publication**	**Trade publication**
Content	Original research	News, opinions, and trends	News and opinions about a specific profession or area
Audience	Researchers	General	People belonging to a particular profession
Author	Researchers	Journalist or professional writer	Member of the field and/or professional writer
Citations	Extensive	Few	Few
Review	Peer-reviewed	Fact-checked	Fact-checked

There are also a lot of sources that are not peer reviewed but are sometimes used in academic writing. For example, researchers might post their work in progress on their website or on repositories before it is peer reviewed. Researchers also present their work in progress at conferences, and sometimes write up their presentation in a conference proceedings paper. Depending on the conference, proceedings papers may or may not be peer reviewed. Occasionally, you may even see someone cite a blog post, a social media post, an email, or even a chat in the hallway (usually cited as *personal communication*, abbreviated *p.c.*). This is because, no matter how prestigious or not your source is, you need to give appropriate credit.

A common place for linguists, in particular, to post and share their work in progress is on LingBuzz. Although many of the articles are legitimate work-in-progress by academic linguists, most of this work is not peer-reviewed (yet) and anyone can post an article.

When you are doing research, part of your job as a researcher is to evaluate the reliability of your sources. Any source potentially has errors, but a source that has been peer reviewed likely has fewer errors than one that has not been peer reviewed, and an informal source prepared by an expert is more likely to have fewer errors and more nuance than one prepared by a non-expert. Part of how you evaluate the source is by checking whether it is peer reviewed and whether it is authored by an expert. Another big part of evaluating sources is using your critical thinking skills to evaluate the content of the source itself. However, as a student, it is best to use peer-reviewed sources as much as possible. First of all, as a student, you are less likely to identify errors and

inconsistencies than someone more experienced. Secondly, if you are doing research for an assignment, likely your instructor wants you to practice identifying and using academic sources.

The peer review process on Wikipedia

Most of us have looked something up on Wikipedia—probably even this week! We rely on it frequently in our daily lives, but professors will often say not to use it for your assignments. How come Wikipedia is good enough for our day-to-day, but not for your homework?

Wikipedia uses a crowd-source volunteer model. That means that anyone with Internet access can create a new Wikipedia page or edit an existing one. You do not even need to login! However, anyone can also flag your edits as incorrect. With millions of users accessing Wikipedia every day, many errors get flagged and corrected very quickly.

However, there are two very big differences between the review process of Wikipedia and the peer review process. First of all, review on Wikipedia is not systematic. It is possible that a minor edit on a little-used page might get missed for quite some time. Second, Wikipedia is not necessarily reviewed by experts. There are some experts on Wikipedia, but there are also lots of amateurs. This means that it is possible that a commonly-believed falsehood might survive on Wikipedia if the majority of volunteer editors believe it to be true.

The kind of false information most likely to survive on Wikipedia is a plausible-sounding fact that most people don't know. For example, in *How I accidentally started a Wikipedia hoax about Amelia Bedelia*, journalist EJ Dickson describes a prank edit she and a friend, Evan, posted in 2009. Their prank edit claimed that the author of the *Amelia Bedelia* children's series, Peggy Parish, spent some of her childhood in Cameroon and based her series on a Cameroonian maid. EJ and Evan expected it to be flagged and removed right away, but not only did it survive Wikipedia's fact-checkers for 5 and a half years, it got repeated by several reputable sources, including Peggy Parish's own nephew!

It is still true that Wikipedia can be a very useful source in many ways. However, for anything that matters, you should double-check the information. Edits in Wikipedia are supposed to have sources listed (although they don't always). Follow the citation to the original source and double-check that the source is reputable and that it actually says what Wikipedia says it says, and then cite the original source.

How to identify an academic source

If you read enough academic papers, you will probably start to notice some patterns about who publishes a lot on certain topics and where they publish, and start to form an understanding about which journals are the most prestigious. However, for most students, this doesn't really start happening unless they pursue graduate studies. How do you decide who is an expert and which journals are reputable when you're just starting out?

The author

One of the first steps is to look at the author. Read their bio and google them. Some things to check:

- Do they have a university affiliation?
- Is their research specialty related to the topic of the source?

Graduate students and professors writing in their research area are likely reliable authors.

The web host

Where did you find the article? If you found it through your university library, that is likely a good sign. If it is hosted on an academic publisher's website, that is also a good sign. However, watch out for websites where people create profiles and share their own research, like ResearchGate and Academia.edu. Many legitimate researchers do post their published and unpublished research on such websites, and Google Scholar searches will find results from those websites. But anyone can create a profile and post an article. Double-check the author's credentials before citing work posted by the author, especially if it is unpublished.

The publisher

If the paper is formally published, who published it? If it's published by a university press, it is likely a scholarly source. Additionally, some for-profit publishers specialize in academic publishing, while others specialize in trade or popular publications, so who the publisher is can be a big clue about what kind of publication it is. You can google a publisher to discover their specialty.

Some academic publishers who publish a lot of work in linguistics written in English include:

- Cambridge University Press
- Elsevier
- John Benjamins

- Kluwer
- MIT Press
- Mouton de Gruyter
- Oxford University Press
- Springer
- Wiley Blackwell

The open access revolution in linguistics

Have you ever been frustrated by the paywalls on academic articles? You're not alone! Academic articles are researched, written, and edited by academics. These academics do not usually get paid by the journals for their work; instead, it is considered necessary service to their profession. Publishers may take care of copy-editing, formatting, printing, and web-hosting of the final product, but over time, more and more of that work has been automated or has been passed on to the unpaid academic editors. And yet, publishers continue to increase the subscription fees that university libraries have to pay in order to get access to the journal articles.

In 2015, the editors of the journal *Lingua* submitted a proposal to Elsevier, the for-profit publisher that operated the journal, to move *Lingua* to an open access model. They proposed that everyone should be able to read *Lingua* for free and that authors would retain copyright on their articles. Elsevier refused, so the entire editorial board of *Lingua* quit and started a new, open access journal called *Glossa*. Elsevier recruited a new editorial board and is still publishing *Lingua*, but it dropped from one of the top journals in linguistics to a marginal one. Meanwhile, *Glossa* has become very prestigious and popular, and is fully open access!

More recently, the editors of another journal, *Syntax,* are making a similar move. In 2024, Wiley Blackwell, the publisher of *Syntax*, eliminated the journal's paid editorial staff and assigned all the tasks to a generalized team who aren't trained in formatting linguistic data. As a result, the founding editor and several others have resigned and are in the process of founding a new open access journal named *Syntactic Theory and Research.*

The editorial board

Academic publishers will have editorial boards for their journals and for book series. It is the editorial board

who arranges for peer review and ensures that the published work meets the standards of the scholarly field (and not just general publication standards). Editorial boards of academic publications should be made up of established academics with expertise in relevant fields, and will typically be professors at universities. Just like you can google the authors of papers, you can google the editorial boards to determine whether they are reputable scholars.

"Publish or perish" and predatory journals

Academia is very competitive, and there is a lot of pressure for early career academics to publish papers in order to get a job and then get promoted to a tenured position—especially since there are far more qualified candidates than positions available. As a result, a number of "predatory journals" have sprung up to exploit graduate students and other inexperienced academics. These predatory journals will often send recruitment emails inviting people to submit articles, with many of the same properties as other phishing emails.

Some signs of a predatory journal include:

- A very broad topic area, or publication of research articles outside of their stated topic area
- A very short review process (e.g., less than three months)
- A cost to the author to publish
- No editorial board or an editorial board that does not consist of established scholars
- Few in-text references (e.g., less than 2-3 per paragraph)
- Short bibliographies that cite mostly general resources
- Low quality content (e.g., presenting opinion as fact, lack of engagement with opposing views, methods are not described in enough detail to be replicated)

Key takeaways

- An academic source typically is written by an expert, presents original research, is peer

reviewed, and includes extensive citation.

- The peer review process is when other experts on the topic read and critique a manuscript and give a recommendation to the editor for whether the manuscript should be published.
- To identify whether a source is academic, look for the following clues:
    - The author has a university affiliation and is writing within their research specialty.
    - The source is posted on a publisher's website, rather than a website where authors share material directly.
    - The publisher is an academic publisher or university press.
    - The editorial board consists of established academics with relevant expertise.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=3260#h5p-49

References and further resources

For a general audience

Dickson, EJ. 29 July 2014. How I accidentally started a Wikipedia hoax about Amelia Bedelia. *Daily Dot*. https://www.dailydot.com/unclick/amelia-bedelia-wikipedia-hoax/

Walkden, George. 9 March 2024. Editors of *Syntax* resign, found new journal. *Open Access Linguistics*. https://oaling.wordpress.com/2024/03/09/editors-of-syntax-resign-found-new-journal/

Wexler, Ellen. 5 November 2015. What a mass exodus at a linguistics journal means for scholarly publishing. *The Chronicle of Higher Education*. https://www.chronicle.com/article/what-a-mass-exodus-at-a-linguistics-journal-means-for-scholarly-publishing/

2.6. BECOMING A LINGUIST: SCIENTIFIC WRITING IN MORPHOSYNTAX

As we mentioned earlier in this chapter, morphosyntax is a science because it follows the scientific method. When we write about morphology and syntax, we use a genre of scientific writing. However, not all scientific fields follow the same conventions. In this section, we will learn about the genre of scientific writing used in morphology and syntax.

The structure of a morphosyntax paper

A lot of scientific fields use a formulaic structure known as the IMRaD structure, which stands for "Introduction, Methods, Results, and Discussion." Although all of those components show up in a morphosyntax paper, we rarely use that structure in morphosyntax. The IMRaD structure is used for papers that discuss formal experiments and analyze their results using statistics. Although morphologists and syntacticians sometimes publish papers about formal experiments in the IMRaD format, the vast majority of morphology and syntax papers follow a less formulaic structure. Papers should still begin with an introduction and end with a conclusion, but the body of the paper typically weaves together description of data, analysis of data, and discussion of the implications of the data throughout.

The introduction and conclusion

The introduction of a paper should typically be less than a fifth of the total length of the paper. In it, you should clearly state the main claim of your paper and summarize your main arguments. You should also contextualize your main claim, explaining why it is important and what theories or frameworks you are working with.

Because the structure of each morphology and syntax paper is unique, most papers end the introduction with a description of the structure of the paper. This is often called the overview, outline, or roadmap. Here's an example from Alhailawani (2023):

> This article is organized as follows. In section 2, I provide an overview of languages with P-drop [Preposition-drop]. I also discuss the two main approaches to P-drop, with representative examples. In section 3, I set out the JA [Jordanian Arabic] P-drop facts and note the similarities and differences between JA and other P-drop languages. In section 4, I discuss and refute an analysis of JA P-drop in terms of pseudo noun incorporation. Section 5 shows that a silent P analysis of P-drop fares better at capturing the JA P-drop data. In section 6, I argue for an analysis of P-drop in JA where PP and DP are present, showing that the analysis neatly captures the JA P-drop facts. Section 7 is a brief conclusion.

Your conclusion should repeat most of the same information as the introduction, except perhaps in a briefer format. You should re-state your main claim and re-summarize your main arguments. You should also return to the contextualization of your paper, restating why it is important and how your paper contributes to the field. Sometimes, a conclusion can also suggest areas for further study.

Here's the conclusion from the same paper about Jordanian Arabic (Alhailawani 2023):

> In this article, I provided a detailed description of P-drop in JA. I showed that the main properties of P-drop found in other P-drop languages hold in JA. I refuted an analysis of P-drop in terms of pseudo noun incorporation (Gehrke and Lekakou 2013, Hall 2019), showing that the prepositionless goal noun in JA does not behave like a pseudo-incorporated nominal. I showed that the noun in JA P-drop exhibits the typical properties of a regular referential DP. I applied some diagnostics of PP-hood to show that there is a silent PP structure involved in JA P-drop. Thus, I argued for an analysis of P-drop in which a PP structure is present but unpronounced (Myler 2013, Biggs 2014, Bailey 2018, among others). The silent P is licensed via incorporation of P into *v* (den Dikken 2010), which in turn explains the absence of P-drop with manner of motion verbs. Also, the goal DP is assigned case in situ via Agree with a higher *p* head in a split *p*/P structure (Svenonius 2007). P-drop is a relatively understudied phenomenon due to its restrictive nature. The arguments and analysis presented in this article contribute to the ongoing debate on the underlying mechanisms involved in P-drop.

Let's practice analyzing this writing so we can learn from it!

- Compare the conclusion from the outline from the introduction. Match the parts of the outline to the restatement in the conclusion. How is the wording different? Which one provides more detail? What other differences do you notice between the outline from the introduction and the conclusion?
- Where does the author contextualize their paper, showing how it contributes to the field?

Remember that you're not writing a mystery novel! It can be tempting to make a big reveal at the end with your final analysis, or describe all the false paths you went down when trying to complete your analysis, but doing so will typically be frustrating and confusing to your reader. Instead, present it simply: say what you're going to do in your paper, do it, and then say what you've done.

It may feel repetitive to write like that, but remember that you've (hopefully) thought about the subject of your paper for a while and are very familiar with it, but your reader is probably learning about it for the first time and does not know your thinking process. Good, clear writing often feels repetitive to the author. For example, notice how the example outline from the introduction and the conclusion from Alhailawani (2023) repeat many of the same points.

If you are reading someone else's paper, it is often a good idea to read the abstract, the introduction, and the conclusion of the paper first. This will help you decide if this is the right paper for you to read right now, and if so, which sections will be the most important or relevant for you. You should also usually be able to get an idea of the main point of the paper, which can help you to understand the paper as a whole.

The body

Before you start writing the body of a morphology or syntax paper, I would recommend making a list of the topics you are going to discuss in your paper and then deciding what order they should go in. Think about what your reader needs to already know in order to understand each part of your paper. Sometimes one argument builds on another argument, which determines which order they should go in.

The body of your paper should be divided into several sections with headers, especially if your paper is longer than about four pages. Each section should also be structured kind of like a whole paper: every section should have an introduction and conclusion that explains the main idea of the section and connects the content of the section back to the topic of the paper. This is recursive, so for even longer papers, your sections should be divided into subsections, and so on.

The headers for each section should be numbered and should have titles that clearly indicate what the section is about. Some papers number the introduction as section 1, but others number it as section 0 or give it no number at all.

Style in morphology and syntax papers

Sometimes when students begin to write university-level papers, they try to use complex vocabulary and sentence structure to look more sophisticated. But the best papers take complex ideas and explain them in simple language. You want to be as clear and concise as you are able. Do not use a more complex vocabulary when simpler terms will do.

> There seems to exist a belief that professional scientific writing *needs* to have a certain tone – impersonal, dense, abstract. I often see it in students who try to imitate that drab style [...] A lot of academics do write that way; nobody thinks that academics *ought* to write that way. In fact, editors and readers usually very much prefer if you *don't* write that way.
>
> -Rohrer 2024

One of the biggest obstacles to writing clearly and concisely in academic writing is that we are trying to explain intricate ideas. Academic writing often turns out complex because the ideas that the author is trying to convey are complex. On the other hand, it is very impressive if you manage to explain a complex idea in simple language, without simplifying the concept.

Another big obstacle to writing clearly and concisely in academic writing is that we are trying to be very precise. We use technical terms as appropriate because technical terms have precise definitions in our academic community. Our sentences may also become complex when we try to accurately convey exactly what we mean. When we aren't sure about our conclusions, or if our conclusions only apply in some cases, we try to be very clear about how sure we are and how our understanding is limited. For example, don't say *always* unless you're sure it actually happens always. Instead, write *usually* or *often* or maybe even just *sometimes*.

Keep in mind that many academics are actually not very good writers. If you find a paper that is hard to read, try to figure out if it's because the writing is not clear or because the ideas are difficult. There are some papers that you should not be imitating. Instead, what you should learn from those papers is what *not* to do.

Academics also tend to use abbreviations, as you no doubt observed in the two examples above from Alhailawani (2023). Try not to use too many abbreviations; it can become very overwhelming for the reader. But it is alright to use a few. However, when you introduce an abbreviation for the first time, make sure you

specify what the abbreviation means. This is usually done by writing out the term in full the first time, with the abbreviation afterwards in parentheses, as shown below (Alhailawani 2023):

> In **Jordanian Arabic (JA)**, directional prepositions like *ʕa* and *li* 'to' can be dropped with certain motion verbs like *yiruuħ* 'go' and *yewsal* 'arrive' ...

The reverse can also be done, with the portion that does not appear in the abbreviation in parentheses, as with the following example from Alhailawani (2023):

> In Greek and some varieties of British English, the complement of some motion verbs can optionally appear without an overt preposition or a determiner. [...] This phenomenon is known as **P(reposition)-drop**
>
> (Ioannidou and den Dikken 2009)

Word choice in scientific and linguistics writing

In scientific writing, we need to be very precise with our language. In this section, I will describe some of the words that we need to be especially careful with.

Proof, evidence, and fact

You should very rarely claim that you have proved something or that some piece of evidence is proof of something. The scientific definition of proof means you can show that it holds in all cases. There is no possible scenario where it could be wrong, not even a very unlikely scenario. Most often, *proof* is used for logical proofs, which is a formal way of demonstrating that certain base assumptions logically lead to particular conclusions.

When students use the words *proof* or *prove*, usually *evidence* or *provides evidence for* is more appropriate wording.

When scientists use the word *fact*, they mean something that has been observed and confirmed repeatedly,

to the point that we accept it as true; however, even when it has been observed hundreds or even thousands of times, it has still not been considered *proven*!

Theory, hypothesis, and model

As we went through the scientific method demonstration in Section 2.1, you may have noticed that we tried to stay skeptical about our conclusions. Because of this, and because we want to be very precise about how sure we are, some words are used differently in everyday speech than they are used in scientific discourse.

For example, in everyday usage, *theory* might mean a hunch, idea, or speculation, but in scientific writing, *theory* means a set of hypotheses about the cause of some event or phenomenon that have been scientifically tested. In some cases, a scientific theory might still be under debate, but in other cases, it might be generally accepted by the scientific community, although ultimately unprovable.

In everyday speech, a *hypothesis* might mean any guess, but in scientific writing, it means an educated guess or an informed prediction, based on observation of data.

When scientists are trying to describe how a complex system works, they might make a series of hypotheses known as a **model**. For example, Figure 1 shows a model of a carbon atom. This is not what a carbon atom actually looks like if you look at it through a very powerful microscope. Instead, it is a schematic representation of the parts of a carbon atom and their relationships to each other.

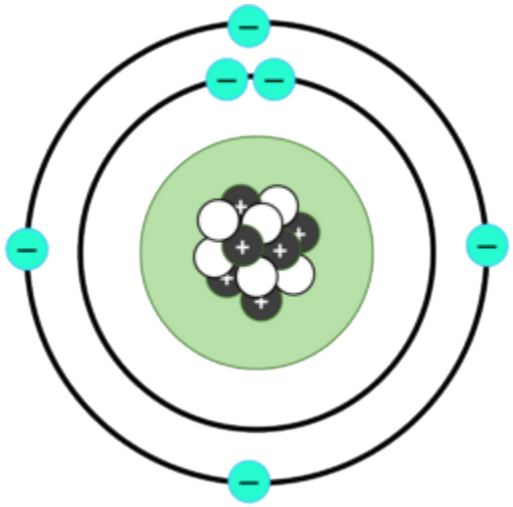


Figure 1: A model of a carbon atom.
Photo by SrKellyOP. Used under CC0 1.0 license.

In this textbook, we will create and test models of language known as *grammars*. They will also be schematic representations of the structure of language, rather than a "picture" of what is happening in our brains.

Significant

In everyday speech, *significant* just means important. However, in scientific writing, *significant* is short for *statistically significant*, which means that there is at least a 95 percent chance that, if you repeat the experiment, you will get results that are the same or more extreme results. In other words, there is a five percent chance or less that your results were a fluke due to random chance. This value is known as the p-value, and the 95% threshold is often described in decimals as $p\text{-}value \leq 0.05$ 'a p-value less than or equal to 0.05'.

To avoid confusion, it is best to reserve the word *significant* for its statistical meaning in scientific writing. Instead, you can use words like *crucial*, *important*, or *critical*.

Clearly, obvious, and without doubt

Sometimes writers will write that something is *clearly* the case, or that an argument is *obvious*, or that we can know something *without doubt*. Even experienced writers do this! Unfortunately, if something is obvious or clear to you, it does not mean it is obvious or clear to everyone. It can be very frustrating as a reader when an author says something is obvious instead of explaining it. If it is so clear and obvious, then it should be easy for you to take a few lines and explain it!

Writing anxiety

You do not need to feel guilty or ashamed if you are anxious about writing. Feeling anxious about writing is normal. Even famous authors get writer's block! The important thing is to not let your anxiety stop you from writing. I know that is easier said than done, but here are some tips to help you:

- Start early and break up the project into small pieces. Perhaps your goal for one day is just to find a research paper, and you will read it another day. Write an outline of your paper on a different day than you start writing.
- Make a more manageable goal. If it feels overwhelming to write 200 words, make the goal to write just 100 words that day. Some days, just writing one word might be a good goal. Often, once you get started, you might get going and write more than your goal, which feels really good! But if not, even a small amount of writing every day can give you lots of momentum.
- Don't worry about being perfect. Just try to start writing. You can come back and edit it later, but it is easier to edit something than nothing.

- Find a format or a genre that is less scary to you, like an email to a loved one or a list. Use that writing to start, and then edit it into formal writing later.
- If you're stuck, write about why you're stuck.
- Record yourself describing your paper out loud, then use speech-to-text to get a first draft.
- Meet up with friends or classmates to work independently alongside each other. Turn off your Wi-Fi and focus on just writing for an hour. Having someone else who is expecting you to show up at a specific time and write can help you avoid procrastination.

Key takeaways

- Experimental linguistics papers follow a formulaic Introduction, Methods, Results, and Discussion structure, but most papers in morphosyntax present the data and the analysis interwoven together throughout the body of the paper. Since the structure is not formulaic, most authors give an overview of the structure of the paper at the end of the introduction.
- In the introduction of your paper, you should clearly state your main claim, explain why your paper matters, and outline the evidence you will use to support your main claim. There should be no surprise reveals in the conclusion!
- Your conclusion should restate most of the information you gave in the introduction.
- The best papers explain complex ideas in the simplest language possible, without simplifying the ideas.
- Be careful about your word choice in scientific writing. For example, the word *proof* should be reserved for logical proofs, and the word *significant* should be reserved for statistical significance.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it

online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1088#h5p-22

References and further resources

For students

Batchelor, LaRoyce. 2024. *Growing a writing practice: Non-extractive writing.* Pressbooks. https://pressbooks.openedmb.ca/nonextractivewriting/

🔍📑 Macaulay, Monica. 2011. *Surviving linguistics: A guide for graduate students*, *2nd edition*. Somerville, MA: Cascadilla Press.

Rohrer, Julia. 2024. Writing about technical topics in an accessible manner. *The 100% CI.* https://www.the100.ci/2024/12/01/writing-about-technical-topics-in-an-accessible-manner/

Sources for examples

Alhailawani, Mohammad. 2023 The syntax of silent directional prepositions in Jordanian Arabic. *Canadian Journal of Linguistics*. 68 (2): 229–249.

2.7. BECOMING A LINGUIST: CITATION IN MORPHOLOGY AND SYNTAX

An important part of academic writing that is distinct from non-academic writing is the practice of **citation**. Citations are the formulaic ways that academic authors attribute the original sources of the information, examples, and quotations in their work.

Different fields use different citation styles. Some common ones you may have encountered in the past are APA (the American Psychological Association style) or MLA (the Modern Language Association style). Because linguistics is interdisciplinary, it doesn't really have a set style. You may find linguistics papers in either APA or MLA style, but most linguistics journals have their own style guide, which is usually similar to but more streamlined than APA. Because of this, most linguistics instructors aren't picky about which citation style students use in their papers as long as they include all of the necessary information in a clear and consistent format.

Why is citation important?

Citation is crucial to academic writing for many reasons.

First of all, academic writing should not be done in isolation. When you do research, you do not want to start from scratch. If you can show that your ideas are related to and build on other people's ideas, it gives your work more validity. In academic writing, you should be entering into a community of scholarship, contributing to the body of work, and combining new ideas and old ideas. Proper citation makes it clear which ideas you have taken from others to build upon and which are your own ideas.

Secondly, academic writing requires accountability to the community. As we are discovering and creating new knowledge, we are responsible to the community to be as accurate as possible. Scientific research, in particular, should be **replicable**, which means that others should be able to repeat our research methods and get similar results. Providing full documentation of our sources allows our colleagues to double-check our work and minimizes the chances that errors will spread.

Thirdly, not providing proper citation is dishonest. When we cite others' research, we give them credit for the work they've done and show respect for their knowledge and effort.

In-text citations

In-text citations are included in the main text of a paper and indicate where each individual piece of information and quotation comes from. This allows you to show the interaction between different people's ideas with each other and with your own ideas.

Some other fields use footnotes, but linguistics exclusively uses parenthetical citations. In-text citations indicate the author's surname (also known as *family name*) and the year of publication. Each in-text citation should have a corresponding entry in the bibliography or references section.

Below you will find some examples of in-text citations, marked in bold, from Culicover (2014). Don't worry too much about what the examples mean; pay attention to the structure of the citation!

The default way to include an in-text citation is to include the author's surname and the year of publication in brackets after the relevant piece of information. In the following example, the original source has two authors, Pearl and Sprouse.

> There are Bayesian approaches, which essentially argue that a structure can be judged unacceptable if there is an alternative structure that is significantly more likely, other things being equal **(Pearl and Sprouse 2013)**.
>
> Culicover (2014: 486)

The author's name is also frequently included in the structure of the sentence, in which case only the year is in parentheses.

> The distinction between grammaticality and acceptability was highlighted by **Miller and Chomsky (1963)**, who observed that a sentence can be well-formed in the sense that it follows the rules of linear ordering, phrase structure, and morphological form, but is nevertheless unacceptable.
>
> Culicover (2014: 466)

When should you incorporate the authors into your sentence and when should they be in parentheses? Most of the time, they should be included in parentheses. You want the source to fade into the background, and have the focus be on the information you're sharing. However, sometimes *who* said the information is important, for example if they were the first person to come up with a certain idea or if you're comparing their ideas with someone else's. In that case, it makes sense to incorporate the author into the sentence structure to give them more prominence.

You can also list multiple sources, as in the following example. In some style guides, they are separated by semi-colons (;), like in this example. In other style guides, they are separated by commas (,). A style guide might also instruct you to list them chronologically or alphabetically.

> The unacceptability of center embedding has been generally attributed to processing complexity and not to grammar **(Gibson 1998; Lewis 1997)**.
>
> Culicover (2014: 466)

In-text citations should make it clear which pieces of information comes from which source, as well as which pieces of information are original to you. In the following example, you can see that the pattern was first described by Postal, and then the pattern was explained in two different papers by Wasow.

> **Wasow (1972; 1979)** proposed traces as a way to solve a puzzling binding fact noted by **Postal (1971).**
>
> Culicover (2014: 475)

If a whole paragraph or chunk of your paper is from the same source, you don't need to repeat the in-text citation after every sentence, as long as it's clear that the whole section comes from the same source. Citations need to show up more frequently when you're incorporating information from many different sources. That being said, synthesizing ideas from many different sources shows an advanced understanding of the subject matter and critical thinking and is something that you should strive for in assignments involving research.

When giving a direct quote or a very specific piece of information, it is good to include the page number. The page number is indicated after the year, separated with a colon (:).

In the following example, there is a direct quote, as indicated by the quotation marks. The citation indicates that the quote is taken from page 161 of Bloomfield (1933). This example also illustrates that when you refer to the same source multiple times close together, you do not need to repeat the full in-text citation as long as it is clear which source you mean.

> This view has its direct antecedents in the theory of immediate constituents (IC), in which the function of syntax is to mediate between the observed form of a sentence and its meaning: **"we could not understand the form of a language if we merely reduced all the complex forms to their ultimate constituents" (Bloomfield 1933: 161)**. **Bloomfield** argues that in order to account for the meaning of a sentence, it is necessary to recognize how individual constituents (e.g., words and morphemes), constitute more complex forms, which themselves constitute more complex forms.
>
> Culicover (2014: 467)

This final example illustrates two conventions. First of all, if there are multiple sources that give the information you are citing, you could list just a few of them as examples (the abbreviation *e.g.* means 'for example,' taken from the Latin phrase *exempli gratia*).

Secondly, when there are more than two or three authors, instead of listing all of the authors, you should list just the first one and include the abbreviation *et al.* This comes from the Latin phrase *et alia* which means 'and others.' The full bibliography entry for Fillmore et al. (1988) shows that the paper has three authors: Charles Fillmore, Paul Kay, and Mary Catherine O'Connor.

> Such a view has been explicitly formalized in Construction Grammar **(Kay 2002)**, and has been widely adopted **(see, e.g., Fillmore et al. 1988; Kay and Fillmore 1999; Goldberg 1995; 2006; Culicover and Jackendoff 2005; Sag 2012)**.
>
> Culicover (2014: 485)

Common abbreviations in in-text citations

There are many abbreviations that may be used in in-text citations.

- *a.o.* for *among others*.
  Use: If you consulted many, many sources that say something similar, you can list just a few, with *a.o.* on the end.
  Example: Syntax textbooks in the generative tradition often use many tree diagrams (Adger 2003, Carnie 2020, Cowper 1992, Larson 2009, *a.o.*).
- *cf.* from the Latin *confer* 'compare'.
  Use: If you want to list several sources that are relevant to what you just wrote but say slightly different things (maybe you agree with them, maybe you don't, but they're worth looking at), you can introduce the citation with *c.f.*
  Example: There are many arguments for and against the theory of Universal Grammar (cf. Boxell 2016, Dąbrowska 2015).
- *e.g.* from the latin *exempli gratia* 'for example'.
  Use: If there are multiple sources that say the same or similar things, and you just want to list one or a few of them.
  Example: Syntax textbooks in the generative tradition often use many tree diagrams (e.g., Carnie 2020).
- *et al.* from the Latin *et alia* meaning 'and others'.
  Use: If your source has more than 2-3 authors, use *et al.* instead of listing them all. Some style guides will have specific instructions for how many authors to list and when to use *et al.*

Example: Charity Hudley et al. (2024) has advanced the discussion on justice in linguistics research.

- *n.d.* for 'no date'.
  Use: If your example does not have a date indicated. Be careful! If there is no date listed, then that is a clue that your source is not an academic source. On the other hand, sometimes my students cite something as *n.d.* when the date is available on the source, just tricky to find.
  Example: Most languages spoken in Europe belong to the Indo-European language family. One notable example is Basque, a language with no known living relatives (Etxepare Euskala Institutua, n.d.).
- *p.c.* from 'personal communication.'
  Use: If you are giving credit to someone who told you directly (not in a published format) the information you are citing, you should give them credit with *p.c.* For example, if a classmate gives you an idea that you use in your assignment, you could give them credit with a *p.c.* citation. This can be used when the information was told to you directly in person, over email, over text, etc. You can also use *p.c.* if a linguist shares data with you that they had collected but not yet published.
  Example: The language Niuean has verb-initial word order (Diane Massam, p.c.).
- *sic* from the Latin 'thus, so'.
  Use: If you're quoting something with an error in it, use *sic* to indicate that the error comes from the original text.
  Example: "There are several subsystems at work in [sic] when we use language" (Carnie 2021: 4).

The references list

The purpose of the bibliography or references list is to provide full citation details for every source in the paper, giving enough information for your reader to find your sources as quickly and efficiently as possible. The bibliography is usually done in a highly formulaic way so that it can be compact, not taking a lot of space, and so that, once you know the pattern, it is very quick and easy to navigate.

When making a bibliography entry, the first thing you have to do is identify what kind of source you are dealing with. Templates for the main kinds of sources are provided below. As you look at templates here and in style guides, pay attention to capitalization, italics, and punctuation.

A book

You use the book citation style if the entire book is written by the same author. If each chapter is written by a different author, you should use the style listed below for a book chapter in an edited volume.

The template for a book is as follows:

> Author's surname, Author's given name. Year. *Title of book in italics*. City of publication: Publishing company.

Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge, MA: MIT Press.

A book chapter in an edited volume

If a book has a collection of book chapters, each written by a separate author, you should cite the chapter you use, not the book as a whole. If you use more than one chapter, you should cite each of them separately.

The template for a book chapter in an edited volume is as follows:

> Author's surname, Author's given name. Year. Title of book chapter. In *Title of Whole Book in Italics,* ed. list of editors, page numbers of chapter. City of publication: Publishing company.

In the above example, the abbreviation "ed." stands for *edited by* and is used to introduce the editors.

Cardinaletti, Anna. 2004. Towards a cartography of subject positions. In *The structure of CP and IP*, ed. Luigi Rizzi, 124–174. Oxford: Oxford University Press.

Chomsky, Noam. 2008. On phases. In *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, ed. Robert Freidin, Carlos P. Otero, and Maria Luisa Zubizarreta, 133–166. Cambridge, MA: MIT Press.

An academic article

An academic article is a paper published in an academic journal.

The template for an academic article is as follows:

Author's surname, Author's given name. Year. Title of article. *Title of Journal in Italics*. Volume number (Issue number): Page numbers.

Not all journals use issue numbers. If you can provide it, it can be very helpful in looking up the article.

Barbosa, Pilar. 2011. *Pro*-drop and theories of *pro* in the Minimalist Program Part 1: Consistent null subject languages and the Pronominal-*Agr* Hypothesis. *Language and Linguistics Compass* 5 (8): 551–570.

Lillo-Martin, Diane. 1986. Two kinds of null arguments in American Sign Language. *Natural Language and Linguistic Theory* 4: 415–444.

A dissertation

A dissertation is a research project completed in order to get a PhD degree, typically the length of a book.

The template for a dissertation is as folllows:

Author's surname, Author's given name. Year. Title of dissertation. Doctoral dissertation, Name of Institution.

Jouitteau, Mélanie. 2005. La syntaxe comparée du breton [Comparative syntax of Breton]. Doctoral dissertation, Université de Nantes.

General patterns in bibliography entries

Here are some general formatting patterns worth noting for bibliographies in linguistics:

- In English, we typically put the given name first and the surname last. However, if we are alphabetizing by last name, we typically reverse the order and put in a comma to indicate that the order has been reversed.

My given name is Julianne. This is the name my parents assigned to me when I was born. My surname is Doner. This is the name that was passed on to me from my parents and that I share with my siblings and parents. I typically write my name as "Julianne Doner." But if I am alphabetizing by last name, I would list it as "Doner, Julianne" with a comma.

- Although many citation styles mandate only including the initial of the given name, when given the choice, many linguists opt for listing the full given name. There are several linguists with the same surname, so it is helpful!
- Unlike APA style, we do not usually put the year in parentheses.
- Titles of book chapters and articles are not in italics. Some style guides may put them in quotation marks, but we do not typically do so in linguistics.
- Titles of books and journals are in italics.
- Titles of one-time publications, including chapters, articles, and books, use sentence capitalization. This means only the first word of the title (and subtitle, if applicable) is capitalized, as well as proper names.
- Titles of repeating publications, like journals, use title capitalization. This means that all the major words of the title are capitalized.
- If the source is in a different language than the rest of the paper, it is useful to provide a translation of the source's title in square brackets, as shown for the Jouitteau example above.

Putting it all together

Here is how to format your bibliography list as a whole:

- Each entry should have a hanging indent. A hanging indent is when the first line is not indented, but the rest of the lines are indented.
- The bibliography entries should be listed alphabetically by author's surname.
- If there are multiple sources from the same author, those sources should be listed chronologically, as shown with the multiple sources by Chomsky.
- Very occasionally, you may find multiple sources from the same author *and* year. In that case, assign a letter of the alphabet (starting with 'a') to each one, which you put immediately after the year, as with the two papers by Barbosa in the example below. Your in-text citations should also include that letter.

Barbosa, Pilar. 2011a. *Pro*-drop and theories of *pro* in the Minimalist Program Part 1: Consistent null subject languages and the Pronominal-*Agr* Hypothesis. *Language and Linguistics Compass* 5 (8): 551–570.

Barbosa, Pilar. 2011b. *Pro*-drop and theories of *pro* in the Minimalist Program Part 2: Pronoun deletion analyses of null subjects and partial, discourse, and semi *pro*-drop. *Language and Linguistics Compass* 5 (8): 571–587.

Chomsky, Noam. 1982. *Some Concepts and Consequences of the Theory of Government and Binding.* Cambridge: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge: MIT Press.

Chomsky, Noam. 2008. On phases. In *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, ed. Robert Freidin, Carlos P. Otero, and Maria Luisa Zubizarreta, 133–166. Cambridge, MA: MIT Press.

Jouitteau, Mélanie. 2005. La syntaxe comparée du breton [Comparative syntax of Breton]. Doctoral dissertation, Université de Nantes.

Lillo-Martin, Diane. 1986. Two kinds of null arguments in American Sign Language. *Natural Language and Linguistic Theory* 4: 415–444.

Schuit, Joke, Anne Baker, and Roland Pfau. 2011. Inuit Sign Language: A contribution to sign language typology. *Linguistics in Amsterdam (ACLC Working Papers)* 4.

Key takeaways

- Citations are a formalized way to give credit to the sources of your work and should be used in all academic writing.
- Each piece of information that you get from another source should be followed by an in-text citation that lists the author's surname and the year of the source in parentheses.
- Academic papers should have a references section at the end of the paper that lists the full bibliographic information for all of its sources, detailed enough so that anyone else should be

able to find the same sources. Each in-text citation in your paper should have a corresponding entry in the references list.
- The field of linguistics doesn't have a single standardized citation format. Most linguistics professors don't care which format you use, as long as you are internally consistent and include all of the relevant information. In this section, we looked at the citation guidelines from the *Canadian Journal of Linguistics*.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=80#h5p-9

References and further resources

Reference materials

Canadian Journal of Linguistics/Revue canadienne de linguistique. Style guide/Protocole de rédaction. https://www-cambridge-org.uml.idm.oclc.org/core/journals/canadian-journal-of-linguistics-revue-canadienne-de-linguistique/style-guide-protocole-de-redaction

Unified style sheet for linguistics. Wayne State University. https://clas.wayne.edu/linguistics/resources/style

Sources for examples

Culicover, Peter. 2014. The history of syntax. In *The Routledge Handbook of Syntax,* ed. Andrew Carnie, Dan Siddiqi, and Yosuke Sato, 465–489. London: Routledge.

Dąbrowska, Ewa. 2015. What exactly is Universal Grammar, and has anyone seen it? *Frontiers in Psychology* 6.

CHAPTER 2 PRACTICE EXERCISES

Data analysis

Exercise 1. Prescriptive or descriptive?

Beginner [2.3]

Consider the following sentences. Some of them are prescriptively incorrect and some are descriptively ill-formed. Which are which?

a. I didn't say nothing.

b. Who should I pass the present to?

c. The cat black chased the mouse.

d. My daughter seems to happily go wherever we take her.

e. Is supper ready yet?

Exercise 2. Applying the scientific method

Intermediate-advanced [2.1, 2.2]

In Section 2.1, we used the scientific method to identify the difference between *we* and *us.* In this exercise, we will use the scientific method to identify the difference between *our* and *ours*.

a. Come up with some sentences that use the words *our* and *ours.*

b. Make an observation about your sentences. Do you notice any patterns?

c. Turn your observation into a hypothesis.

d. Check to see if your hypothesis is **falsifiable**. What would be the **necessary** and **sufficient** evidence to support your hypothesis? What is the opposite hypothesis? What would be the necessary and sufficient evidence to support the opposite hypothesis?

e. Come up with some sentences that test your hypothesis. Make sure you use both positive and **negative evidence**.

f. Do you need to revise your hypothesis? Why or why not?

Exercise 3. Mental vs. written lexicon

Beginner [2.4]

Which of the following pieces of information would you expect to be encoded in the mental lexicon? Which would typically be encoded in a dictionary?

a. The plural of *desk* is *desks*.

b. The words *guarantee* and *warranty* have the same historical origin.

c. The idiom *raining cats and dogs* means that it's raining very hard.

d. The abbreviation *tl;dr* means 'too long; didn't read' and is written at the beginning of long internet posts followed by a summary of the post.

e. My two-year-old niece calls *juice* /d͡ʒud͡ʒu/.

f. The plural of *child* is *children*.

Communication and study skills

Exercise 4. Scientific terminology

Beginner [2.6]

The following paragraph misuses scientific terminology. Rewrite the paragraph using terminology correctly.

Hint: If you're stuck, re-read the subsection on word choice in scientific writing from Section 2.6.

There is a significant distinction between the pronouns *we* and *us*. The following examples prove that they are not interchangeable.

(1) a. We are happy.
b. *Us are happy.
c. Joni saw us.
d. *Joni saw we.

From looking at the examples in (1), it's obvious that *we* and *us* are used differently.

Exercise 5. Personal language reflection

Intermediate [2.3]

Write a paragraph describing your language background. Some questions you may answer in your response include:

a. What language(s) do you speak?

b. What language(s) did your ancestors speak?

c. Do the language varieties you speak have prestige in your society?

d. How is your identity connected to your language(s)?

e. How did you and your family come to speak the languages and language varieties you speak?

f. How do prescriptivist attitudes affect you, your family, and your community?

Exercise 6. Citation practice

Beginner [2.6]

The following definition is taken from the 4th edition of *Syntax: A Generative Introduction* by Andrew Carnie on page 27.

> *Syntax* (as part of grammar): The level of linguistic organization that mediates between sounds and meaning, where words are organized into phrases and sentences.
>
> -Carnie 2021: 27

a. How would you write and cite a direct quote of this definition? Write it out using proper format for an in-text citation using parentheses.

b. How would you cite a paraphrase of this definition? Re-write the definition in your own words, using proper format for an in-text citation using parenthesis.

c. All of the information needed to write a bibliography entry for this textbook is listed below. Write a bibliography entry in the style of the *Canadian Journal of Linguistics*.

Title: Syntax: A Generative Introduction

Edition: 4th

Author: Andrew Carnie

Publisher: Wiley Blackwell

Year: 2021

Publication location: Malden, MA

Exercise 7. Evaluating sources

Intermediate [1.3, 2.5, 2.7]

Consider the following sources. For each source, determine (i) if it is an academic source, (ii) the primary sub-field(s) of linguistics relevant to the source, and (iii) write a citation for the source using *Canadian Journal of Linguistics* citation style.

a. Domains in Michif phonology by Nicole Rosen

b. Blackfoot language by Glenbow Museum

c. Modeling the noun morphology of Plains Cree by Conor Snoek, Dorothy Thunder, Kaidi Lõo, Antti Arppe, Jordan Lachler, Sjur Moshagen, and Trond Trosterund

d. The Mohawk language by Marianne Mithun

e. English-Cayuga/Cayuga-English dictionary by Frances Froman, Alfred Keye, Lottie Keye, and Carrie Dyck

Research and application

Exercise 8. Dictionary vs. lexicon

Intermediate [2.4]

In this exercise, we will compare a dictionary with your mental lexicon.

a. Think of a word you know that you expect to not be in the dictionary. Why not? Then look it up in a dictionary and see if it's there.

b. Find a word in a dictionary that is not in your mental lexicon. Why do you think is it not in your mental lexicon?

Language Journal

At the end of Chapter 1, we introduced the Language Journal. You should have picked a language to build a profile for as you work through this textbook. You can add to your profile with these questions.

Exercise 9. Prescriptivism

Advanced [2.3]

Research the prescriptivist attitudes of your language. Here are some questions to guide you. Not all of them will be applicable to all languages.

a. What are some of the different language varieties of your language? Is one more prestigious than the others?

b. How was your language affected by colonialism? Has it been a colonial language? Is it in contact with a colonial language?

c. Is there a language academy that determines what is "proper" use of the language?

Exercise 10. Pronouns

Beginner [2.1]

What pronouns are used in your language? Does your language use different pronouns in subject and object position?

3. THEORIES OF GRAMMAR AND LANGUAGE ACQUISITION

In this chapter, we will introduce the basics of a **generative grammar** model and learn about some of the arguments for and against Chomsky's theory of Universal Grammar.

Learning objectives

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

- Describe grammar as a model of the descriptive rules of language.
- Distinguish between I-language and E-language, between **competence** and **performance**, and between synchronic and diachronic approaches to explaining language properties.
- Describe the evidence that children acquire language as a rule-based system.
- Explain arguments for and against generative grammar and the theory of Universal Grammar.
- Contrast computer use of language with human use.
- Distinguish theoretical from empirical arguments and identify supporting evidence for different kinds of arguments.

3.1. WHAT IS A GRAMMAR?

If you look up the word *grammar* in a dictionary, you will find that this word has many different meanings and usages, three of which are particularly relevant to the study of morphology and syntax.

Grammar as a description of a particular language

One definition of *grammar* is a collection of rules that describe a particular language. Sometimes when people refer to the grammar of a language, they mean just the syntax rules, but other times, they mean all of the rules of the language, including phonological rules.

However, remember that when we are talking about the rules of the language, we mean **descriptive** rules, not **prescriptive** rules. Descriptive rules are not rules like a speed limit, where some authority tells you how fast you should drive and you can choose to ignore or follow the rule—and if you ignore it, you might have to pay a fine or even go to jail.



Figure 1. A speed limit sign. Photo by Shahnoor Habib Munmun. Used under CC BY 3.0 license.

Rather, these are rules like the law of gravity, which exists because scientists studied the properties of the world and wrote a law that describes what's happening. You don't choose whether or not to follow the law of gravity. If you do break it, and start floating away, you won't get in trouble! Instead, scientists will be very curious about what's going on!

Some examples of descriptive rules of English are listed in example (1).

(1)	a.	Nasalize vowels that precede a nasal consonant.	*phonological rule*
	b.	Mark plurals with the suffix *-s*.	*morphological rule*
	c.	Include a subject before the verb in a sentence.	*syntactic rule*



Figure 2. A bookshelf at a university library in the Indigenous languages section. This section contains reference and pedagogical grammars, as well as dictionaries, texts in Indigenous languages, and academic books on specific topics in Indigenous linguistics. What kinds of books can you spot in this picture? Photo by Julie Doner.

Pedagogical grammars are designed to help people learn a language in order to speak, write, or sign it. They often contain example dialogues, vocabulary, and practice exercises. They will also often explain grammatical patterns using non-technical vocabulary, ideally starting with the most useful constructions and vocabulary.

Lesson 3 - Ẹ̀kọ́ Kẹta:

Asking for Age

When you want to ask 'how old' you will ask: 'ọmọ ọdún mélòó?' For example, if you want to ask how old Tèmi is, you will ask: Ọmọ ọdún mélòo ni Tèmi?

And the response:

Ọmọ ọdún méjì ni Tèmi.
Tèmi is 2 years old.

Telling how old you are

Túndé fẹ́ mọ̀ nípa ọjọ́-orí àwọn ẹbí Yẹmí.
Túndé wants to know how old Yẹmí's family members are.

. Ìsọ̀rọ̀ngbèsì *(Dialogue)*

Yẹmí, Ọmọ ọdún mélòó ni ẹ́?
Ọmọ ọdún méjìdínlógún ni mí.
Ọmọ ọdún mélòó ni bàbáà rẹ?
Ọmọ ọdún mẹ́tàlélógójì ni wọ́n.
Ọmọ ọdún mélòó ni màmáà rẹ?
Ọmọ ogójì ọdún ni wọ́n.
Ọmọ ọdún mélòó ni àbúròo bàbáà rẹ?
Ọmọ ọgbọ̀n ọdún ni wọ́n.
Ọmọ ọdún mélòó ni ẹgbọ́nọ̀n rẹ?
Ọmọ ogún ọdún ni wọ́n.
Ọmọ ọdún mélòó ni àbúròò rẹ?
Ọmọ ọdún mẹ́rìndínlógún ni.



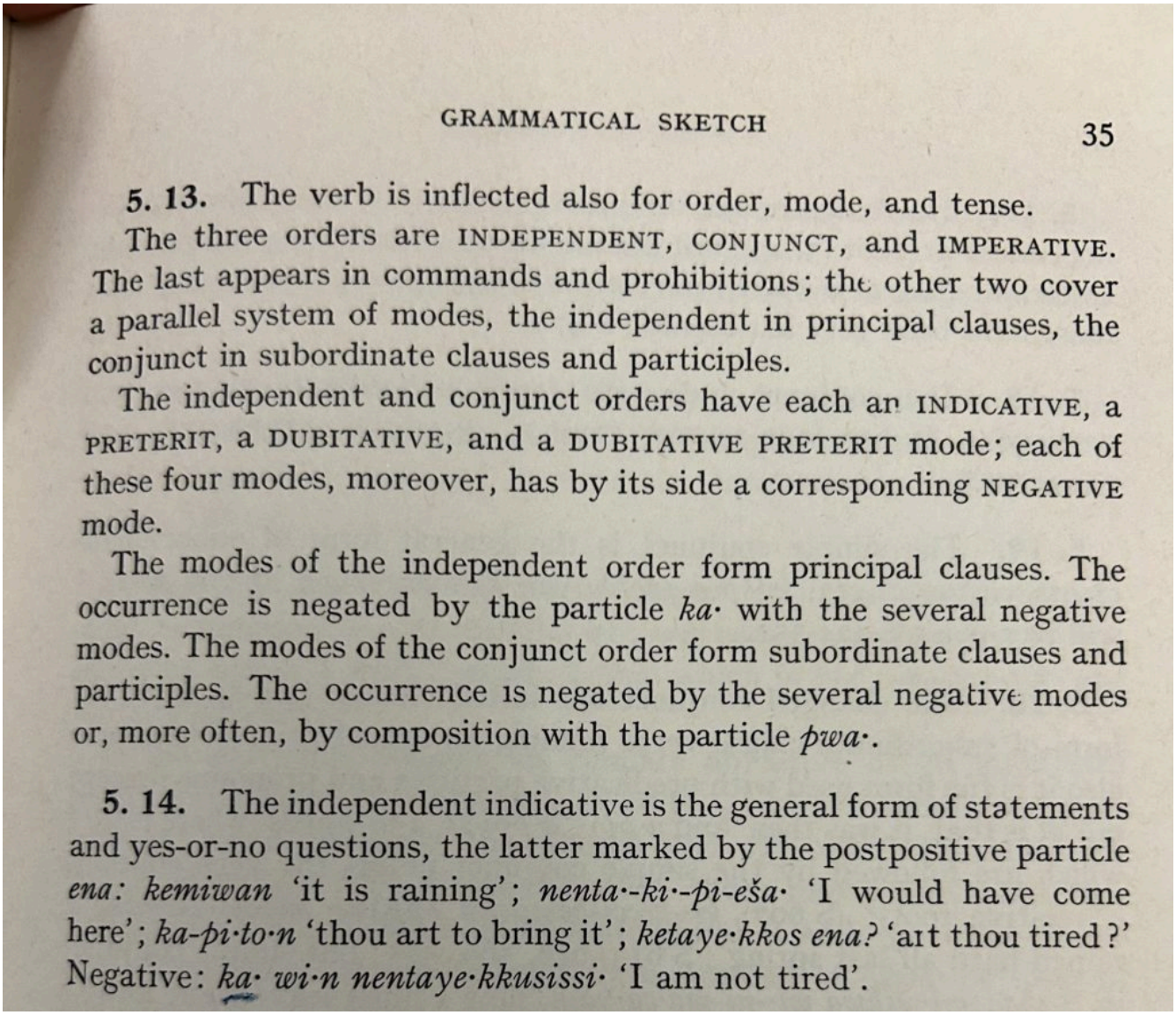
Túndé fẹ́ mọ̀ nípa ọjọ́-orí àwọn ẹbí Yẹmí.

Túndé wants to know how old Yẹmí's family members are.

 CC – 2011 The University of Texas at Austin

Figure 3: A page from a pedagogical grammar of Yoruba (Mosadomi 2011). Yoruba is a member of the Niger-Congo language family and is spoken primarily in Nigeria with over 43 million native speakers (Eberhard et al. 2023).

A reference grammar is designed for linguists or other scholars to look up the properties of a particular language. Reference grammars may be purely descriptive, simply describing the properties of the language in question, or they may have a theoretical approach, using the properties of a particular language to make claims about human language in general. Some reference grammars will only cover morphology and syntax, but others may cover phonology as well.



GRAMMATICAL SKETCH 35

5. 13. The verb is inflected also for order, mode, and tense.

The three orders are INDEPENDENT, CONJUNCT, and IMPERATIVE. The last appears in commands and prohibitions; the other two cover a parallel system of modes, the independent in principal clauses, the conjunct in subordinate clauses and participles.

The independent and conjunct orders have each an INDICATIVE, a PRETERIT, a DUBITATIVE, and a DUBITATIVE PRETERIT mode; each of these four modes, moreover, has by its side a corresponding NEGATIVE mode.

The modes of the independent order form principal clauses. The occurrence is negated by the particle *ka·* with the several negative modes. The modes of the conjunct order form subordinate clauses and participles. The occurrence is negated by the several negative modes or, more often, by composition with the particle *pwa·*.

5. 14. The independent indicative is the general form of statements and yes-or-no questions, the latter marked by the postpositive particle *ena: kemiwan* 'it is raining'; *nenta·-ki·-pi-eša·* 'I would have come here'; *ka-pi·to·n* 'thou art to bring it'; *ketaye·kkos ena?* 'art thou tired?' Negative: *ka· wi·n nentaye·kkusissi·* 'I am not tired'.

Figure 4: A picture of page 35 of Bloomfield (1957), a classic reference grammar of Nishnaabemwin (also known as Eastern Ojibwa). Nishnaabemwin belongs to the Algonquian language family and is spoken by about 220 people in Canada, according to the 2021 census. It is spoken in and around Lake Huron, especially Manitoulin Island (Eberhard et al. 2023). Photo by Julie Doner.

Grammar as a model of Language

In the study of syntax and morphology, the most important meaning of the word *grammar* is a scientific model of the entire system and structure of the human capacity for language. In other words, any theory of morphosyntax which accounts for how all of the components of language structure work together to result in the possible languages of the world could be considered a grammar.

A scientific **model** is a representation of a system, an organism, a process, or a phenomenon. In this case, what we are modelling is the human capacity for Language. This does not mean we are studying a particular language, like French, Anishinaabemowin, or American Sign Language, but the cognitive ability of humans to use language.

This model of Language has to account for the ways that diverse languages are similar, but also for the ways that languages are different. It needs to account for how children acquire language, both the process and the timing. It should also be compatible with all of the individual rules of particular human languages, since all of those languages are the result of the human capacity for Language.

A model like this consists of a collection of hypotheses. Each of these hypotheses needs to be tested and revised to help us develop the most accurate model of Language.

As we test and revise our models, we look to see what kind of data it can explain. Chomsky (1965) proposed three levels of adequacy:

1. An **observationally adequate** grammar accounts for all of the real-world data that we observe.
2. A **descriptively adequate** grammar not only accounts for the real-world data that we observe, but also speaker judgments about what is or is not ungrammatical, therefore including **negative evidence**.
3. An **explanatorily adequate** grammar accounts for all of the above, while also explaining how language is acquired.

We aim for an explanatorily adequate grammar, as that kind of model enables us to explain the most kinds of data.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1081#h5p-18

References and further resources

Reference materials

Bloomfield, Leonard. 1957. *Eastern Ojibwa: Grammatical sketch, texts and word list.* Ann Arbor: The University of Michigan Press.

Mosadomi, Fehintola. 2011. *Yorùbá Yé Mi: A Beginning Yorùbá Textbook*. Austin: University of Texas at Austin. https://coerll.utexas.edu/yemi

Academic sources

Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2023. *Ethnologue: Languages of the world*. Twenty-sixth edition. Dallas, TX: SIL International. http://www.ethnologue.com.uml.idm.oclc.org

3.2. GENERATIVE GRAMMAR

In this textbook, the model that we will be learning together belongs to the family of models called **Generative Grammar**. It was Noam Chomsky that came up with the idea that models of grammar should be generative. He defines generative grammar as "a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences" (Chomsky 1965: 8). In other words, a generative grammar uses rules to generate or "build" the structure of sentences. Santorini and Kroch (2007) define it as "an **algorithm** for specifying, or generating, all and only the grammatical sentences in a language."

What's an algorithm? It's simply any finite, explicit procedure for accomplishing some task, beginning in some initial state and terminating in a defined end state. Computer programs are algorithms, as well as recipes, knitting patterns, the instructions for assembling an Ikea bookcase, or a list of steps for balancing your checkbook.

An important point to keep in mind is that it is often difficult to construct an algorithm for even trivial tasks. A quick way to gain an appreciation for this is to describe how to tie a bow. Like speaking a language, tying a bow is a skill that most of us master around school age and that we perform more or less unconsciously thereafter. But describing (not demonstrating!) how to do it is not that easy, especially if we're not familiar with the technical terminology of knot-tying. In an analogous way, constructing a generative grammar of English is a completely different task than speaking the language, and much more difficult (or at least difficult in a different way)! Just like a cooking recipe, a generative grammar needs to specify the ingredients and procedures that are necessary for generating grammatical sentences.

Not all models of grammar use a generative framework. In other kinds of grammar models, language is produced by repeating memorized fragments or by probabilistic modeling, which is more similar to how Large Language Models produce language.

Noam Chomsky

Noam Chomsky (1928-) is perhaps the most well-known linguist in the world. Beginning with his 1955 dissertation *Transformational Analysis* and his 1957 book *Syntactic Structures*, Chomsky revolutionized the ways we think about language and linguistics and founded the modern field of linguistics.

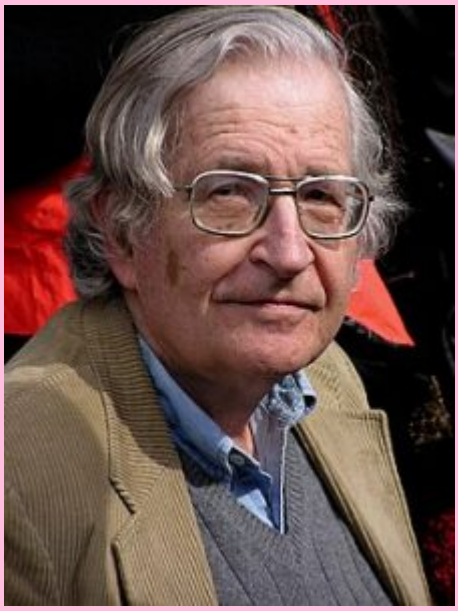


Figure 1: Noam Chomsky in 2004. Photo by Duncan Rawlinson. Used under CC BY-NC 2.0 license.

As we already learned in this section, Chomsky was the first to explicitly model language as a rule-governed system, launching the study of Generative Grammar. Another foundational proposal by Chomsky was the idea of **Universal Grammar**, which is the idea that humans are genetically endowed with the capacity for language. We will learn more about Universal Grammar in the remainder of this chapter.

Grammatical theory has changed a lot since 1955, but Chomsky has been a key player throughout this time. Most of the grammatical models that have been developed since then were either developed within a Chomskyan framework, sometimes by Chomsky himself, or in direct opposition to it.

Chomsky is also well-known for his political writings and activism.

At the time of this writing, Chomsky, although retired, is still publishing papers and giving talks. He is professor emeritus at the Massachusetts Institute of Technology and a laureate professor at the University of Arizona.

Some useful distinctions

When we are talking about our model of grammar, there are some useful distinctions we should make.

Competence vs. performance

Sometimes when we produce language, words don't come out exactly the way we intend. Because of this, we need a distinction between competence and performance. If you have linguistic **competence** in a language,

then you have acquired the grammatical rules necessary to produce the language in question. If you have linguistic **performance**, the language you produced conforms to the rules of the grammar you are using.

Most of the time, we have both competence and performance. But it is possible to have competence without performance. For example, if you are drunk or sleepy, you are more likely to misspeak. You still have the rules of grammar in your head, so you have competence. But you may have trouble accessing or implementing the grammar rules, so the end product of your language use does not conform to the rules of grammar in your head, so you do not have performance.

You can also have performance without competence. For example, say you memorize a sentence from a language you don't speak. You can repeat it, perhaps even flawlessly, so you have linguistic performance. However, you do not have the grammatical rules in your mind necessary to construct that sentence from scratch — you are only repeating what you have memorized — so you don't have competence in that language.

I-language vs. E-language

The next distinction we should make is between **I-language** and **E-language**. I-language stands for *internal language* and refers to the system of grammatical rules that an individual language user has in their mind. Everyone has a slightly different I-language. E-language, on the other hand, stands for *external language*, and refers to how language is externalized, including how it is used in a community. Since everyone has slightly different I-languages and because of the effects of linguistic performance, the E-language might not be consistent. In Generative Grammar, what we are trying to model is the properties of human I-languages. However, we cannot access I-language directly, since it is a cognitive object. Instead, we infer the properties of I-language from the properties of particular E-languages.

I-languages is also sometimes called Language (with a capital L) while e-language can be called language (with a small l).

Synchronic vs. diachronic

It is sometimes useful to look at how language changes over time, which is called the **diachronic** study of language. Although historical linguistics can be very interesting indeed, our model of grammar needs to be a model of language at a particular time, which is called the **synchronic** study of language. Often, this means studying modern languages, but it can also mean studying a historical language at a particular period.

The history of a language is not encoded in its grammar. Most speakers, unless they have specifically studied it, do not know the history of the languages they speak (and if they have studied it, they very likely did so *after* their language was acquired as young children). Because the history of the language is not part of what most speakers know, we cannot use a historical explanation in our model of grammar. The historical explanation can

be useful for explaining why the grammar has one set of rules and not another, but the rules themselves need to work as a system independent of where they came from.

Let's use riding a bicycle as an analogy. I could know the history of the bicycle, where each piece of metal was mined and smelted, and where the bike was assembled. All of those processes had to happen in order for the bike to exist for you to ride, but that knowledge isn't necessary to be able to ride the bike. What is necessary is that your bike is properly assembled, with the peddles linked to one of the wheels with a chain. When you peddle your bike, the peddles move the chain, which in turn rotates the wheel and moves the bike forward. This chain reaction between the parts of your bike is kind of like a grammar. The pedals, the chain, and the wheel are the different parts of a system that work together to make your bike work. It doesn't matter whether the chain on your bike is the original one or has been replaced, it just matters that it is working now.

In the same way, it is not necessary to know the history of your language in order to use the language. When we are trying to explain how our model of Language works, it doesn't matter if a particular rule was original to the language or borrowed from a different language — it is part of the system now, and we need to explain how it works now.



Figure 2: A bicycle. Photo by LUM3N. Used under Pixabay Content License.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it

online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1269#h5p-19

References and further resources

Attribution

Portions of this section are adapted from the following CC BY NC source:

Santorini, Beatrice, and Anthony Kroch. 2007. *The syntax of natural language: An online introduction.* https://www.ling.upenn.edu/~beatrice/syntax-textbook

For a general audience

Enos, Casey. No date. Noam Chomsky. *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/chomsky-philosophy

Academic sources

Chomsky, Noam. 1955. *Transformational Analysis*. Doctoral dissertation, University of Pennsylvania.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague/Paris: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

3.3. THE ACQUISITION OF MORPHOSYNTAX

As mentioned in Section 3.1, the best model of grammar should be **explanatorily adequate**, which means that it should be able to account for all the data we observe, all of the grammaticality judgments we collect, and how the rules of grammar are acquired. Because of this, although Language Acquisition is an entire field of study on its own, morphologists and syntacticians also care about the language acquisition process.

Learning vs. acquisition

You may have noticed that we have said a lot about language "**acquisition**" but not about language "learning." This is because, in linguistics, we distinguish between the two. When we say "acquisition," we mean the development of something (in this case, language) subconsciously, by instinct. When we say "learning," on the other hand, we mean that it developed consciously, through study and deliberate practice.

One piece of evidence that children acquire language, rather than learn it, is the course and timing of language development. Children typically acquire the majority of their grammar by the age of 4, before they go to school. This happens regardless of the language(s) they are exposed to. It also happens in different cultural contexts, in communities with vastly different traditions on how to interact with and speak to young children (see Figueroa in press and references therein). Children even begin learning the **prosody** of the language in their environment before they're even born (Gervain 2013)! Children will usually not even repeat utterances that have grammatical structures that they have not yet acquired—to the point that asking children to repeat an experimenter's utterance can be used in acquisition studies (for example, Kidd et al. 2006).

Compare that to the study of language in a classroom as a teenager or adult. In language classrooms, we often memorize and recite vocabulary and grammatical patterns. It takes effort and often makes your brain hurt! If you've ever studied a language as an adult or a teenager, think about how the process differed for you from acquiring a language as a child. If you're like me, there were times that you could recite the verb conjugations for a test, but when it came time to use them in a real conversation, you couldn't find the right form in your mind quick enough. This is what it feels like when you've *learned* the grammar consciously, but you haven't *acquired* it subconsciously.

One of the puzzles of language acquisition is the fact that children do not seem to receive negative evidence. The majority of the utterances produced by the adults in their environment are grammatical. If the adults misspeak and produce an ungrammatical utterance, there is often nothing to indicate that they misspoke. It is difficult to identify even indirect ways children might get negative evidence. Some studies show that parents comprehend children's ungrammatical utterances as frequently as their grammatical utterances, and

that parents rarely correct children's ungrammatical utterances, and instead respond to the content of their child's speech (Brown and Hanlon 1970, as cited in Marcus 1993). When parents do correct their children's grammar, it doesn't usually work very well. Some famous examples include the following:

> For experimental purposes, I have occasionally made an extensive effort to change the syntax of my two children through correction. One case was use by my two-and-a-half-year-old daughter of *other one* as a noun modifier. Over a period of a few weeks I repeatedly but fruitlessly tried to persuade her to substitute *other* + N for *other one* + N. With different nouns on different occasions, the interchanges went somewhat as follows:
>
> "Want other one spoon, Daddy."
> "You mean, you want THE OTHER SPOON."
> "Yes, I want other one spoon, please, Daddy."
> "Can you say 'the other spoon'?"
> "Other one spoon."
> "Say 'other'."
> "Other."
> "Spoon."
> "Spoon."
> "Other spoon."
> "Other. spoon. Now give me other one spoon?"
>
> Further tuition is ruled out by her protest, vigorously supported by my wife. Examples indicating a similar difficulty in using negative information will probably be available to any reader who has tried to correct the grammar of a two- or three-year-old child.
>
> (Braine 1971: 160-161, as cited in Marcus 1993)

> Child: Nobody don't like me.
> Mother: No. say "nobody likes me."
> Child: Nobody don't like me.

> [Eight repetitions of this dialogue follow.]
> Mother: No, now listen carefully, say "NOBODY LIKES ME."
> Child: Oh! Nobody don't likes me.
>
> (McNeill 1966: 69, as cited in Marcus 1993)

If we, as linguists, require negative evidence to figure out how grammar works, how can children manage without it?

Rule-based language learning

Let's consider a toy version of English that contains three-word sentences consisting of a noun, a verb, and another noun. The toy version contains sentences like (1) that are sensible given the real world as well as sentences like (2) that aren't, but that might be useful in fairy tale or science fiction contexts.

(1) a. Cats detest lemons.
b. Children eat tomatoes.
c. Cheetahs chase gazelles.

(2) a. Lemons detest cats. ("Secret life of citrus fruits")
b. Tomatoes eat children. ("Attack of the genetically modified tomatoes")
c. Gazelles chase cheetahs. ("Avenger gazelle")

Again for the sake of argument, let's assume a (small) vocabulary of 1,000 nouns and 100 verbs. This gives us a list of 1,000 x 100 x 1,000 (= 100 million) three-word sentences of the type in (1) and (2). Numbers of this magnitude are difficult to put in human perspective, so let's estimate how long it would take a child to learn all the sentences on the list. Again, for the sake of argument, let's assume that children can memorize sentences quickly, at a rate of one sentence a second. The entire list of three-word sentences could then be memorized in 100 million seconds, which comes to 3.17 years. So far, so good. However, the minute we start adding complexity to Toy English, the number of sentences and the time it would take to memorize them quickly mushrooms. For instance, adding only 10 adjectives to the child's vocabulary would cause the number of five-word sentences of the form in (3) to grow to 10 billion (100 million x 10 x 10).

(3) a. Black cats detest green peas.
b. Happy children eat ripe tomatoes.
c. Hungry cheetahs chase speedy gazelles.

Even at the quick rate of one sentence per second that we're assuming, the list of all such five-word sentences would take a bit over 317 years to learn. Clearly, this is an absurd consequence. For instance, how could our child ever come to know, as every English speaker plainly does, that the sentence in (4) is ungrammatical? If grammatical knowledge were based purely on rote memorization, the only way to determine this would be to compare (4) to all of the 10 billion five-word sentences and to find that it matches none of them.

(4) *Cats black detest peas green.

And even after performing the comparison, our fictitious language learner still wouldn't have the faintest clue as to why (4) is ungrammatical!

In addition to this thought experiment with its comically absurd consequences, there is another reason to think that language acquisition isn't entirely based on rote memorization – namely, that children use what they hear of language as raw material to construct linguistic rules. How do we know this? We know because children sometimes produce rule-based forms that they have never heard before.

One of the earliest demonstrations that children acquire linguistic rules, rather than simply imitating the forms of adult language, was the well-known **wug test** (Berko 1958). In it, the psycholinguist Jean Berko used invented words to examine (among other things) how children between the ages of 4 and 7 form plurals in English. She showed the children cards with simple line drawings of objects and animals and elicited plurals from them by reading them accompanying texts like (5).

(5) This is a wug. Now there is another one. There are two of them. There are two ___.

More than 75% of the children pluralized the invented words *cra, lun, tor,* and *wug* in exactly the same way that adults did in a control group: they added the sound *-z* to the word (Berko 1958: 159-162). Since none of the children had encountered the invented words before the experiment, their response clearly indicates that they had acquired a plural rule and were using it to produce the novel forms.

Children are also observed to produce novel rule-based forms instead of existing irregular adult forms (for instance, *comed* or *goed* instead of *came* or *went*). This process, which is known as **overregularization**, is further illustrated in (6) (Marcus et al. 1992: 148-149, based on Brown 1973).

(6)	a.	beated, blowed, catched, cutted, doed, drawed, drived, falled, feeled, growed, holded, maked, sleeped, standed, sticked, taked, teached, throwed, waked, winned (Adam, between the ages of 2 and 5)
	b.	drinked, seed, weared (Eve, between the ages of 1½ and 2)

Overregularized forms don't amount to a large fraction of the forms that children produce overall (less than 5% in the case of past tense forms, according to Marcus et al. 1992: 35), but they are important because they clearly show that even the acquisition of words can't be completely reduced to rote memorization.

Skinner vs. Chomsky and the debate on how children acquire language

B.F. Skinner was a psychologist who worked within a theory called behaviourism. In the theory of behaviourism, you can train people and animals to behave in certain ways by using positive and negative reinforcement (giving or taking away a desired stimulus) and positive and negative punishment (giving or taking away an undesired stimulus).

In 1957, Skinner published a book called *Verbal behaviour* in which he argued that children's language development could be explained by behaviourism. Basically, a babbling baby would one day, by coincidence, babble *dada* and the parents would be excited and encourage the baby to repeat those sounds. Over time, the baby would learn to associate *dada* with their father, and then to learn increasingly complex language structures.

Two years later, in 1959, Chomsky published a review of Skinner's book that eventually became even more famous than Skinner's original book! When Chomsky's review first came out, Chomsky was still relatively unknown, and Skinner mostly ignored it. He never formally refuted it, but continued to claim that Chomsky misunderstood his argument.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1280#h5p-20

References and further resources

Attribution

Portions of this section are adapted from the following CC BY NC source:

🔄 Santorini, Beatrice, and Anthony Kroch. 2007. *The syntax of natural language: An online introduction.* https://www.ling.upenn.edu/~beatrice/syntax-textbook

Comedy and satirical linguistics

🎉 Dresher, B. Elan. 2010. The Geese Rethink Innateness. T*oronto Working Papers in Linguistics* 32. https://twpl.library.utoronto.ca/index.php/twpl/article/view/13954

For a general audience

🧠 CrashCourse. 2020. Language Acquisition: Crash Course Linguistics #12. https://www.youtube.com/watch?v=Ccsf0yX7ECg

🧠 NOVA PBS Official. 2010. The secret life of scientists and engineers: Jean Berko Gleason and wugs. https://www.youtube.com/watch?v=ElabA5YICsA

Academic sources

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14: 150–177.

Brown, Roger. 1973. *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Braine, Martin D. 1971. On two types of models of the internalization of grammars. In Dan Slobin (ed.), *The ontogenesis of grammar.* New York: Academic Press. 153–186.

Brown, Roger & Camille Hanlon. 1970. Derivational complexity and order of acquisition in child speech. In John R. Hayes (ed.), *Cognition and the development of language.* New York: Wiley. 11–53.

Chomsky, Noam. 1959. Review of Verbal behaviour by B.F. Skinner. *Language* 35(1): 26–58.

Figueroa, Megan. (to appear). Language Development, Linguistic Input, and Linguistic Racism. *WIREs Cognitive Science*.

Gervain, Judit. 2018. The role of prenatal experience in language development. *Current Opinion in Behavioral Sciences* 21: 62–67.

Kidd, Evan, Elena Lieven, & Michael Tomasello. 2006. Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development* 21(2): 93–107.

Marcus, Gary, Steven Pinker, Michael Ullman, Michelle Hollander, T. John Rosen, and Fei Xu. 1992. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* 57(4). Chicago: University of Chicago Press.

Marcus, Gary F. 1993. Negative evidence in language acquisition. *Cognition* 46(1): 53–85.

McNeil, David. 1966. Developmental Psycholinguistics. In Frank Smith and George A. Miller (eds.), *The Genesis of Language: A Psycholinguistic Approach.* Cambridge: MIT Press. 15–84.

Skinner, B.F. 1957. *Verbal behaviour.* New York-London: Appleton-Century.

3.4. POVERTY OF THE STIMULUS AND UNIVERSAL GRAMMAR

Look at this sequence and guess what the next number in it should be:

(1) 1, 2, 3, 4, 5, ?

Most people would guess that it is '6,' assuming the formula in (2a). That is a very reasonable hypothesis, but the fact of the matter is, the next number in the sequence could also be '126,' if you use the formula in (2b).

(2) a. x = y

b. [(x-5) * (x-4) * (x-3) * (x-2) * (x-1)] + x = y

Without the 6th element in the sequence, there is no way to tell which formula is correct. Even if we have the 6th element, though, this problem doesn't go away. We could easily modify the formula in (2b) so that the 6th element is '6' but the 7th element — or any other term — diverges from the formula in (2b).

This illustrates the fact that an infinite rule-governed system is unlearnable. When a system is infinite, we can never be sure that the next element in the system will follow the pattern we think we have found.

The problem of infinite systems is relevant to language because language is **productive**, allowing us to combine the pieces of language in unlimited ways. This means that language is also an infinite system. Anytime you make a sentence, you can take that sentence and transform it into another sentence by adding "I believe that" or "Javier says that" or "Amita thinks that", etc..., to the beginning, as shown in (3).

(3) a. Language is infinite.

b. I believe that language is infinite.

c. Javier says that I believe that language is infinite.

d. Amita thinks that Javier says that I believe that language is infinite.

There is no limit to the amount of times you can do this, in principle, although you may eventually run out of breath or forget where you are in the sentence. But those are physical limitations. The grammar itself doesn't have any limits. In addition to that, using **embedded clause**s like we did in (3) is only one way we could add to

sentences infinitely. This kind of infinite building is in every part of the grammar, including in **possessive**s (4) or with **prepositional phrases** (5).[1]

(4) a. Zofia's idea
b. Zofia's brother's idea
c. Zofia's brother's friend's idea
d. Zofia's brother's friend's mother's idea

(5) a. the hole in the ground
b. the tree in the hole in the ground
c. the limb on the tree in the hole in the ground
d. the branch on the limb on the tree in the hole in the ground

This leaves us with a problem. Language is infinite, but infinite systems are unlearnable. Somehow, though, most people seem to learn language. This is known as the **poverty of the stimulus argument** since the language we receive in our environment (the stimulus) is insufficient for us to acquire language. In other words, it is impoverished.

Foundational assumptions: Grammar in the abstract vs. language in use

The argument that language is infinite because you can, in principle, always add more words to a sentence makes a pretty big assumption. The assumption is that there is an abstract notion of language or grammar independent of language in use. In other words, even if no one ever actually produces an infinitely long sentence, we are assuming that there is an abstract grammatical system that *can* produce an infinitely long sentence. The grammar has this possibility in its toolbox, even if no one uses it. This is similar to the distinction between **competence** and **performance** from Section 2.5.

Not all linguists would agree with this assumption. They might argue that, if language is restricted

1. This example is inspired by the children's song And the Green Grass Grew All Around.

by physical limitations, than these physical limitations are also part of grammar. In other words, there is no such thing as "abstract grammar" outside of language in use. Including physical limitations in your grammar does add some complications, though, as physical limitations vary. Grammar rules in such a system would have to be a bit fuzzy instead of **categorical**.

What do you think? Should our physical limitations be accounted for within grammar or outside of grammar? Are physical limitations a counter-argument against the infinity of language? What sort of evidence would convince you to change your mind?

Noam Chomsky's solution to this problem is the idea that some parts of language are innate. That is, some components of language are built into human DNA. In fact, he argued that there are three components to the human capacity for language: general cognitive capacities, experience, and a language-specific genetic component.

- **General cognitive capacity.** There are many parts of human cognition that can help us learn language but aren't specific to language. These are general cognitive capacities that we already have for other skills, but that we can also use to help us learn language. They include the ability to analyze data and spot patterns, developmental constraints, and principles of efficient computation.
- **Experience.** We know that experience plays a role in language acquisition. If you are surrounded by Japanese speakers as a child, you will learn Japanese. But if you are surrounded by Swahili speakers, you will learn Swahili. The language in your environment is called the input. You will learn a different language—both different grammatical rules and different vocabulary—depending on what language you're exposed to. This is where all the differences between languages come from.
- **Language-specific genetic component.** This genetic component is a capacity for language that we are born with, but unlike general cognitive capacities, it is specialized for language. Chomsky named it **Universal Grammar**, which is often abbreviated as UG. Universal Grammar should be uniform for all humans, since we have approximately the same genetics. It allows us to take our input and analyze it by instinct as an infant. The timing of language acquisition might also be constrained by genetics.

Besides the poverty of the stimulus argument, there are a few other pieces of evidence for Universal Grammar:

- Even children who are not exposed to a rich language environment tend to develop language of some kind. For example, deaf children in a hearing family who do not otherwise have access to a signed language community tend to develop a homesign system.
- There is evidence from acquisition that children need to be exposed to language within a certain timeline known as the Critical Period in order to be successful at acquiring language.

- There is a dissociation between language development and cognitive development. Some people (those with Specific Language Impairment) show deficient language skills but normal cognitive development, while others (those with Williams Syndrome) seem to show delayed cognitive development but normal language skills.
- Specific Language Impairment has even been traced to a particular gene called FOXP2!
- Certain parts of the brain are specialized for language use, such as Broca's and Wernicke's Areas.

In summary, according to the poverty of the stimulus argument, language is infinite, and although infinite systems should be unlearnable, we still manage to learn language. The solution to this is Universal Grammar, which is the hypothesis that there is an innate, genetically-endowed cognitive capacity that is specific to language.

On the other hand, the theory of Universal Grammar and the poverty of the stimulus argument are not accepted by all linguists. However, it is a foundational assumption in the framework of Generative Grammar, which is the approach taken in the textbook. In the next section, we will learn a bit more about the debate between those who adopt UG and who don't.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=118#h5p-13

References and further resources

Comedy and satirical linguistics

🎉 Wug Life. 2 Oct 2017. This Meowth can speak human language! Tumblr. https://wuglife.tumblr.com/post/165979527843/hahaha-oh-man-this-is-beautiful-a-quick

Academic sources

🔍 Boxell, Oliver. 2016. The place of Universal Grammar in the study of language and mind: A Response to Dąbrowska (2015). *Open Linguistics* 2: 352–372.

🔍 Dąbrowska, Ewa. 2015. What exactly is Universal Grammar, and has anyone seen it? *Frontiers in Psychology* 6: 852.

3.5. DEBATING UNIVERSAL GRAMMAR

Since Chomsky proposed **Universal Grammar** in the 1960s, it has split the field of linguistics in two—those who adopt his hypothesis and those who don't. Unfortunately, both sides of the debate are really bad at communicating and collaborating with each other. Both sides tend to argue against and make fun of (over)simplified versions of the other side.

For example, here is a pair of excerpts from an op-ed and a blog post in the wake of Vyvyan Evans' 2014 book *The Language Myth*.

> "I was, of course, expecting a negative reaction from Chomsky's disciples. But, given the practice in normal science, I was expecting a debate, and in particular, a challenge, with argument and evidence, to the substantive criticisms I levelled against the Chomskyan enterprise. This, it seemed to me, would have been a profitable way to take the field forwards. But things, at least, amongst some of Chomsky's die-hard followers, haven't panned out like that. In various fora, the reaction has sometimes appeared to border on the hysterical. For instance, in various public on-line venues I have been described as an 'idiot' and a 'quack', and *The Language Myth* has been repeatedly derided as 'junk' and 'junk science'; there has been a public debunking exercise led by a 'panel' of experts, embroidered with scorn and disdain, conducted in venues ranging from Facebook to Reddit. One prominent, and influential Chomskyan supporter, Professor Norbert Hornstein, has devoted, by now, an extraordinary number of posts to what amounts to a slim volume aimed at the general public, in his Faculty of Language Blogspot. He has repeatedly railed against *The Language Myth*, even issuing a call to (intellectual) arms: 'criticize this in all venues, especially where non-linguists gather. Consider it part of your linguistic public service' (December 2014).
>
> -Evans 2015 (Source)

Hornstein responds in the same vein:

> I was going to write something long and pointed in response to this new piece by Vyvyan Evans (VE). I was going to analyze the article noting its inflated combination of self-pity (those 'Chomsky disciples' are saying that my work is junk) and self-aggrandizement (I am the leader of a Kuhnian paradigm shift that will overthrow the Chomskyan orthodoxy that, prior to the heroic efforts of people like me, who are now being viciously pilloried by the intellectually enervated Generative establishment, has had a stranglehold on linguistic pedagogy). I was going to note that VE has yet to address a single criticism leveled against his work, preferring instead to personalize the disagreement in a two fold manner; first by noting his personal travails in fighting the Chomskyan dragon and second by citing the approval of various authorities and luminaries that think him and his work wonderful. However, despite the obvious temptations that VE's public demeanor make almost irresistible, his latest piece has convinced me that nothing much will be gained by doing this (fun as it may be).
>
> -Hornstein 2015 (Source)

Besides being rude, this kind of exchange is actually harmful to the advancement of science.

First of all, it is important to remember that just because someone adopts a hypothesis doesn't mean they think it is true. Often, the only way to test a hypothesis is to assume it is true, and see how far that will take you. Sometimes researchers will follow this line of research because they're pretty sure it is true, but sometimes researchers will adopt a hypothesis because they think it is wrong. In either case, they would be applying the hypothesis to different sets of data to try and find one where the hypothesis fails to predict the data. The **Scientific Method** is built on the assumption that your first hypothesis (and probably your second, and your third) will not be correct, but that each time we formulate a hypothesis and test it against new data, our hypothesis will become refined and hopefully closer to the truth.

Second of all, science is usually strongest when there are scientists working on the same problem from multiple perspectives. For example, if we assume that the Universal Grammar hypothesis is false, we might miss hard-to-spot patterns and similarities among the languages of the world. On the other hand, if we assume Universal Grammar is true, we might force languages into boxes where they don't quite fit. Because scientists are humans who generally become emotionally attached to their research, having skeptics look over your work from the opposite perspective helps to reduce errors.

Thirdly, by not taking each other seriously, we are missing out on half of the available linguistic research! All of our data is relevant to other theories, and even some of our theories can be reformulated to fit into other approaches. In fact, sometimes researchers from different approaches converge on the same solution, which provides a strong case that they are both on the right track. For example, the trend right now in **Chomskyan** syntax is to try to minimize the Universal Grammar component and explain as much as possible through

general cognitive capacities, in a framework known as **Minimalism**. It is possible that this line of research might lead to the conclusion that Universal Grammar can be attributed entirely to general cognitive capacities.

In sum, Universal Grammar is controversial and has resulted in a lot of contentious debate. However, the field of linguistics benefits from researchers from both sides working together and taking each other's research seriously.

Universal Grammar, recursion, and Pirahã

In the last section, we learned that Universal Grammar is the **language-specific genetic** component of the human language ability. In the last few decades, Chomsky begun arguing that, from an evolutionary perspective, it would make sense if this language-specific genetic component only consists of a single mutation that provided an immediate advantage to the human species.

In 2002, Chomsky published a paper with two evolutionary biologists named Marc Hauser[1] and Tecumseh Fitch, in which they argue that this single component is the capacity for **recursion**. Recursion is a property of grammar that allows us to iteratively apply a rule to its own output. This is what allows us to put possessives inside of other possessives, or prepositional phrases inside of other prepositional phrases, like the examples that illustrated the infinity of language from last section. We will learn more about recursion in Chapter 16.

A few years later, in 2005, a fieldworker named Daniel Everett published a paper claiming that the language he was working on did not have recursion. The language's name was Pirahã and it is spoken by a tribe in the Amazon rainforest that is still relatively isolated. Some researchers took Everett's claims as evidence that UG does not exist and that Chomsky is wrong. Many others, though, were not convinced.

Generative grammarians have criticized many elements of Everett's work. Some have looked through previous work on Pirahã by Everett himself and found data that they think contradicts his 2005 paper. Others have pointed out that Chomsky never claimed that all languages must have recursion, simply that recursion is a human capability. The Pirahã people are still able to learn Brazilian Portuguese, which definitely has recursion, which means that the Pirahã people are

1. In 2010, Marc Hauser was found guilty of falsifying data and scientific misconduct. However, the paper mentioned here is still cited and used, likely due to the prestige of the other two authors.

capable of recursion even if they don't use it. Similarly, others have pointed out that there are lots of different kinds of recursion, but not all languages use all of them. For example, German allows recursive noun-noun **compound**s but other languages do not (see Chapter 10). Still others have pointed out that Everett's work is problematic because it primitivizes and exocticizes the Pirahã people and their culture.

Everett's work and the Pirahã people still get attention in popular linguistics books and blogs and on mainstream media. The impact on research, on the other hand, was not so large. Because the Pirahã people are so isolated, it is difficult to independently test Everett's claims. After a few years debating the merits and weaknesses of Everett's work, many linguists have moved on to other things.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://pressbooks.openedmb.ca/wordandsentencestructures/?p=133#h5p-14

References and further resources

For linguistics students

Hornstein, Norbert. 2015. My (HOPEFULLY) last ever post on Vyvyan Evans and his endless dodging of the central issues. *Faculty of Language*. http://facultyoflanguage.blogspot.com/2015/05/my-hopefully-last-ever-post-on-vyvyan.html

Evans, Vyvyan. 2015. The structure of scientific revolutions: Reflections on radical fundamentalism in language science. *Psychology Today*. https://www.psychologytoday.com/ca/blog/language-in-the-mind/201504/the-structure-scientific-revolutions

Academic sources

Hauser, Marc, Noam Chomsky, and Tecumseh Fitch. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298* (5598), 1569–1579.

Everett, Daniel. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46(4): 621–646.

Nevins, Andrew, David Pesetsky, & Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* 85(2): 355–404.

3.6. COMPUTER USE OF LANGUAGE

Large language models (LLMs), one type of a broad category often called "artificial intelligence," collect a large amount of data and use probabilistic and statistical methods to process language that they haven't seen before and respond to it. In the last few years, LLMs have received a lot of attention because they seem to interpret and use language in a human-like way, at first glance. However, if you look closely, these programs have some significant differences compared to the human use of language.

How LLMs learn language

Humans learn language through live interaction with other humans, in context. When infants as young as 10 months old are learning language, they pay attention to where the adults who are speaking or signing are looking, and use that information to help them figure out what words mean (Brooks and Meltzoff 2005).

In contrast, LLMs typically learn language through the statistical analysis of large collections of text, called **corpora**. Text corpora have several important differences compared to live interactive speech and sign. For example, text typically indicates where words and sentences begin and end with spaces and periods, unlike speech. Many recent LLM programs are trained on very large corpora, such as large portions of the Internet. The Internet is several orders of magnitude larger than the total amount of language a human hears in their lifetime, let alone how much language input a child receives in the first four years of life, when they acquire the majority of their grammar. We can tell that LLMs do not learn language in the same way as humans because of the differences in the size and type of input.

But perhaps the biggest difference in the ways LLMs use language is that they do not have the shared context and real-world knowledge that we have. The language they produce can be very convincing, but there is evidence that LLMs do not understand the language they are producing, if you know where to look. Consider exchange 1, in which ChatGPT tells you that someone wearing a blindfold is not easy to see. Even a young child can tell you that this is false. The blindfold obstructs Mary from seeing, not from being seen.

Exchange 1

Prompt: Mary and Sue are talking and Mary is blindfolded. Is Mary easy to see?

Response: Mary being blindfolded implies that she cannot see, regardless of how easy or difficult she might be to see physically. The blindfold prevents her from seeing anything. So, from a literal perspective, Mary is not easy to see because she is blindfolded.

ChatGPT 3.5, May 24, 2024
Prompt by Hadas Kotek

When humans communicate, we try to recreate our interlocutor's perspective and anticipate their train of thought. We try to determine *why* they are communicating in the way they chose, and respond sometimes to the purpose of the communication rather than the literal meaning. An LLM does not have access to any of the real-world context or empathy needed for this kind of communication. When we, as users, access information through an LLM, we lose the relationship aspect of information seeking. The information we get from an LLM becomes separated from its source, which makes it more difficult to evaluate its reliability. This problem is amplified to the level of absurdity in situations where LLMs reproduce satirical sources as fact. For example, Google AI overview has suggested eating a rock a day based on an article by satirical news source *The Onion* and that adding non-toxic glue to pizza can keep the cheese from sliding off based on a Reddit comment which was presumably meant as a joke (Bender 2024).

How LLMs process and produce language

As we already saw in Section 3.3, there is evidence that children produce language using a rule-based approach. When humans use language, they usually understand it and think about its meaning. LLM programs, on the other hand, focus on key words and use statistical analysis. The way they produce language can be called a stochastic parrot, that is, "a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning" (Bender et al. 2021: 617). In other words, it considers what words are in the input it receives and then calculates what words are likely to appear as an output.

The probabilistic model of language used by LLMs results in some key differences compared to human language. First of all, LLMs are not good at extending patterns to novel forms. For example, in exchange 2, ChatGPT fails a **wug test** assessing how to pronounce the plural of the novel word *mee*. Since *mee* ends in a vowel, the plural form should be pronounced with a /z/ sound, rhyming with *freeze* not *lease*. Recall that even young children are able to correctly predict the plural form of unfamiliar words.

Exchange 2

Prompt: If "mee" is the name of a new type of toy, and I have two of them, I have two mees. Does the plural word "mees" rhyme with "lease" or "freeze"?

Response: The plural word "mees" would typically rhyme with "lease." It follows the standard English pluralization rule where adding an "s" to the end of a noun typically results in a pronunciation that rhymes with "lease." So, "mees" would rhyme with "lees" rather than "freeze."

ChatGPT 3.5, May 24, 2024
Prompt by Nathan Sanders

Additionally, LLMs are not able to distinguish between rare and ungrammatical constructions, unlike human children (Nathan Sanders, p.c.). In Section 3.4, we learned that, in human language learning, we are able to use a relatively small amount of data to learn language, even though language is infinite, which is called the **poverty of the stimulus**. What this means is that we can determine whether a sentence is ungrammatical even if we've never encountered the sentence before. LLMs, on the other hand, are not good at identifying ungrammaticality.

The trouble LLMs have in identifying ungrammatical utterances is illustrated through exchange 3. The prompt asks ChatGPT to evaluate two *wh*-questions, shown in (1) and (2) below.

(1) What will you sleep after the cat eats some?

(2) You will fall asleep after the cat eats some what?

Wh-questions are normally formed by moving a *wh*-word such as *what* to the beginning of a sentence. However, there are constructions in which *wh*-movement is blocked, called *wh*-islands, including the **embedded clause** *after the cat eats some tuna.* This means that the *wh*-word has to stay in place, as in sentence (2). It is ungrammatical to move it, as in sentence (1). However, ChatGPT fails to recognize that sentence (1) is ungrammatical.

Exchange 3

Prompt: I said to Alice: "I will sleep after the cat eats some tuna". However, Alice didn't hear the last word I said.
Which of the following would be valid ways for Alice to respond?

(1) What will you sleep after the cat eats some?

(2) You will fall asleep after the cat eats some what?

Response: Both responses are valid ways for Alice to seek clarification:

(1) "What will you sleep after the cat eats some?" – Alice is asking for clarification on what you'll sleep after the cat eats, as she didn't hear the last word.

(2) "You will sleep after the cat eats some what?" – Alice is asking for clarification on what specifically the cat will eat before you sleep.

In both cases, Alice is attempting to fill in the missing information from what she heard you say.

ChatGPT 3.5, May 24, 2024
Prompt by Nathan Sanders

Note the interpretation that ChatGPT gives for sentence (1): "Alice is asking for clarification on what you'll sleep after the cat eats." This interpretation would arise if the *wh*-word was moved from the object position of the main clause, rather than from the embedded clause. This movement is not blocked by being a *wh*-island, however, it is still ungrammatical because the main clause verb *sleeps* is **intransitive** and does not take an object.

I followed up with ChatGPT to see if I could get it to also provide the interpretation where the *wh*-word comes from the embedded clause, describing what the cat has eaten, which would violate the *wh*-island. I did this by asking if sentence (1) was ambiguous. ChatGPT provided two interpretations. The first interpretation provided is that the speaker is asking what activity or event you will engage in after the cat eats some food. This is similar to the interpretation provided for sentence (1) in exchange 3, although it removes the reference to sleeping. The second interpretation provided in exchange 4, on the other hand, is that the speaker is asking what the cat will eat. This second interpretation has the *wh*-word moving from object position of *eat* inside the embedded clause. In other words, interpretation 2 violates the *wh*-island.

Exchange 4

Prompt: Is the sentence "What will you sleep after the cat eats some?" ambiguous?

Response: Yes, the sentence "What will you sleep after the cat eats some?" is ambiguous because it could be interpreted in two different ways:

(1) "What will you sleep after the cat eats some?" – This interpretation suggests that the speaker is asking what activity or event you will engage in after the cat eats some food.

(2) "What will you sleep after the cat eats some?" – This interpretation suggests that the speaker is asking what specific thing the cat will eat before you sleep.

The ambiguity arises from the lack of clarity about whether "some" refers to an action (e.g., eating) or a noun (e.g., food).

ChatGPT 3.5, May 24, 2024

Remember, LLMs are trained on **corpus** data, which means they do not encounter **negative evidence**. This means that LLM training data would not include sentences like (1) and so it is not surprising that LLMs don't know how to handle them. However, children *also* do not have access to negative evidence, and children are able to distinguish between rare and ungrammatical utterances.

When problems like these become known, the LLM developers often work on adjusting the design of the LLM, so it is possible that these issues may no longer arise in future iterations of ChatGPT. Even so, LLMs need specialized programming to address such issues, whereas children figure it out on their own.

Some dangers of LLMs: Hallucinations and bias

LLMs are programmed to create plausible-sounding text, but not necessarily to be accurate, truthful, or even helpful. It is not uncommon for LLMs to confidently assert false information or to make things up. When an LLM makes up false information, it is called a hallucination. Sometimes LLM hallucinations are partially true, which makes them harder to spot. For example, an LLM might create a reference list using real authors' and journals' names, but the articles themselves are not

real. However, because LLMs create output that sounds like a human, it is easy sometimes to forget that it was not created by a human and should not be trusted like a human. Furthermore, the LLM will not indicate its sources, which makes it even trickier to fact-check.

We don't always forget that LLMs are not human. Sometimes we remember that we are interacting with a computer, and so we treat their output as objective and universal. But this is also not true. LLMs are each trained on a particular corpus, and if the corpus has any biases, the LLM will reproduce that bias. In fact, because LLMs are statistical, it may even amplify the bias. If you play with an LLM long enough, it is easy to find these biases. For example, they may assume people's gender based on their occupation and text-to-image programs will often default to white people.

At the time of this writing, nearly all LLMs are trained primarily on English data. This is another major sources of bias.

LLMs are very interesting tools with lots of potential applications—but if we are to use them responsibly, we need to keep their hallucinations and their biases in mind.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1438#h5p-24

References and further resources

For a general audience

🔍 Bender, Emily. 28 May 2024. Information is relational. *Mystery AI Hype Theater 3000: The Newsletter*. https://buttondown.email/maiht3k/archive/information-is-relational/.

🔍 Bender, Emily. 14 June 2022. Human-like programs abuse our empathy — even Google engineers aren't

immnue. *The Guardian*. https://www.theguardian.com/commentisfree/2022/jun/14/human-like-programs-abuse-our-empathy-even-google-engineers-arent-immune

Bender, Emily and Chirag Shah. 13 Dec 2022. All-knowing machines are a fantasy. *Institute of Art and Ideas News.* https://iai.tv/articles/all-knowing-machines-are-a-fantasy-auid-2334

🔍 Biever, Celeste. 25 July 2023. ChatGPT broke the Turing test — the race is on for new ways to assess AI. *Nature* 619: 686-689. https://www.nature.com/articles/d41586-023-02361-7

Chomsky, Noam, Ian Roberts, and Jeffrey Watumull. 8 Mar 2023. Noam Chomsky: The false promise of ChatGPT. *The New York Times*. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

🔍 Dede, Chris. 6 Aug 2023. What is Academic Integrity in the era of generative Artificial Intelligence? *Silver Lining for Learning*. https://silverliningforlearning.org/what-is-academic-integrity-in-the-era-of-generative-artificial-intelligence

🔍 Epstein, Robert. 18 May 2016. The empty brain. *Aeon*. https://aeon.co/essays/your-brain-does-not-process-information-and-it-is-not-a-computer

Kotek, Hadas. 6 Nov 2023. Text-to-image models are shallow in more ways than one. Personal blog. https://hkotek.com/blog/text-to-image-models-shallow-parsing

🔍📗 Millière, Raphaël and Charles Rathkopf. 23 Nov 2023. Why it's important to remember that AI isn't human. *Vox*. https://www.vox.com/future-perfect/23971093/artificial-intelligence-chatgpt-language-mind-understanding

🔍 Nkonde, Mutale. 22 Feb 2023. ChatGPT: New technology, same old misogynour. *Ms. Magazine.* https://msmagazine.com/2023/02/22/chatgpt-technology-black-women-history-fact-check

🔍 O'Brien, Matt. 1 Aug 2023. Chatbots sometimes make things up. Is AI's hallucination problem fixable? *Associated Press.* https://apnews.com/article/artificial-intelligence-hallucination-chatbots-chatgpt-falsehoods-ac4672c5b06e6f91050aa46ee731bcf4

O'Neil, Cathy. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Books.

Academic sources

Bender, Emily, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Smitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.

Bender, Emily, and Alexander Koller. 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th annual meeting of the association for computational linguistics*. 5185–5198.

Brooks, Rechele and Andrew Meltzoff. 2005. The development of gaze following and its relation to language. *Developmental Science* 8 (6): 535–543.

Dingemanse, Mark. 2024. Generative AI and research integrity. Manuscript, Radboud University Nihmegen. https://osf.io/preprints/osf/2c48n

Hicks, Michael Townsen, James Humphries, and Joe Slater. 2024. ChatGPT is bullshit. *Ethics and Information Technology* 26 : 38. https://link.springer.com/article/10.1007/s10676-024-09775-5

Kotek, Hadas, Rikker Dockum, and David Sun. 2023. Gender bias and stereotypes in Large Language Models. *Proceedings of The ACM Collective Intelligence Conference.* 12–24.

3.7. BECOMING A LINGUIST: EMPIRICAL AND THEORETICAL ARGUMENTS

There are two main kinds of arguments, empirical arguments and theoretical arguments. Empirical arguments are based primarily on observation of data, while theoretical arguments are based primarily on ideas. In this class, we will be using both. Often, more complex arguments combine both theoretical and empirical arguments.

Empirical arguments

Empirical arguments are based on data. They have three parts: make relevant observations, make a theoretical claim, and explain how your observations are related to your claim. The parts don't have to be in this order, but they should be in an order that makes sense.

In this chapter, the fact that children's knowledge of language is rule-based (as discussed in Section 3.3) is an example of an empirical argument. We observed examples of children's behaviour, such as their response to corrections from their parents, their pluralization of **nonce words**, and their use of **overregularization**. Based on our observations, we concluded that children's language is rule-based rather than memorized.

Making observations

When you make an empirical argument, you need to start by describing the surface properties of your data. You should be answering the question, "What are the non-controversial facts?" These are mostly observations that a non-linguist could make!

Some examples of observations that you might start which are relevant to morphology and syntax include:

- What does the sentence mean?
- Which nouns and pronouns in the sentence refer to the same entity?
- Is the sentence ambiguous? In other words, does it have more than one meaning?
- What is the word order? Which words precede or follow which other words?
- Are other word orders possible?
- Is the sentence grammatical?
- What similarities are there between this word and other words?

There are lots of observations that you could make about any sentence, but you need to stick to the ones that are relevant to your theoretical claims.

Making a theoretical claim

Although an empirical argument is based on some observation, it leads you to make some theoretical claim about the structure of your data. In this part, you should be answering the question, "What conclusions can you draw based on your observations?" This is where you talk about things you can't see or hear, but that you infer based on what you do see or hear.

Some examples of theoretical claims you might make in this course include:

- Whether a word is **simplex** or **complex**.
- The classification of a word or sentence based on its structure.
- The part of speech of a word (e.g., whether a word is a noun or a verb), as we will learn about in Chapter 5.
- Whether a string of words behaves as a unit, which we will learn about in Chapter 16.

Backing up your claim

To make a complete argument, you also need to explain how your your observations provide evidence for your claim. You cannot assume that your readers will be able to draw the same conclusions as you!

In some cases, you should also think about other claims people might think to make about the same data, and why your analysis is better.

Theoretical arguments

Theoretical arguments are based on ideas, logic, and reasoning, instead of data. Because of this, we can never trust them as much as empirical arguments. However, we wouldn't be able to get as far in science without theoretical arguments, too!

In this chapter, the Poverty of the Stimulus argument in Section 3.4 is an example of a theoretical argument.

A theoretical argument has two parts: premises and a conclusion.

Laying out your premises

A theoretical argument is based on premises, which are the ideas that you use as a foundation to draw your

conclusion. They are your claims about how things work. A premise can be a conclusion from a previous (theoretical or empirical) argument, or it can be an easily observable fact (e.g., language is infinite; humans acquire language; not all languages are the same).

For your argument to work, you need to be relatively certain that your audience will accept your premises.

The Poverty of the Stimulus argument has two premises: that language is infinite and that infinite systems are unlearnable.

One of the ways to argue against a theoretical argument is to claim that one (or more) of the premises is false.

Drawing a conclusion

Your premises should logically lead to a conclusion. It's important that you actually state what that conclusion is, though, instead of making your readers figure it out.

Another way to argue against a theoretical argument is to claim that the conclusion does not actually follow from the assumptions.

Consider your assumptions

Every time you make an argument, you make theoretical assumptions. These can be foundational, like "I am assuming generative grammar is correct," or structural, like "I'm assuming that two things can't occupy the same spot," or methodological, like "I'm assuming that grammaticality judgments are reliable." A lot of times, we are making assumptions without realizing it, so it's always good to spend some time thinking about what assumptions we are making.

Decide whether you *should* be making your assumptions, or whether you should dig deeper and investigate some of them. Sometimes it's useful to make assumptions, though, to see where a theory will lead us, or because if we had to start from scratch with no assumption every time, we wouldn't make it very far very fast.

Decide whether you need to state your assumptions. You don't need to state any assumptions that you are pretty sure your audience shares. This happens a lot in a class setting—we share the assumptions of the class materials. You also don't need to state any assumptions that are pretty obvious from your argument. But beware! Sometimes what is obvious to you will not be obvious to your reader. When in doubt, state your assumptions.

Presenting your argument

When you write an argument, you should clearly state all of the components of your argument and how they are connected.

Sometimes you will have to make a chain of arguments. For example, if we were to argue that a language has prepositions, we first have to identify an adposition and which noun it goes with, and then show that the adposition goes before that noun.

Sometimes a chain of argumentation will include both *empirical arguments* and *theoretical arguments.* For example, our argument that children use rule-based language actually has both empirical and theoretical components.

As you lay out all of the arguments in your line of reasoning and each of their components, think about what order it is best to lay out your claims. Do any of your points depend on another point?

Always make sure you use technical terminology where appropriate. Technical terminology is useful because it is normally concisely defined to avoid ambiguity. For example, saying *un-* is a morpheme is more precise than saying *un-* is part of a word.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1282#h5p-23

CHAPTER 3 PRACTICE EXERCISES

Data analysis

Exercise 1. Competence vs. performance

Beginner [3.2]

Consider each of the following situations. Does the language user in each case have competence only, performance only, both, or neither? Explain your answer.

a. Sandy has just been to the dentist and now her mouth is numb. When she talks, it is difficult to understand her, but the numbness will wear off soon and Sandy will be able to talk clearly again.

b. Kai is 18 months old. He knows several words but doesn't combine them into sentences.

c. Cherry the dog knows the word *walk* and gets very excited every time she hears the word *walk*. When Cherry's owner twists his ankle, he tells his mom on the phone, "I can't walk." Cherry hears this, barks excitedly, and fetches her leash.

d. Nissa and Remi are chatting at a party. Nissa is telling Remi about her shopping trip, when Remi spills his drink. Nissa stops midsentence to help Remi clean up and forgets what she had been talking about.

e. A company has just released a new AI program called ASLBot and everyone is talking about it. The program is able to produce videos of a cartoon character signing in American Sign Language based on a Large Language Model. You can even turn on your webcam and have a relatively natural two-way conversation with ASLBot that seems fairly human-like. You research how ASLBot is programmed, and you discover it is programmed based on a probabilistic model of language, rather than a generative one.

Communication and study skills

Exercise 2. Theoretical and empirical arguments

Advanced [3.7]

Pick an argument from the textbook or from the lecture to analyze.

a. Is it a theoretical argument or an empirical argument? Or does it have bits of both?

b. For the empirical arguments: What is the observable evidence? What is the claim? How was the claim backed up? Can you think of another way to analyze the same data?

c. For the theoretical arguments: What are the assumptions, the premises, and the conclusions? Do you agree with the premises and the assumptions?

Research and application

Exercise 3. Trick an LLM

Advanced [3.6]

Explore a freely-available LLM program and see if you can find some of its weaknesses. Explain what you find.

a. Can you find a tricky grammatical construction that it misunderstands?

b. Did it give you any false information?

c. Did its responses reveal any biases or stereotypes?

Language journal

At the end of Chapter 1, we introduced the Language Journal. You should have picked a language to build a profile for as you work through this textbook. You can add to your profile with these questions.

Exercise 4. Find a grammar

Beginner [3.1]

Are there any published grammars about your language available in your university library? Are they pedagogical, descriptive, or theoretical grammars?

4. LINGUISTIC DATA IN MORPHOSYNTAX

In this chapter, we look at what types of data we consider in morphology and syntax. We will learn about how to collect it, and the strengths and weaknesses of the different kinds of data collection methods. We will also learn about how to format and present data according to linguistic convention.

Learning objectives

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

- Describe different modalities and varieties of language using technical terms.
- Describe the strengths and weaknesses of corpus studies, elicitation, and experimentation.
- Discuss ethical issues in data collection and use.
- Present linguistic data in a reader-friendly format, following convention.

4.1. LANGUAGE MODALITY

In linguistics, language **modality** is the medium or format through which language is conveyed. Human language shows up in at least four different modalities:

- Spoken language, also known as vocal-auditory language, is produced by making sounds with the vocal tract and is perceived through hearing.
- Signed language, also known as manual-visual language, is produced by making movements with the hands, face, and body, and is perceived through seeing.
- Tactile language, also known as manual-tactile, is produced by making movements with the hands, face, and body, and is perceived through touch.
- Graphemic language is produced by making markings such as symbols on some material and is perceived through seeing.

Most often, in natural settings, language is multi-modal, meaning we use multiple modalities simultaneously to communicate (Henner and Robinson 2023). For example, someone communicating through spoken language will also frequently use body posture and hand gestures at the same time.

Our model of language needs to take into account all of the attested modalities of human language. Studying different modalities can give us different clues about which aspects of language are a result of the physical constraints of a particular modality, and which are a result of our cognitive dispositions. Historically, the field of linguistics has overlooked signed and tactile languages and focused primarily on spoken language.

A terminological note on the words *signed languages* and *deaf*

Signed languages are sometimes called *sign languages*. Both terms are generally acceptable, so you may encounter either one in linguistics writing. *Sign languages* has long been the more common term, but *signed languages* has recently been gaining popularity among deaf scholars.

Another piece of relevant terminology that is in flux is the long-standing distinction in capitalization between uppercase *Deaf* (a sociocultural identity) and lowercase *deaf* (a physiological status). However, this distinction has been argued to contribute to elitist gatekeeping within deaf

communities, so many deaf people have pushed to eliminate this distinction (Kusters et al. 2017, Pudans-Smith et al. 2019).

In this textbook, we follow these prevailing modern trends by using *signed languages* and by not using the *Deaf/deaf* distinction. However, the alternatives are still widespread in linguistics writing, so you may still encounter them.

For these issues, it is important to proceed with caution and follow the lead of anyone more knowledgeable than you, especially if they are deaf. If you are uncertain what usage is appropriate in a given situation with a given deaf person, ask what they prefer.

Why writing is different

Spoken, signed, and tactile languages are all primary modalities, but writing, a form of graphemic language, is a bit different. Unlike the other three modalities, writing is secondary to another modality, because it is a written representation of a language that occurs primarily in another modality, usually spoken language. For example, English writing is a representation of spoken English. Individual letters in written English represent different sounds in spoken language.

Here are some ways that writing is different:

- Other modalities of language are transient. That is, they disappear as soon as they are produced. Writing is the only language modality that can last independently of the language user.[1] Because of this, writing is often not as interactive as other modalities of language, which, in turn, affects discourse structure and other properties.
- Other modalities of language are acquired subconsciously by instinct, while writing must be consciously learned.
- Unlike spoken and signed languages, writing has only been independently invented a handful of times throughout human history. Many languages do not have a written form at all. Most of the languages that do have a written form borrowed at least the idea from another language community. In contrast, spoken languages occur in every known human community with hearing populations, and deaf children who are not exposed to signed languages will develop a homesign language with their families.

1. Since the invention of audio and video recording devices, we can also keep records of spoken, signed, and tactile language. However, these are all recordings of a language user producing the language—they do not separate the language from the language user in the same way as writing.

- Writing has different genre conventions than spoken language. Writing tends to be more formal and language change in writing lags behind change in spoken language.

Because of these differences, linguists do not treat written language the same. Because it is not subconsciously acquired, there is no reason to think it would be part of **Universal Grammar**. For these reasons, it is often intentionally excluded from our model of language. However, we still sometimes use written language as a clue to patterns in spoken language, especially for older varieties of language for which only a written record remains.

We have to be careful, however, that we do not dismiss all graphemic forms of language as secondary to spoken language. There are many forms of graphemic communication that are not directly linked to a spoken language, at least some of which display some complex properties (see Henner and Robinson 2023).

> Dr. Melanie McKay-Cody, a deaf Cherokee professor, demonstrates that petroglyphs—stone carvings—from the Northern Ute in Utah and the Kiowa in Texas incorporate signs from North American Indian Sign Language (NAISL). She argues that these petroglyphs are a form of written signed language. Although a few different transcription methods for signed language have been proposed in modern times, none have really caught on. But perhaps there has been signed language writing for hundreds, if not thousands, of years!

Ethics and modality

As previously mentioned, the field of linguistics has primarily focused on spoken languages, which Henner and Robinson (2023) call *modality chauvinism*. This focus on spoken modality has its roots in ableism, since other modalities are often used by people with disabilities. The solution seems simple—include more signed languages in our studies! But even well-intentioned linguists will face some difficulties incorporating signed languages in their research. For example, the lack of a writing system for signed languages makes them more difficult to write about, contributing to their marginalization in research (Henner and Robinson 2023). Secondly, signed languages are best portrayed by video, and so signed language data cannot be anonymized and require different copyright permissions. Most importantly, however, signed and tactile modalities are used by communities which are marginalized both within linguistics and in wider society, and therefore require extra care from researchers. We will discuss more about how to conduct research with marginalized communities in Chapter 5. These extra difficulties do not excuse us from including signed and tactile languages in our research! But they must be taken into account when planning a research study.

Other modalities for other species?

Some animal species use different modalities for their communication systems. For example, ants are known to use scents to communicate. However, as far as we can tell, these communication systems do not seem to have all of the properties of language.

In some science fiction settings, though, authors have also invented full language in other modalities. For example, in Becky Chambers' *The Wayfarers* series, a species called the Aeluon communicate through patches on their cheeks that change colours. In the movie *Arrival*, the alien species have a form of written language that does not seem to be secondary to spoken language. Many, many science fiction stories also have some sort of telepathic communication in their fictional worlds.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1467#h5p-26

References and further resources

Attribution

The call-out box 'A terminological note' *is adapted from the following CC BY NC source:*

Anderson, Catherine, Bronwyn Bjorkman, Derek Denis, Julianne Doner, Margaret Grant, Nathan Sanders, and Ai Taniguchi. 2022. Section 3.1: Modality. *Essentials of Linguistics, 2nd edition.* Pressbooks. https://ecampusontario.pressbooks.pub/essentialsoflinguistics2/chapter/3-1-modality/

Linguistics in the media

🎉 Chambers, Becky. 2014–2021. *The Wayfarers* series. Hodder & Stoughton. https://www.otherscribbles.com

- Book 1: The long way to a small, angry planet
- Book 2: A closed and common orbit
- Book 3: Record of a spaceborn few
- Book 4: The galaxy, and the ground within

🎉 Villeneuve, Denis (director). 2016. *Arrival*. FilmNation Entertainment, Lava Bear Films, and 21 Laps Entertainment.

For a general audience

🔍 Canadian Language Museum. n.d. Sign Languages of Canada 3D. Virtual museum exhibit. https://languagemuseum.ca/online-exhibits/sign-languages-of-canada-3d

🔍 📘 Heumann, Judy and Melanie McKay-Cody. 2023. Indigenous sign languages with Dr. Melanie McKay-Cody. *The Heumann Perspective*. https://judithheumann.com/melaniemckaycody/

Edwards, Terra. 2024. *Going tactile: Life at the limits of* language. Oxford: Oxford University Press.

🔍 Leland, Andrew. 12 May 2022. DeafBlind communities may be creating a new language of touch. *The New Yorker*. https://www.newyorker.com/culture/annals-of-inquiry/deafblind-communities-may-be-creating-a-new-language-of-touch

🔍 Quartz. 2016. Pro-tactile ASL: A new language for the DeafBlind. YouTube video. https://www.youtube.com/watch?v=9GrK3P15TYU

For students

🧠 Anderson, Catherine, Bronwyn Bjorkman, Derek Denis, Julianne Doner, Margaret Grant, Nathan Sanders, and Ai Taniguchi. 2022. Section 3.1: Modality. *Essentials of Linguistics, 2nd edition*. Pressbooks. https://ecampusontario.pressbooks.pub/essentialsoflinguistics2/chapter/3-1-modality

Academic sources

🔍📘 Henner, Jon, and Octavian Robinson. 2023. Unsettling languages, unruly bodyminds: A Crip

Linguistics manifesto. *Journal of Critical Study of Communication and Disability* 1(1): 7–37. https://criticalstudycommunicationdisability.org/index.php/jcscd/article/view/4

Kusters, Annelies, Maartje De Meulder, and Dai O'Brien. 2017. Innovations in Deaf Studies: Critically mapping the field. In *Innovations in Deaf Studies: The role of deaf scholars*, ed. Annelies Kusters, Maartje De Meulder, and Dai O'Brien, Perspectives on Deafness, 1–56. Oxford: Oxford University Press.

McKay-Cody, Melanie. 2019. *Memory comes before knowledge—North American Indigenous Deaf: Socio-cultural study of rock/picture writing, community, sign languages, and kinship*. PhD dissertation, University of Oklahoma. https://shareok.org/handle/11244/319767

Pudans-Smith, Kimberly K., Katrina R. Cue, Ju-Lee A Wolsley, and M. Diane Clark. 2019. To Deaf or not to deaf: That is the question. *Psychology* 10(15): 2091–2114.

4.2. METHODS OF DATA COLLECTION

There are three main ways we collect data for use in morphology and syntax: **corpus** studies, **linguistic elicitation**, and experimentation.

Types of data collection

Corpus studies

A **corpus** study is based on a collection of found real-world data. A corpus (plural: *corpora*) can be compiled from written text, such as a collection of social media posts, newspaper articles, or books, or it can be compiled from video or audio recordings, such as a set of talk show interviews, political speeches, or podcasts.

Sometimes researchers compile their own corpora, but there are also several corpora that have already been compiled and are available for use, such as the Corpus of Contemporary American English (COCA).

If a word or construction has been found in a corpus, we say it is **attested**. If it has not been found, we say it is **unattested**.

Linguistic elicitation

In linguistic elicitation, linguists work with a user of a language to collect linguistic data from that language. Elicitation is often bilingual, which means that data collection is mediated by another language. The linguist may ask the language user to translate a word, phrase, or sentence into the target language, or the linguist may construct examples in the target language and ask the language user if the constructed examples are acceptable.

Linguistic elicitation can also be performed even if the linguist and the language user have no language in common, although it is more difficult. Daniel Everett gave a demonstration of how to perform a monolingual elicitation at the 2013 Linguistic Institute, which was recorded and posted here. To perform a monolingual elicitation, you can use props or images to elicit words and act out situations or use videos to elicit sentences.

Although bilingual elicitation is a lot easier, it also increases the likelihood that the language user will be influenced by the mediating language. For example, if two word orders are permitted in the target language, a language user might be more likely to use the word order that matches with the word order of the mediating language, even if it is the less common one in the target language.

Another method of data collection that is closely related to elicitation is introspection. If you are a speaker

of the language that you are studying and you make up your own data based on your own personal judgments, that is called introspection.

Experimentation

An experiment is a highly controlled procedure, usually done in a research lab. The researcher will recruit multiple research participants, and ask them to perform the same task, in as close to the same conditions as possible, and observe their behaviour.

Experiments in morphosyntax can be simple, such as a survey which collects grammaticality judgments from a large number of people, or more complex, such as an eye-tracking study that pays attention to where research participants are looking when reading a sentence. A pause or slower reaction time can be a clue that the structure is more complex.

Classification of data collection methods

We can classify these methods of data collection by whether they collect observational data or targeted data. Targeted data is data that is specifically sought out by the researcher to test a hypothesis. Both elicitation and experimentation are targeted. In elicitation, the researcher asks the language user for specific constructions that provide evidence for or against their hypothesis. In experiments, the researcher designs the methods of their experiment to provide evidence supporting or contradicting their hypothesis. Observational data, on the other hand, is naturally occurring data that has been observed. Corpus studies use observational data, as the data was created for independent purposes and later analyzed.

We can also classify these methods of data by whether they are categorical or quantitative. Categorical data is sorted into categories, such as grammatical or ungrammatical. Elicitation results in categorical data, based on the judgments of the language user. Quantitative data is data that has been counted and statistically analyzed. Corpus studies result in quantitative data, such as the frequency of a particular word or construction. Experiments also result in quantitative data, such as the rate at which research participants exhibit a particular behaviour.

Table 1. Classification of data collection methods

	Categorical data	**Quantitative data**
Observational data		Corpus study
Targeted data	Elicitation	Experiment

Which methods are best?

Each of the methods have different strengths and weaknesses. Which method is most appropriate for a given study will vary based on the research question and several other factors. It's best, if possible, to use multiple methods–hopefully the different methods will all converge on the same results!

In this next section, we will look at two factors that should be considered when choosing a method: negative evidence and the resources of the language community.

Negative evidence

In Section 2.2, we learned that **negative evidence**, which is evidence that something is not possible, did not occur, or is absent, is important for linguistic analysis. Which methods might produce negative evidence?

Let's do a little thought experiment. Let's say we looked through the entire publishing history of *The Toronto Star* to look for the two sentences in (1) and didn't find them.

(1) a. The prime minister of Canada ate a salami sandwich.

b. *The ate of salami a Canada ministerial prime sandwich.

Does this mean that both sentences (1a) and (1b) are not possible? According to my intuition, sentence (1a) is a possible sentence, but it just hasn't been used by *The Toronto Star*. On the other hand, sentence (1b) is not a possible sentence of English.

We can't tell the difference between a sentence that is possible but **unattested** and a sentence that is not possible at all just by looking at the sentences that have been produced in the past. Since language is creative and **productive**, not all possible sentences have been produced in the past. Arguably, *most* possible sentences haven't been produced yet. Because of this, no conclusions can be drawn if the pattern you are looking for doesn't show up in your corpus. As Carl Sagan famously said, "Absence of evidence is not evidence of absence."

Corpus studies, since they result in observational data and not targeted data, cannot provide negative evidence. Elicitation and experimentation, on the other hand, can provide negative evidence, depending on their design.

Resources of the language community

Another important factor to consider when choosing a data collection method is the resources of the language community. If you are doing a study on English, it is easy to recruit a large number of English speakers to participate in a research study, and there are multiple electronic corpora that can be searched with just a few clicks. But English is one of the best-resourced languages in the world, and the situation is not the same

for many minority and endangered languages. However, minority and endangered languages are crucial for helping us understand the breadth of the diversity of human language!

> Minority and endangered languages are exactly the languages for which it may be impossible to gather large numbers of participants, and where the absence of literacy, the age of the speakers, and other factors make certain types of experiments unfeasible. (Davis et al. 2014: e187)

Even within well-resourced languages like English, corpora have biases. There are numerous dialects of English, which may all get mixed together in a single corpus, obscuring the differences between the dialects and making it unclear which groups of people use which language patterns. Minority dialects might not show up at all, or show up infrequently enough that they get excluded by the statistical analysis.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1068#h5p-25

References and further resources

For linguistics students

🔍 Everett, Daniel. 2013. Monolingual demonstration. *Linguistic Institute.* https://www.youtube.com/watch?v=sYpWp7g7XWU

Academic sources

Davies, Mark. (2008–) *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca

Davis, Henry, Carrie Gillon, and Lisa Matthewson. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language 90*(4), e180–e226

4.3. LANGUAGE, DIALECT, AND GRAMMAR

One language, many grammars

When we are collecting linguistic data to study, we are usually collecting data from a particular language. But it turns out that the notion of *language* is very messy. The notion of language seems straightforward because we are used to thinking and speaking of "the English language," "the French language," "the Swahili language," and so forth. But these terms are actually much vaguer than they seem at first glance because they cover a large number of varieties, including ones that differ enough to be mutually unintelligible. For instance, *Ethnologue* distinguishes 32 dialects of English in the United Kingdom alone. In addition, distinct dialects of English are spoken in former British colonies, including Australia, Canada, India, New Zealand, South Africa, the United States, and many other African, Asian, and Caribbean nations, and many of these dialects have subdialects of their own. Similarly, *Ethnologue* distinguishes 11 dialects of French in France and 10 dialects of Swahili in Kenya, and there are further dialects in other countries in which these languages are spoken. Moreover, we use terms like "the English language" to refer to historical varieties, which may differ profoundly from modern-day varieties. For example, present-day English differs from Old English about as much as modern English differs from German.

Although the most salient differences between dialects are often phonological (that is, speakers of different dialects often have different accents), dialects of a so-called single language can differ syntactically as well. For instance, in standard French, as in the Romance languages more generally, adjectives ordinarily follow the noun that they modify. But that order is reversed in Walloon, a variety of French spoken in Belgium. The two parametric options are illustrated in (1).

(1) a. *Standard French*

un chapeau noir

a hat black

'a black hat'

b. *Walloon*

on neûr tchapê

a black hat

'a black hat'

(Bernstein 1993: 25-26)

In this example, we have two different dialects of "the same language", French, differing with respect to a language parameter. The opposite is also possible: two "different languages" that are basically the same. For example, the same linguistic variety spoken on the Dutch-German border may count as a dialect of Dutch or German depending on which side of the political border it is spoken, and the same is true of many other border dialects as well. This is called a dialect continuum, where each adjacent language variety is quite similar, but the extreme ends of the continuum are quite different from each other.

According to Max Weinreich, "a language is a dialect with an army and a navy." A striking (and sad) confirmation of this aphorism concerns the recent terminological history of Serbo-Croatian. As long as Yugoslavia was a federal state, Serbo-Croatian was considered a single language with a number of regional dialects. The 14th edition of Ethnologue, published in 2000, still has a single entry for Serbo-Croatian. In the 15th edition, published in 2005, the single entry is replaced by three new entries for Bosnian, Croatian, and Serbian.

As the previous discussion has shown, the notion of language is based more on sociopolitical considerations than on strictly linguistic ones. By contrast, the term 'grammar' refers to a particular set of parametric options that a speaker acquires. The same language label can be associated with more than one grammar (the label "French" is associated with grammars that have adjective-noun order and ones with noun-adjective order), and a single grammar can be associated with more than one language label (as in the case of border dialects). In other words, one particular **E-language** might be a mishmash of many different **I-languages**, as discussed in Section 3.2.

Mutual intelligibility

It is important to distinguish the concept of shared grammar from **mutual intelligibility**. To a large extent, standard English and many of its nonstandard varieties are mutually intelligible even where their grammars differ with respect to one parameter or another. On the other hand, it is perfectly possible for two or more varieties that are mutually unintelligible to share a single grammar. For instance, in the Indian village of Kupwar (Gumperz and Wilson 1971), the three languages Marathi, Urdu, and Kannada, each spoken by a different ethnic group, have been in contact for about 400 years, and most of the men in the village are bi- or trilingual. Like the standard varieties of these languages, their Kupwar varieties have distinct vocabularies, thus rendering them mutually unintelligible to monolingual speakers, but in Kupwar, the considerable grammatical differences that exists among the languages as spoken in other parts of India have been virtually eliminated. The difference between standard French and Walloon with respect to prenominal adjectives is another instance of this same convergence phenomenon. Here, too, the adjective-noun order in Walloon is due to language contact and bilingualism, in this case between French and Flemish, the other language spoken in Belgium; in Flemish, as in the Germanic languages more generally, adjectives ordinarily precede the nouns that they modify.

One speaker, many grammars

Finally, we should point out that it is perfectly possible for a single speaker to acquire more than one grammar. This is most strikingly evident in balanced bilinguals. Speakers can also acquire more than one grammar in situations of syntactic change. For instance, in the course of its history, English changed from an object-verb order language to a verb-object order language, and individual speakers during the transition period (which began in late Old English and continued into Middle English) acquired and used both parametric options. Speakers can also acquire more than one grammar in situations of diglossia or stable syntactic variation. For instance, English speakers who speak a non-prestigious dialect of English at home might acquire the prestigious parametric variants at school, and switch between them depending on context.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1469#h5p-27

References and further resources

Attribution

This section is adapted from the following CC BY NC source:

Santorini, Beatrice, and Anthony Kroch. 2007. *The syntax of natural language: An online introduction.* https://www.ling.upenn.edu/~beatrice/syntax-textbook

Academic sources

Bernstein, Judy. 1993. Topics in the syntax of nominal structure across Romance. Ph.D. dissertation, City University of New York.

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2023. *Ethnologue: Languages of the World*. Twenty-sixth edition. Dallas, Texas: SIL International. http://www.ethnologue.com.uml.idm.oclc.org

Gumperz, John J., and Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In Dell Hymes, ed., Pidginization and creolization of languages. Cambridge: Cambridge University Press. 151–167.

4.4. WHAT IS GRAMMATICALITY?

The aim of a generative grammar is to generate all and only the grammatical sentences of a language. Since the notion of grammaticality is basic to syntactic theory, it is important to distinguish it from notions with which it is easily confused.

If you are collecting data on a language, and a speaker tells you that a sentence seems odd, they don't usually know why. It is up to the linguist to figure out why!

Grammaticality vs. making sense

First and foremost, 'is grammatical' is not the same thing as 'makes sense.' The sentences in (1) all 'make sense' in the sense that it is easy to interpret them. Nevertheless, as indicated by the asterisks, they are not grammatical.

(1) a. *Is our children learning?

b. *Me wants fabric.

c. *To where are we be taking thou, sir?

d. *The introduction explained that "the Genoese people, *besides of hard worker, are good eater too, and even 'gourmand,'* of that honest gourmandise which will not drive a man to hell but which is, after all, one of the few pleasures that mankind can enjoy in this often sorrowful world."

Conversely, sentences can be grammatical, but not 'make sense.' Two examples are given in (2). Since the sentences are grammatical, they aren't preceded by an asterisk. Their semantic anomaly can be indicated, if desired, by a prefixed pound sign (hash mark).

(2)	a.	#Colorless green ideas sleep furiously. (Chomsky 1965: 149)	– cf. Revolutionary new ideas appear infrequently.
	b.	#I plan to travel there last year.	– cf. I plan to travel there next year.

Grammaticality vs. difficulty of processing

Second, 'grammatical' must be distinguished from 'easily processable by human beings.' This is because it turns out that certain well-motivated simple grammatical operations can be applied in ways that result in sentences

that are virtually impossible for human beings to process. For instance, it is possible in English to modify a noun with a relative clause, and sentences containing nouns that are modified in this way, like those in (3), are ordinarily perfectly acceptable and easily understood. (Here and in the following examples, the relative clauses are bracketed and the modified noun is underlined.)

(3) a. The mouse [that the cat chased] escaped.

b. The cat [that the dog scared] jumped out the window.

But now notice what happens when we modify the noun within the relative clause in (3a) with a relative clause of its own.

(4) The mouse [that the cat [that the dog scared] chased] escaped.

Even though (4) differs from (3a) by only four additional words and a single additional level of embedding, the result is virtually uninterpretable without pencil and paper. The reason is not that relative clause modification can't apply more than once, since the variant of (3a) in (5), which contains exactly the same words and is exactly as long, is perfectly fine (or at any rate much more acceptable than (4)).

(5) The mouse escaped [that the cat chased] [that the dog scared].

The reason that (4) is virtually uninterpretable is also not that it contains recursive structure (the relative clause that modifies *mouse* contains the relative clause that modifies *cat*). After all, the structures in (6) are recursive, with up to 5 levels of embedding, yet they don't throw us for a loop the way that (4) does.

(6) a. She won.

b. The Times reported that [she won].

c. John told me that [the Times reported that [she won]].

d. I remember distinctly that [John told me that [the Times reported that [she won]]].

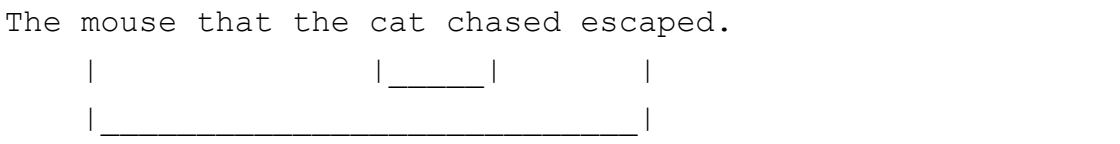
e. They don't believe that [I remember distinctly that [John told me that [the Times reported that [she won]]]].

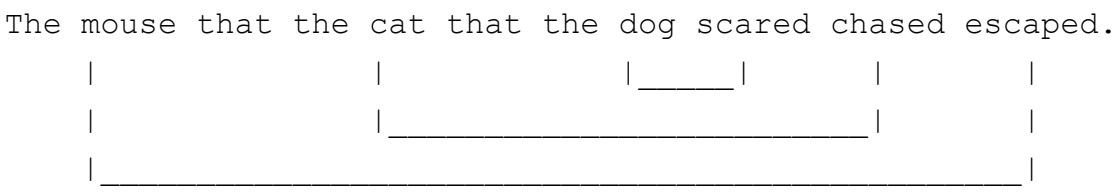
f. I suspect that [they don't believe that [I remember distinctly that [John told me that [the Times reported that [she won]]]]].

Example (4) is unacceptable not because it is ungrammatical, but because of certain limitations on human short-term memory (Chomsky and Miller 1963: 286, Miller and Chomsky 1963: 471). Specifically, notice that in the (relatively) acceptable (5), the subject of the main clause *the mouse* doesn't have to "wait" (that is, be kept active in short-term memory) for its verb *escaped* since the verb is immediately adjacent to the subject. The

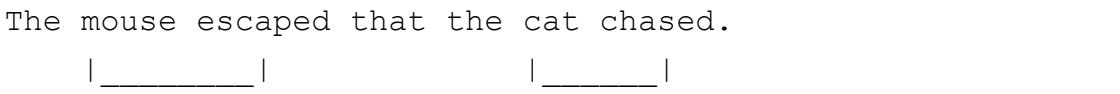
same is true for the subjects and verbs of each of the relative clauses (*the cat* and *chased*, and *the dog* and *scared*). In (4), on the other hand, *the mouse* must be kept active in memory, waiting for its verb *escaped,* for the length of the entire sentence. What is even worse, however, is that the period during which *the mouse* is waiting for its verb *escaped* overlaps the period during which *the cat* must be kept active, waiting for its verb *chased.* What makes (4) so difficult, then, is not the mere fact of recursion, but that two relations of exactly the same sort (the subject-verb relation) must be kept active in memory at the same time. In none of the other relative clause sentences is such double activation necessary. For instance, in (3a), *the mouse* must be kept active for the length of the relative clause, but the subject of the relative clause (*the cat*) needn't be kept active since it immediately precedes its verb *chased*.

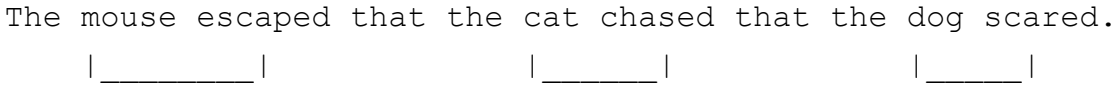
Sentences like (3) and (4) are often referred to as **centre-embedding** structures, and the dependencies between the subjects and their verbs are said to be **nested**.





By contrast, the corresponding dependencies in (5) are not nested.





Grammaticality across contexts

A final important point to bear in mind is that any sentence is an expression that is paired with a particular interpretation. Grammaticality is always determined with respect to a pairing of form and meaning. This

means that a particular string can be grammatical under one interpretation, but not under another. For instance, (7) is ungrammatical under an subject-object-verb (SOV) interpretation (that is, when the sentence is interpreted as *Sue hired Tom*).

(7) Sue Tom hired.

Example (7) is grammatical, however, under an object-subject-verb (OSV) interpretation (that is, when it is interpreted as *Tom hired Sue*). On this interpretation, *Sue* receives a special intonation marking contrast, which would ordinarily be indicated in writing by setting off *Sue* from the rest of the sentence by a comma. In other words, the grammaticality of (7) depends on whether its interpretation is analogous to (8a) or (8b).

(8) a. *She him hired.

b. √Her, he hired. (The other job candidates, he didn't even call back.)

Related to this idea is that context matters. For example, sentence (7), even though it has a grammatical interpretation, is a fairly odd way to word that sentence. However, in some contexts, it sounds more natural than in other contexts. Sometimes it is fairly easy to think of a context where a sentence is grammatical, but other times it can be quite tricky.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1479#h5p-28

References and further resources

Attribution

This section is adapted from the following CC BY NC source:

Santorini, Beatrice, and Anthony Kroch. 2007. *The syntax of natural language: An online introduction.* https://www.ling.upenn.edu/~beatrice/syntax-textbook

Comedy and satirical linguistics

🎉 Bakery, Yreka. 2006. New speech disorder linguists contracted discovered! *Speculative Grammarian* CLI(2). https://specgram.com/CLI.2/03.bakery.disorder.html

Academic sources

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam, and George A. Miller. 1963. Introduction to the formal analysis of natural languages. In R. Duncan Luce, Robert R. Bush, and Eugene Galanter, eds., *Handbook of mathematical psychology*, vol. 2, 269–321. New York: Wiley.

Miller, George A., and Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush, and Eugene Galanter, eds., *Handbook of mathematical psychology*, vol. 2, 419–491. New York: Wiley.

4.5. THE VALIDITY OF SPEAKER JUDGMENTS

Who should we ask for judgments?

Common wisdom among linguists is that we should only be asking **native speakers** for linguistic judgments. We are trying to get at the subconscious knowledge about language that is acquired by instinct during early childhood. If a speaker learned the language later in life, then their language may have subtle differences compared to a native speaker, because their judgments may be influenced by their first language. But is this common wisdom all there is to say about it?

First of all, it's important to keep in mind that everyone's judgments can give us information about language. A multilingual language user can give us information about how multilingual brains work and how different grammars interact with each other. A neurodivergent language user or a language user with a language disorder can give us information about variation in language processing. However, if we are trying to develop a grammar of a particular language, neurotypical monolinguals can give us data that has the fewest confounding factors.

But it is not always as straightforward as it sounds to find a native speaker with no confounding factors. If you are studying a language using bilingual elicitation, you will need to find someone who has advanced proficiency in *both* languages, and there will be influence between the two languages. Even monolinguals almost always have command of more than one dialect or variety, and so there is always a chance of influence between multiple grammars.

In contrast, some native speakers do not have sufficient command of their native language to act as a language informant. For example, someone who immigrates in childhood may become dominant in their second language and may even lose their native language completely if they stop having opportunities to use it.

There are also some situations where the idea of a native speaker itself is problematic. For example, approximately 90% of deaf children are born to hearing parents and many of these do not learn to sign until they go to school, and thus arguably do not have a native language at all. When studying a signed language, is it better to find one of the rare "native" speakers, or is it better to work with a typical signer, one who learned the signed language a little bit later in life?

Racism also plays a role in who is considered a "native speaker." It is not uncommon for "native speaker status" to be denied to those who speak less prestigious varieties of the language, especially varieties associated with people of colour, even if it is their first language. For example, consider the following excerpt from Nuzhat Amin's (1997) short reflection on how her native dialect of Pakistani English is perceived in Canada.

> Another source of disempowerment for South women is the preference given to native speakers of English, and the assumption that only white people are native speakers, and therefore, that only white people know "good" English. I consider myself to be a native speaker of English on the grounds that English is the language I know best, but my colleagues — teachers of English and English as a second language (ESL), linguists, and applied linguists — often position me as a non-native speaker, I would say, because I am non-white and I have a Pakistani accent. When I say that I am a native speaker, there is a look of bewilderment and disbelief on their faces.
>
> (Amin 1997: 141)

Judgments can be messy

When you read a finished, typed-up linguistics paper, the data is usually all organized, with no sign of controversy. Each piece of data is marked clearly either as grammatical or ungrammatical.

However, if you ever try to do collect your own data, it soon becomes obvious that it does not start out quite so tidy. When you are working with real people, a lot of things can confuse the data. You might get two different people giving you two different judgments for the same sentence. Sometimes even the same person will give you a different judgment from one day to the next.

Here are some of the reasons judgments are messy:

- The language has more than one variety or dialect. Two speakers may have slight differences in their grammar. Even a single speaker might alternate between their multiple grammars from one situation to the next.
- Sometimes a sentence is only grammatical in certain contexts. It's possible that the context your language consultant is imagining has changed your results. One way to reduce this problem is to set up a context for them.
- If you ask several judgments in a row, especially if they're similar to each other or tricky constructions, a speaker may start to have judgment fatigue. Judgment fatigue can either cause everything to sound wrong, or cause everything to sound right.
- Some constructions are marginally grammatical, which means they're not great but they're also not terrible. Marginal sentences are often marked with a question mark (?) instead of an asterisk (*).

Statistically testing judgments

The practices of using elicitation and introspection as the basis for linguistic research is sometimes criticized as not being sufficiently rigorous. Sometimes linguists who use introspection as their primary method of study get called "armchair linguists" disparagingly.

However, we don't need to just assume that a particular method is or is not valid. A study by Sprouse, Schütze, and Almeida (2013) responded to these criticisms by collecting nearly 300 different examples from published papers in one of the top academic journals in linguistics and testing them using multiple different methodologies. They found that these different methodologies, including informal methods like elicitation and introspection, gave the same result 95% of the time, showing that these different methodologies meet a reasonably reliable benchmark.

However, this study only tested English data published in an English journal. When a research article in linguistics goes through the peer-review process, the reviewers typically also consider the data in the paper, as well as the argument. So, when a paper is reviewed by reviewers who know the language that the paper is about, the data goes through an additional process of verification, compared to a paper that is reviewed by those who do not know the language. Therefore, errors in data from lesser-studied and lesser-known languages are more likely to slip through the review process than errors in well-known languages like English. Furthermore, making mistakes about lesser-studied languages have more severe consequences. With fewer resources available on the language, it is more likely that the mistake will be perpetuated instead of corrected.

Thus, even though elicitation and introspection can provide us with reliable data, we do need to make sure we use them carefully. We should try to replicate our results—double-checking our data with the same speaker in follow-up sessions, as well as checking with multiple speakers. And if a grammaticality judgment is marginal or otherwise tricky, or if there seems to be a lot of dialectal variation, we should consider using alternative methods alongside elicitation and introspection.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1523#h5p-29

References and further resources

Academic sources

🔍📚 Amin, Nuzhat. 1997. South Englishes, North Englishes. *Canadian Woman studies* 17(2). https://cws.journals.yorku.ca/index.php/cws/article/view/8930

Birkeland, Anne, Adeli Block, Justin T. Craft, Yourdanis Sedarous, Sky Wang, Alexis Wu, and Savithry Namboodiripad. 2024. Rejecting nativeness to produce a more accurate and just Linguistics. *Language* 100 (3): e156-e194.

Sprouse, Jon, Carson T. Schütze, and Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from *Linguistic Inquiry* 2001-2010. *Lingua* 134: 219–248.

Truan, Naomi. 2024. Whose language counts? Native speakerism and monolingual bias in language ideological research: Challenges and directions for further research. *European Journal of Applied Linguistics* 12 (1): 34-53.

4.6. BECOMING A LINGUIST: DISCUSSING DATA

In this section, we are going to discuss how to write about linguistic data. In morphology, our data is often words or lists of words. In syntax, our data is often phrases and sentences.

Two technical terms will be useful in this section. The **object language** is the language under discussion, or the language being analyzed. The **metalanguage** is the language that we are using in order to discuss the object language. In this section, I am using examples from Gonzalez (2023), which is a paper written in English about questions in Finnish and Turkish. The object languages of this paper are Finnish and Turkish, while the metalanguage is English.

Numbering data

When we include linguistic data in our papers, we usually separate it out from the main text or prose of our argument. Examples that are separated out this way are each numbered sequentially throughout the text. An example is shown below from the opening paragraph of Gonzalez (2023).

> The strategy used for forming polar questions varies across languages. For instance, in languages like English, polar questions are formed using a raising intonation (marked as ↑) and subject-auxiliary inversion, as illustrated in (1). In contrast, in other languages, polar questions are formed using an interrogative particle. Japanese is an example of such a language. As shown in example (2), polar questions can be formed using the interrogative particle *ka*.

(1) Are you leaving ↑ ?

(2) *Japanese* (Uegaki 2018)

Hanako-ga	hashitta-ka?
Hanako-NOM	ran-KA

'Did Hanako ran?'

(Gonzalaz 2023: 2)

Related examples may share the same number, with sub-examples labelled with letters.

(4) a. *Finnish*

Lähti-**kö** Mari?

left-PolQP Mari

'Did Mary leave?'

b. *Turkish*

Oya Dilara'yı **mı** öp-tü?

Oya Dilara-ACC PolQP kiss-PST

'Did Oya kiss Dilara?'

(Gonzalez 2023: 3)

Definitions, formulas, and syntax trees are also sometimes numbered alongside the examples.

If you repeat an example several pages after you first discuss it, it is good to repeat the example so your reader does not need to flip back and forth. Typically, authors will give the repeated example a new number, but also tell you the original number, as shown below.

Consider again example (5), repeated below as (23).

Embedded examples

We also sometimes include examples embedded directly into a sentence, instead of setting them apart and numbering them. This is generally only done with really short examples, such as one-word examples. This

format can also be used when discussing portions of a longer example. This should really only be done when the example does not need to be **glossed**, either because the example is simple enough to not need it, or because the example is glossed elsewhere close by in the text.

When we include examples embedded into a sentence, the word(s) in the object language are typically in italics. In fact, any time you mention a word (rather than use the word), it should be in italics. If the object language is different than the metalanguage, you should also include a translation of the word(s) in single quotes.

In the following example, *pu:ch* 'ask' and *ja:n* 'know', both Hindi-Urdu words, are formatted in this way. On the other hand, *kya:* is not translated, because this sentence is part of a larger passage about the details of the meaning and use of *kya:*.

> For instance, example (15) shows that *kya:* can be embedded under the rogative predicate *pu:ch* 'ask', but cannot be embedded under the responsive predicate *ja:n* 'know'.
>
> (15) a. *Hindi-Urdu* (Bhatt & Dayal 2020)
>
> | | | | | | | | | |
> |---|---|---|---|---|---|---|---|---|
> | Ṭi:car=ne | Anu=se | pu:ch-a: | [ki | kya: | vo | ca:i | piyegi:]. |
> | teacher=ERG | Anu=from | ask-PFV | that | KYA | s/he | tea | drink.FUT.3FSG |
>
> 'The teacher asked Anu whether she would drink tea.'
>
> b.
>
> | | | | | | | | |
> |---|---|---|---|---|---|---|---|
> | *Anu | ja:n-ti: | hai | [ki | kya: | tum | ca:i | piyoge]. |
> | Anu.F | know.HAB.F | be.PRS.SG | that | KYA | you | tea | drink.FUT.2MPL |
>
> (Int.) 'Any knows whether you will drink tea.'
>
> (Gonzalez 2023: 8)

How to discuss data

Every time you include a linguistic example, you should introduce your example, present your example, and then discuss your example.

Not all examples necessarily need discussion before *and* after. If it's a very simple example, it may be enough simply to explain it all before you present it. Use your judgment and think about what your reader needs to

know to understand your example and how it fits into your argument, as well as when they need to know which information.

Introduce your example

Before you give an example, you should at least tell your reader that you're about to show them an example, referring to the example by number. In can be very disorienting if you're reading along and all of a sudden there's an example with no context. Even better, though, is if you tell your reader why you're about to show them an example and tell them what to look for. This is illustrated above with Gonzalez's introduction to her examples (1) and (2).

Present your example

After you've introduced the example, present the example. The example should be numbered, as discussed above, and **glossed**, as we will discuss in Section 4.7.

If the example is at all complex, guide your reader to which parts of the example they should be looking at. You can bold or underline the relevant parts. For example, Gonzalez bolded the question markers in her example (4), shown above. Ideally, you should explicitly state what you've bolded and underlined when you introduce the example. Instead of bolding or underlining, you can also describe in the text where your readers should be looking, before or after you present the example.

Always clearly indicate the language that the data is from. If the entire paper is about the same language, you can do this just once, in the introduction. But if you're discussing multiple languages, every single example should be labelled with the language it comes from. Some authors do this in the prose before the example, but it's even clearer if it's incorporated into the offset example itself. Two common ways of doing so are shown below in examples (1)-(2). In (1), the name of the language is right-aligned next to the example. In (2), the name of the language appears above the example. Whichever format you choose, you should be consistent throughout your paper.

(1) This is an example. [English]

(2) *English*
This is another example.

Discuss your example

After you have presented your example, you should discuss it. Do not leave it to your reader to analyze the data

— that's your job as the author. Describe the pattern you see in the data and explain why the pattern matters for your argument. And be specific! Mention specific words or morphemes from your example and why they matter.

Attributing your data

Whenever you use data, you must attribute your source.

- If you collected the data via introspection, this is the only time that you don't need to attribute your sources.
- If you collected the data via elicitation, you should thank your speakers in your acknowledgments section. It is also a good idea to discuss your methodology at some point within your paper.
- If your data comes from a corpus, the data collection methods should be explained at some point in your paper. If you use multiple corpora, each example should be marked with which corpus it came from.
- If your data comes from another published source, you should provide an in-text citation next to or under the example that indicates its source. Since an example is a kind of direct quote, you should include the page number of the original source. If the source from which you got the example is not itself the original source, you should cite *both* the original source and the source you got it from, in the format "[original source], as cited in [source you got it from]".

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1475#h5p-30

References and further resources

Sources for examples

Gonzalez, Aurore. 2023. Interrogative particles in polar questions: The view from Finnish and Turkish. *Glossa* 8(1): 1–47.

4.7. BECOMING A LINGUIST: GLOSSING SPOKEN LANGUAGE DATA

If the object language you are discussing differs from your metalanguage, you should gloss your examples. This means that you provide a morpheme-by-morpheme translation of your data. This is what enables us, as linguists, to analyze the structure of languages we ourselves do not speak.

Normally glosses are provided in a three- or four-line format.

In three-line glosses, the first line provides the object language data. The second line provides a morpheme-by-morpheme translation of the object language into the metalanguage. The third line provides a natural, idiomatic translation into the metalanguage. Gonzalez's example (4a) from the previous section is repeated below to illustrate.

(1) *Finnish*

Lähti-kö Mari?

left-PolQP Mari

'Did Mary leave?'

(Gonzalez 2023: 3)

A 4-line gloss is most often used when the phonological processes in the object language make it difficult to identify the morpheme boundaries. In these cases, the first line would provide the object language without modification and the second line would provide the **underlying form** of each morpheme. The last two lines are the same as in a three-line gloss.

(2) *Blackfoot*

Nitsiikaakaahsi'taki					aotoyaakihtsiniki			
nit-	iik-	aak-	yaahssi	-i'taki	a-	oto-	yáakihts	-iniki
1-	intend-	FUT-	be.good.VTI	-feel.emotion	DUR-	go.to.do-	go.to.bed	-SUBJ

'It will make me happy when you go to bed.'

(Déchaine and Wiltschko 2014: 74)

The first line: The object language data

If the object language of your paper is not written with the Latin alphabet but your metalanguage is, then it is convention to transliterate the object language into the Latin alphabet or transcribe it with the IPA. Many languages have a standardized transliteration system which you can use. For example, Mandarin is often transliterated in the system known as pinyin. If you wish to include the original orthography for any reason, you can use a 4-line gloss with the orthography as the first line and a transliteration or transcription in the second line.

Secondly, you must always mark the sentence for its acceptability. If the sentence has no acceptability mark, it means that it is fully acceptable. The most common acceptability marks are shown in Table 1.

Table 1. Acceptability markings

Name	Symbol	Meaning	Example
Asterisk	*	ungrammatical	*Cat the apple ate.
Question mark	?	weakly ungrammatical	?I have dived into the pool.
Percentage sign	%	variation in acceptability	%I eat meat anymore.
Number sign	#	semantically or pragmatically ill-formed	#My toothbrush ate an apple.

The most common acceptability mark is the asterisk (*), which means that it is ungrammatical, or, in other words, morphologically or syntactically ill-formed. It is also sometimes used when it is unclear why a sentence is unacceptable. The question mark (?) is used when a sentence feels weakly ungrammatical. In other words, it feels kind of off, but not fully ungrammatical. The percentage sign (%) is used when there is variation. The example in Table 1 uses positive *anymore*, which is a grammatical feature of some dialects of US American English and Irish English. Finally, the number sign (#), also sometimes called the hash or pound sign, is used for sentences that are semantically or pragmatically ill-formed. The example in Table 1 is semantically ill-formed because toothbrushes cannot eat.

Sometimes linguists will use multiple asterisks or question marks, or combinations of asterisks and question marks, to indicate degrees of unacceptability.

Sometimes linguists use parentheses to abbreviate multiple examples into one example. If a word is included in the example in parentheses, as in (3a), it means that the word in parentheses is optional. Thus, (3a) means that the sentence is acceptable either without the word *quickly,* as in (3b), or with the word *quickly*, as in (3c).

(3) a. Jenna ate the pickles (quickly).

b. Jenna ate the pickles.

c. Jenna ate the pickles quickly.

If a word is included in parentheses with an asterisk also inside the parentheses, as in (4a), that means that the word in parentheses may not be included in that position in the sentence but that the sentence is otherwise grammatical. Thus, the notation in (4a) means that the sentence in (4b), without *quickly*, is grammatical, but the sentence in (4c), with *quickly*, is ungrammatical.

(4) a. Jenna ate the (*quickly) pickles.
b. Jenna ate the pickles.
c. *Jenna ate the quickly pickles.

Finally, if a word is in parentheses with an asterisk outside of the parentheses, as in (5a), it means that the word in the parentheses is obligatory. In other words, the notation in (5a) means that a sentence with the word *ate*, as in (5b), is grammatical, but the same sentence, with *ate* omitted, as in (5c), is ungrammatical.

(5) a. Jenna *(ate) the pickles.
b. Jenna ate the pickles.
c. *Jenna the pickles.

The second line: Morpheme-by-morpheme glossing

In the second line of a three-line gloss, we include a morpheme-by-morpheme translation. The format for the second-line gloss is highly conventionalized. The most common conventions are summarized by the Leipzig Glossing Rules. If you continue on in linguistics, you should refer to the glossing rules when you read and write papers. This explanation is consistent with the Leipzig glossing rules.

Each word in the first line should be aligned with their translation in the second line. When they are not lined up, the examples are much harder to read, as shown by the difference between (6a) and (6b) below. You can line them up manually using spaces and tabs or you can use an invisible table, with each word in a different column.

(6) a. *Hindi-Urdu*

Ṭi:car=ne	Anu=se	pu:ch-a:	[ki	kya:	vo	ca:i	piyegi:].
teacher=ERG	Anu=from	ask-PFV	that	KYA	s/he	tea	drink.FUT.3FSG

'The teacher asked Anu whether she would drink tea.'

b. Ṭi:car=ne Anu=se pu:ch-a: [ki kya: vo ca:i piyegi:].
teacher=ERG Anu=from ask-PFV that KYA s/he tea drink.FUT.3FSG
'The teacher asked Anu whether she would drink tea.'

(Gonzalez 2023: 8)

In the second line, we put content words in lowercase letters, and we put grammatical markers in small caps. If small caps cannot be used for some reason, uppercase letters can be used instead.

Morphemes are separated out by hyphens (-) in both the first and second lines. Other markings may be optionally used for special kinds of morphemes (=, ~). If a morpheme cannot be translated by a single word, use periods (.) to separate the pieces of the gloss that all translate parts of the same morpheme.

Authors often use abbreviations for the grammatical markings in the gloss. The Leipzig glossing rules include a list of common abbreviations. If you use any abbreviations that differ from Lepizig's, you need to include a key or glossary of your abbreviations somewhere in your paper. Some common places authors put their abbreviation list include a footnote at the beginning of the paper, a footnote on the first example in the paper, a note at the end of the paper, or an appendix in a book. All of the abbreviations used in this textbook are included in a back matter section.

The third line: A natural translation

The third line of a gloss should provide a natural translation into the metalanguage. The third line tells you what the example means.

Sometimes an author will also include a literal translation with the abbreviation "lit.".

If an example is ungrammatical, it might not really have a meaning. In this case, an author might indicate that the meaning is "intended".

Finally, if you have multiple examples in a row that all mean the same thing, sometimes the author will only include the translation once, after either the first or last example.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1594#h5p-31

References and further resources

Reference materials

📑 Max Planck Institute for Evolutionary Anthropology Department of Linguistics. 2015. Leipzig Glossing Rules. https://www.eva.mpg.de/lingua/resources/glossing-rules.php

Sources for examples

Déchaine, Rose-Marie and Martina Wiltschko. 2014. Micro-variation in agreement, clause-typing and finiteness: Comparative evidence from Blackfoot and Plains Cree. In J. Randolph Valentine and Monica Macaulay, eds. *Papers of the 42nd Algonquian Conference*. SUNY Press. 69101.

Gonzalez, Aurore. 2023. Interrogative particles in polar questions: The view from Finnish and Turkish. *Glossa* 8(1): 1–47.

CHAPTER 4 PRACTICE EXERCISES

Data analysis

Exercise 1. Glossing

Beginner [4.7]

Imagine that you are conducting elicitations in Nigeria and you have just collected the following Yoruba data in (1).[1] You have collected the pronoun paradigm and several sentences. You've also looked up all the words in the dictionary, which are listed in the table below. Gloss the following three sentences according to linguistic convention to prepare them for inclusion in a paper.

Note 1: *Bolu*, *Wale,* and *Shade* are proper names.

Note 2: Yoruba is a tonal language, but the tones are omitted in these examples.

(1) a. *Yoruba*
a ti jo
'We danced'

b. a ti jo
'We danced'

c. a ti jo
'We danced'

1. The data in this exercise was provided by Tolani Akinlade.

Table 1. Mini Yoruba dictionary

Yoruba data	English translation
adiyẹ	'chicken'
ati	'and'
gbọ	'hear'
je	'eat'
jinan	'ready' (for food)
jo	'dance'
ko	'not'
n	action in progress
ti	past tense marker

Table 2. Yoruba pronoun paradigm

	Subject pronouns	Object pronouns
1sg	mo	mi
2sg	iwọ	e
3sg	o	Ø
1pl	a	wa
2pl	eyin	yin
3pl	wọn	wọn

Communication and study skills

Exercise 2. Writing about data

Intermediate [4.6]

In Chapter 2, Exercise 2, you developed some data to demonstrate the behaviour of *our* and *ours*. Use the introduce, present, discuss format described in Section 4.6 to describe your data.

Research and application

Exercise 3. Source analysis

Intermediate [4.7, 4.8]

Look up an academic article, a book chapter in an academic book, a dissertation, and/or a single-authored academic book in morphology or syntax. Look at how the examples are formatted and answer the following questions.

a. Find the abbreviations list. Where is it?

b. What abbreviations are common/uncommon?

c. Does the source use three or four line glossing?

d. What similarities or differences do you notice compared to the Leipzig conventions?

e. Do they restart the example numbering in every chapter/section?

f. Can you tell which examples they collected themselves vs. got from other sources? How can you tell?

g. How do they cite examples from other sources?

h. If there are examples from multiple languages, how do they indicate which examples are in which language? Is it effective or confusing?

i. Are syntax trees numbered like examples or treated like figures?

Language Journal

At the end of Chapter 1, we introduced the Language Journal. You should have picked a language to build a profile for as you work through this textbook. You can add to your profile with these questions.

Exercise 4. Language modality

Beginner [4.1]

What is the modality of your language?

Exercise 5. Dialects

Intermediate [4.3]

What are the different dialects of your language? Is your language mutually intelligible with another language?

Exercise 6. Glossing

Intermediate [4.7, 4.8]

Find three sentences from your language and create properly formatted glosses for them.

5. DECOLONIZING LINGUISTICS

[placeholder]

5.1. LANGUAGE AND COMMUNITY

As social scientists, we have a responsibility to think critically about where our data comes from—not only so that we can ensure it is reliable and that the collection methods fit our research objectives, but also to ensure that our data has been collected in an ethical manner. We are studying human language, which means that we are collecting our data from people and from communities, and we need to consider the impact of our research on these people and communities. In this chapter, we will look at some of the ethical issues involving using and collecting data, especially data from Indigenous languages, within the context of colonialism.

In the past, linguistics as a field has tended to separate the data we collect from the communities that use them, to treat the data as abstract ideas that belong to no one, or perhaps to the linguist who records them. However, such an approach can cause harm; the languages we study often have emotional, cultural, historical, and spiritual significance to the communities using them. As Christine Mallinson writes, "language and society are inseparable, language is a cultural artifact that belongs to those people and communities who use it, and addressing power dynamics and inequalities surrounding language is central to social justice, equity, and inclusion" (Charity Hudley, Mallinson & Bucholtz 2024b: xiv).

Likewise, we, as researchers, bring our own history, culture, and context to bear as we do research. As Clemons and Lawrence (2020: e259) state, "research subjectivities affect the kinds of questions that get asked as well as the interpretation of findings."

> Emotionless, passionless, abstract, intellectual, academic research is a goddamn lie, it does not exist. It is a lie to ourselves and a lie to other people. Humans—feeling, living, breathing, thinking humans—do research. When we try to cut ourselves off at the neck and pretend an objectivity that does not exist in the human world, we become dangerous, to ourselves first, and then to the people around us.
>
> Eber Hampton (1995: 52, as cited in Wilson 2008: 100-101)

As such, I will begin this chapter by introducing myself and my position with regards to linguistic research and Indigenous communities.

Positionality statement

I, the author of this textbook, am a white settler person of mixed Germanic and Celtic descent. Most of my ancestors settled in Canada between 5 and 8 generations ago in the 1700s and 1800s. Most bought, were given, or rented land to farm. Some were the first white settlers in their area, such as David and Mary Thompson on my paternal grandmother's side, who came from Scotland and were the first to settle in Scarborough, Ontario. My ancestors settled on land covered by many different treaties in Ontario, which were signed on their behalf, including the Williams treaties, the Toronto Purchase, the Haldimand Tract and Simcoe Patent, the Huron Tract, and the Robinson-Huron Treaty. There are many ethical issues and disputes regarding these treaties.

As I write this section, I cannot separate myself from my family's history. What I can do, however, is move forward in a posture of learning, one that acknowledges this history and my place in it and that seeks to move forward in a better way. This learning journey will continue for the remainder of my life. This section represents some of the things I have learned so far, from my perspective as a settler linguist who would like to honour the people who speak the languages I study as best as I can.

As I wrote this section, I tried to base it on the words and shared experiences of Indigenous scholars and community members. However, the very act of collecting and summarizing these views colours them with my perspective. If at all possible, you should learn about the experiences of Indigenous people directly from them. Check out the sources marked with a ➡ in the references sections throughout this chapter for some suggestions.

What is an Indigenous language?

In brief, an Indigenous language is a language used by or associated with an Indigenous people group. Often, Indigenous languages have historical, cultural, and/or spiritual significance for the Indigenous group(s) that use them.

In its simplest meaning, *indigenous* is simply an adjective describing originating in a particular place. Under this definition, all languages are indigenous to somewhere, since all languages arose in a particular community in a particular place. More often, though, the term *indigenous* is used in a more specific manner. However, as there are more than 5,000 distinct Indigenous groups in the world (according to the United Nations),

it is difficult to provide a precise definition. Instead of providing a definition of indigeneity, the United Nations Declaration on the Rights of Indigenous Peoples states that Indigenous peoples have a right to self-identify as Indigenous. This can lead to tricky situations, though, when groups with dubious claims to Indigeneity demand political Indigenous status. However, some commonalities between Indigenous groups can be identified: continued occupation of a land since before colonization, distinct cultures and political systems, and the desire to maintain institutions distinct from the majority population.

> Indigenous peoples have in common a historical continuity with a given region prior to colonization and a strong link to their lands. They maintain, at least in part, distinct social, economic and political systems. They have distinct languages, cultures, beliefs and knowledge systems. They are determined to maintain and develop their identity and distinct institutions and they form a non-dominant sector of society.
>
> -United Nations

Indigenous peoples (and their languages) are found in every region of the world. The majority of the world's languages are used by Indigenous peoples.

Indigenous peoples challenge!

How many Indigenous people groups are you aware of? Try to name as many Indigenous people groups as you can from memory, such as the following:

- The Indigenous people groups of where you currently live
- The Indigenous people groups of the places you grew up
- The Indigenous people groups of where your ancestors are from
- An Indigenous people group from every continent of the world
- An Indigenous people group for every letter of the alphabet
- An Indigenous people group from every country of the world

Consider reading about the Indigenous people groups from regions that are especially relevant to you! Many of you will be surprised by what you find.

Key takeaways

- Language is inseparable from the communities that use them.
- As linguists, we have an ethical obligation to ensure that the way we collect, store, and use language data does no harm to the communities who use the languages our data is from.
- It is impossible to stay completely objective; everyone has conscious and subconscious biases.
- Indigenous people groups typically have the following characteristics: they have occupied their land since before colonization, are a minority population, and desire to maintain distinct cultural and political institutions.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1473#h5p-41

References and further resources

The references in this section that have at least one Indigenous author are indicated with an ➡.

For a general audience

🔍 Oxford, Will. 2019. Indigenous Languages in Canada. Canadian Language Museum. https://languagemuseum.ca/wp-content/uploads/2022/12/indigenous_languages_in_canada_0.pdf

United Nations. No date. Indigenous peoples. Accessed May 16, 2024. https://www.un.org/en/fight-racism/vulnerable-groups/indigenous-peoples

United Nations. 2007. United Nations declaration on the rights of Indigenous peoples. https://social.desa.un.org/issues/indigenous-peoples/united-nations-declaration-on-the-rights-of-indigenous-peoples

Academic sources

Charity Hudley, Anne, Christine Mallinson, and Mary Bucholtz, eds. 2024a. *Decolonizing Linguistics*. Oxford: Oxford UP. https://global.oup.com/academic/product/decolonizing-linguistics-9780197755259?cc=ca&lang=en&#

Charity Hudley, Anne, Christine Mallinson, and Mary Bucholtz, eds. 2024b. *Inclusion in Linguistics*. Oxford: Oxford UP. https://global.oup.com/academic/product/inclusion-in-linguistics-9780197755303?cc=ca&lang=en&

Clemons, Aris, and Anna Lawrence. 2020. Beyond position statements on race: Fostering an ethos of antiracist scholarship in linguistic research (Response to Charity Hudley et al.). *Language* 96 (4): e254-e267.

➡ 🔍 Wilson, Shawn. 2008. *Research is ceremony: Indigenous research methods.* Halifax: Fernwood Publishing.

➡ 📑 Younging, Gregory. 2018. Elements of Indigenous Style. Brush Education Inc.

For instructors

Singerman, Adam. 2025. Talking about language endangerment and Indigenous languages in the classroom: Some *do*s and *don't*s I have learned through fieldwork in the Brazilian Amazon. In Lauren Clemens, Vera Gribanova, and Gregory Scontras (ed). *Syntax in uncharted territories: Essays in honor of Maria Polinsky*. Open Access Publications from the University of California. 525-545. https://escholarship.org/uc/item/3081×700#main

5.2. COLONIALISM AND DECOLONIZATION

The Canadian Encyclopedia defines colonialism as "the process through which a foreign people establish control over a territory and, if applicable, its Indigenous peoples" (Hale 2023). Colonial control can be exercised in a variety of ways, and often includes control of a people's language.

Shoemaker (2015) describes 12 different forms of colonialism, two of which are relevant to us: settler colonialism and extractive colonialism.

Settler colonialism

> **Settler Colonialism.** Large numbers of settlers claim land and become the majority. Employing a "logic of elimination," as Patrick Wolfe put it in the *American Historical Review*, they attempt to engineer the disappearance of the original inhabitants everywhere except in nostalgia.
>
> -Shoemaker 2015

Settler colonialism was the major form of colonialism deployed in Canada and many other nations. Throughout much of Canadian history, "desirable" settlers were actively recruited, while "undesirable" settlers were discouraged and Indigenous peoples were killed, displaced, or assimilated by force in order to clear the land for settlers.

One major tool of forced assimilation used in Canada and elsewhere is the Indian Residential School System. Children were forcibly removed from their families and taken to boarding schools where they were horribly mistreated, as well as forbidden from speaking their languages and practicing their cultures. The stated aim of these residential schools was assimilation, as stated openly by Prime Minister John A. Macdonald in parliament.

> When the school is on the Reserve the child lives with its parents, who are savages; he is

> surrounded by savages, and though he may learn to read and write his habits and training and mode of thought are Indian. He is simply a savage who can read and write ... Indian children should be withdrawn as much as possible from the parental influence, and the only way to do that would be to put them in central training industrial schools where they will acquire the habits, modes and thought of white men.
>
> -John A. Macdonald, House of Commons, 9 May 1883
> as cited by Starblanket 2018: 179

The Truth and Reconciliation Commission classifies these practices as a "cultural genocide" (2015). The act of forcibly transferring children from one group to another group is also classified as an act of genocide according to the United Nations (1948).

The last residential school didn't close until 1996 and communities are still healing and rebuilding. Many community elders and language users are residential school survivors. Other forms of forced assimilation are still in force, as well, such as the placement of Indigenous children in foster care with non-Indigenous families.

When we do linguistic research with communities that have experienced forced assimilation, we need to keep this in mind. There may be only a few language users left, whose time and energy belongs to the community first, not to outsider linguists. Because language use was punished in the residential schools, survivors may have trauma associated with the language. Others may have purposefully not taught the language to their children in order to protect them. Community members may be rightfully upset if settlers and other outsiders seem to have more access to their own language than they do.

Extractive colonialism

> **Extractive Colonialism.** All the colonizers want is a raw material found in a particular locale: beaver fur, buffalo hides, gold, guano, sandalwood. The desire for natural history specimens and ethnographic artifacts could also be considered extractive colonialism. A slash-and-burn operation, extractive colonialism does not necessarily entail permanent occupation, but it often seems to follow. Extractive colonizers might destroy or push away indigenous inhabitants to

> access resources but more typically depend upon native diplomatic mediation, environmental knowledge, and labor.
>
> -Shoemaker 2015

Although aspects of settler colonialism are still at play in Canada, we have in many cases transitioned to an extractive colonial society.

Extractive colonialism in Manitoba

Extractive colonialism is alive in well in Canada. It is part of our everyday lives, even though it may be invisible to many of us. Let's look more closely at two instances of extractive colonialism that is ongoing in Manitoba: Winnipeg's water supply and Manitoba Hydro.

Winnipeg aqueduct *coming soon*. Read *Aqueduct* by Adele Perry or watch Freedom Road to learn more.

Starting in the 60s, Manitoba Hydro has diverted a major river and built several dams in Northern Manitoba. It is continuing to build expensive dams, even though we are already generating more power than we need. The initial construction of these dams has a huge ecological impact, involving clear cutting forests, moving entire communities, and building quarries and roads, in addition to the dam itself. The ongoing impact of Manitoba Hydro's management of water levels also has a huge impact. Large fluctuations in water level have swept away entire beaches, downed millions of trees, killed wildlife, and have destroyed the local fishing population and fishing-based economy. The rapid erosion also introduces mercury into the water. The ongoing erosion makes it dangerous to travel along the shorelines, while the trees in the water and flooded islands make it dangerous to boat. The town of South Indian Lake used to have the second largest whitefish fishery in North America, but now 85% of the community relies on social assistance (Wilt 2016). Meanwhile, local hydro employees receive discount heating while Indigenous populations pay a "rural"

premium—which, together with the destruction of the fishing industry, creates a wealth gap largely along racial lines. Meanwhile, Manitoba Hydro continues to request and receive augmented flow licenses every year, allowing them to manipulate water levels more than the initial agreements. Read "How green energy has hurt First Nations in the North" by James Wilt or watch Green Green Water to learn more.

Extractive colonialism doesn't only apply to physical resources like fur and oil, though; it also applies to knowledge. When we do research on a language, we need to be careful to protect the well-being of the community the data comes from, keeping in mind that we are guests and that the language belongs to the community, not to the researchers. We also need to remember that our research depends on community members' knowledge, labour, and expertise, and that we would not be able to do our research without them.

Decolonization

Decolonization, on the other hand, refers to the removal of colonial elements. This means that we have to first recognize the ways that our systems are colonial in nature and then restructure them to remove those colonial structures. In the next few sections, we will first learn about the colonial history of linguistics and then we will look at models for how to move forward in a good way.

Key takeaways

- Settler colonialism is when settlers attempt to become the majority population group of a land, displacing or assimilating the original population.
- The Indian Residential School System in Canada was a form of genocide used to force the assimilation of Indigenous peoples.
- In Indian Residential Schools, children were typically forbidden from speaking their languages, which still has far-reaching impact on Indigenous language use today.
- Extractive colonialism is when settlers collect resources from the settled land, often using Indigenous labour or expertise, prioritizing their desire for resources over the well-being of Indigenous communities.

- Research conducted on Indigenous peoples can be a form of extractive colonialism if it is not done in a good way.
- Decolonization is the process of identifying colonial systems and removing them from our society.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=3411#h5p-42

References and further resources

The references in this section that have at least one Indigenous author are indicated with an ➡.

For a general audience

Hale, Karl. 2023. Colonialism in Canada. *The Canadian Encyclopedia.* https://www.thecanadianencyclopedia.ca/en/article/colonialism-in-canada

➡ McLeod, Angelina. 2019. Freedom Road. National Film Board of Canada. https://www.nfb.ca/series/freedom-road-series/

Mikkelson, Dawn, and Jamie A. Lee (Directors). 2005. Green green water. Culture unplugged. https://www.cultureunplugged.com/play/6661/Green-Green-Water

Perry, Adele. 2016. *Aqueduct: Colonialism, resources, and the histories we remember.* Winnipeg: ARP Books.

➡ Shoemaker, Nancy. 2015. A Typology of Colonialism. Perspectives on History.

Truth and Reconciliation Commission of Canada. 2015. Honouring the truth, reconciling for the future: Summary of the final report of the Truth and Reconciliation Commission of Canada. https://ehprnh2mwo3.exactdn.com/wp-content/uploads/2021/01/Executive_Summary_English_Web.pdf

United Nations. 1948. Convention on the prevention and punishment of the crime of genocide. https://www.un.org/en/genocideprevention/genocide-convention.shtml

➡ Wa Ni Ska Tan Alliance. 2024. Augmented flow: License to destroy documentary. https://www.youtube.com/watch?v=ZdrEwEt-V7o

Wilt, James. 19 October 2016. How green energy has hurt First Nations in the North. *Vice.* https://www.vice.com/en/article/4w58mq/how-green-energy-has-hurt-first-nations-in-the-north

Academic sources

Carleton, Sean. 2021. 'I don't need any more education': Senator Lynn Beyak, residential school denialism, and attacks on truth and reconciliation in Canada. *Settler Colonial Studies* 11 (4): 466-486. https://www.tandfonline.com/doi/full/10.1080/2201473X.2021.1935574

➡ Fontaine, Lorena. 2017. Redress for linguicide: Residential schools and assimilation in Canada. *British Journal of Canadian Studies* 30(2): 183–204.

MacDonald, David & Graham Hudson. 2012. The Genocide question and Indian residential schools in Canada. *Canadian Journal of Political Science* 45: 427–449.

➡ Starblanket, Tamara. 2017. 'Kill the Indian in the child': Genocide in international law. In*Indigenous peoples as subjects of international law*, ed. Irene Watson. Abingdon, Oxon: Routledge. 171-200.

CHAPTER 5 PRACTICE EXERCISES

Data analysis

Exercise 1

Classify each of the following scenarios according to its model of community engagement from Czaykowska-Higgins (2009) The options are: (a) linguist-focused model, (b) advocacy research, (c) empowering research, or (d) community-based language research.

a. Joshil Patel has been writing his PhD dissertation on the Dene language. He has lived in one particular Dene community for months at a time as he has collected his data, during which time the local school has been implementing a Dene language immersion program, but they are short on language resources. Joshil has helped the community apply to government grants for the language immersion program. As part of his research, Joshil recorded several stories told by elders in Dene. Joshil types up the stories, prints and binds them, and gifts them to the school as a thank you for their contribution to his dissertation research.

b. Sam Klein just got hired as a professor at the University of Manitoba. Once they got settled in, Sam researches which local Indigenous organizations are involved in language revitalization. Sam reaches out to several of those organizations and offers their help and expertise. One of the organizations asks for help with their ongoing dictionary project. Sam begins by attending several meetings as a spectator, and eventually volunteers to join the committee working on the morphological analysis of the complex verbs of the language. Sam is using their status and experience as a university professor to get the committee's discoveries published in an academic journal, with several community members as co-authors. One of the volunteers on the project is a bright high school student. Sam helps the high school student apply to the linguistics program at UM.

c. Jane Clark is very interested in wordhood and in developing a cross-linguistic definition of words. She wants to test her hypothesis on a polysynthetic language. She reads up on Inuktitut in the university library. She applies to the university ethics board for conducting an experiment on speakers of Inuktitut. She then travels to an Inuit community and recruits Inuktitut speakers to participate in her experiment. She pays participants $30 for completing her 30-minute experiment.

d. Yuxi Liu and Zihao Wu are writing a grammar of the Taiwanese language Amis. Yuxi and Zihao have visited the Amis communities frequently as they work on their project. They had several community members give comments on the outline and each draft of their grammar. Based on feedback

from the community, Yuzi and Zihao have included a glossary of technical linguistic terms so that the grammar can be used by community members and not just by other linguists. Yuzi and Zihao will thank the community members in the acknowledgements section of their book.

Communication and study skills

Exercise 2

Research a language and write an introduction to that language that emphasizes the relationship between the language and its speakers and humanizes them.

Exercise 3

Read the Truth and Reconciliation Commission's Calls to Action and/or the Missing and Murdered Indigenous Women and Girls Calls for Justice. Choose either one call to action or one call to justice and discuss how it is relevant to linguistic research. How can you create a reciprocal and respectful relationship with the communities whose languages we are studying?

Research and application

Exercise 4

Choose a place that is significant to you, such as where you live, where you grew up, or where your ancestors are from. What languages are/were spoken there? Spend extra time looking for languages and varieties that are not well-known. Are there Indigenous or endangered languages that you've never heard of before? How has colonialism and history affected the language landscape of this place?

Language journal

At the end of Chapter 1, we introduced the Language Journal. You should have picked a language to build a profile for as you work through this textbook. You can add to your profile

with these questions.

Exercise 5

Is your language an Indigenous language? How do you know?

Exercise 6

Research the history of linguistic research on your language. In what ways were colonial practices in use?

Exercise 7

Write a profile about one of the communities that use your language.

6. KINDS OF MORPHEMES AND MORPHOLOGICAL PROCESSES

[Placeholder.]

6.1. THE BUILDING BLOCKS OF MORPHOLOGY

A word can be broken up into several **morphemes**, which, as we learned in Section 1.1, are the smallest meaningful units of language. There is more than one kind of morpheme. In this section, we will learn about roots and affixes and about free and bound morphemes.

Roots, stems, and affixes

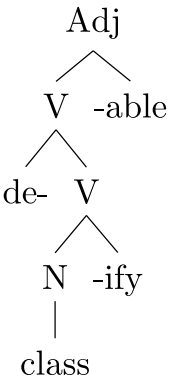
A **root** is the core morpheme of a word, while **affixes** are the morphemes that are attached to it. Some examples are shown in (1).

(1)		**Word**	**Root**	**Affixes**
	a.	*librarian*	*library*	*-an*
	b.	*unthinkable*	*think*	*un-*, *-able*
	c.	*workers*	*work*	*-er*, *-s*
	d.	*dismissiveness*	*miss*	*dis-*, *-ive*, *-ness*
	e.	*declassifiable*	*class*	*de-*, *-ify*, *-able*

Words are built by starting with the root and adding one morpheme at a time. The piece of the word that hosts an affix at each stage is called the **stem** or the **base**.

Sometimes students mix up roots and stems. A root is always a single morpheme, but a stem may be complex, consisting of multiple morphemes. The root of a word does not change as you add more affixes, but the stem is different for each affix. Let's consider the word *declassifiable* as an example, which contains four morphemes: *de-, class, -ify,* and *-able,* with the hierarchical structure shown in (2).

(2) a.



b. [Adj [V [de-] [V [N[class]] [-ify]]] [-able]]

The root of the word *declassifiable* is *class*. The first affix that is attached to the root is *-ify*. The stem of *-ify* is the root *class*. Next, *de-* is attached; the stem of *de-* is the complex form *classify*. Finally, *-able* is affixed; the stem of *-able* is the complex form *declassify*.

As you can see, complex words like *declassifiable* are built from the inside out, starting with the root. Affixes are added to either edge of the root, one at a time. We will discuss more about how to determine which order affixes are attached in Chapter 11.

Free and bound morphemes

Another way we can classify morphemes is based on whether they are **free** or **bound**. A free morpheme can stand on its own, while a bound morpheme must always be attached to another morpheme. Some examples of free morphemes in English include *library, think, blue, class,* and *couch*. Some examples of bound morphemes in English include *-s*, *-ify*, *-ed*, *in-*, *dis–* and *-ish*.

We use hyphens to show whether a morpheme is free or bound. Free morphemes are not marked with hyphens while bound morphemes are marked with hyphens. The hyphens are always placed on the side of the affix that attaches to the stem.

Affixes should always have a hyphen. In some languages, like English, most roots are free and do not need a hyphen. In other languages, most roots are bound, and do need a hyphen. For example, in Spanish, most nouns are marked with a gender suffix. Since the roots cannot occur on their own without the gender suffix, these would be considered bound roots. Some examples, shown in (3), are the feminine noun *casa* 'house' which consists of the bound root *cas-* and the feminine suffix *-a* or the masculine noun *cerezo* 'cherry' which consists of the bound root *cerez-* and the masculine suffix *-o*.

(3)	a.	*Spanish*	
		cas	-a
		house	-F
		'house'	

	b.	cerez	-o
		cherry	-M
		'cherry'	

Key takeaways

- Roots are the core morpheme of a word. Affixes are morphemes that are attached to the root.
- The piece of a word that hosts an affix is called a stem or base.
- Morphemes that can stand on their own are called free morphemes and morphemes that cannot stand on their own are called bound morphemes.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=2849#h5p-40

6.2. AFFIXES

Affixes are bound morphemes that are attached to a stem, usually linearly. There are 5 kinds, categorized based on where they are attached: prefixes, suffixes, infixes, circumfixes, and suprafixes.

Prefixes

Prefixes are attached to the beginning of a stem, as shown in (1). In (1a), the prefix *re-* is attached to the stem *work* to mean 'work again.' In (1b), the prefix *un-* is attached to the stem *lock* to mean 'reverse the action of locking.'

(1) a. re-work
b. un-lock

There can be more than one prefix in the same word. For example, in (2), the prefix *un-* is attached to the stem *lock* to mean 'reverse the action of locking,' and then the prefix *re-* is attached to the stem *unlock* to mean 'reverse the action of locking again.' These are both prefixes, since they are attached to the beginning of the stem, even if it is no longer at the beginning of the stem after other affixes have been added.

(2) re-un-lock

Prefixes are marked with a hyphen at the end, as with *re-* and *un-,* to show that they are bound and that they must be attached to a stem on their right edge.

Suffixes

Suffixes are attached to the end of a stem, as shown in (3). In (3a), the suffix *-ity* is attached to the stem *equal* to mean 'the state of being equal'. In (3b), the suffix *-s* is attached to the stem *cat* to mean 'multiple cats.'

(3) a. equal-ity
b. cat-s

Just as with prefixes, there can be more than one suffix, as shown in (4). In (4), the suffix *-ity* is attached to the stem *equal* to form *equality* meaning 'the state of being equal,' and then the suffix *-s* is attached to the stem *equality* to mean 'multiple states of being equal.' These are both suffixes, since they are attached to the end of the stem, even if *-ity* is no longer at the end of the stem in the final form of the word. Note that *-ity* changes spelling to *-itie* when the plural morpheme *-s* is added, but it is the same morpheme (it is pronounced the same and it has the same meaning).

(4) equal-itie-s

Suffixes are marked with a hyphen at the beginning, as with *-ity* and *-s*, to show that they are a bound morpheme that attaches to the stem on its left side.

Infixes

Infixes are attached inside a stem, as shown in the Hoava data in (5). Hoava is an Austronesian language spoken in the Solomon Islands. In Hoava, the infix *-in-* is inserted before the first vowel of the stem, turning verbs and adjectives into nouns. In (5a)-(5d), this results in an infix. However, in vowel-initial words like (5e)-(5f), this looks like a prefix, because the position before the first vowel of the word happens to also be the edge of the word. We still call it an infix, though, because the overall pattern that can explain the data as a whole is an infixation pattern.

(5) *Hoava*

a.	to	'alive'	t<in>o	'life'
b.	hiva	'want'	h<in>iva	'wishes'
c.	va-bobe	'cause to be full'	v<in>a-bobe	'filled object'
d.	ta-poni	'be given'	t<in>a-poni	'gift'
e.	asa	'grate'	<in>asa	'pudding of grated cassava'
f.	edo	'happy'	<in>edo	'happiness'

(original source, Davis 2003, adapted from Kalin 2022: 644)

Note also that the infixation pattern applies to the stem, not the root. This means that the infix might sometimes appear inserted inside the root, as in (5a)-(5b), or might sometimes apper in an affix that has already been attached to the root, as in (5c)-(5d), where it appears inside the causative prefix and the passive prefix, respectively.

Many students confuse infixes with prefixes or suffixes that are not at the word edge, like *un-* in (2) or *–ity* in

(4). Infixes are fairly rare. If you think you have found an infix in your analysis of a language, you should look closely to see if you really can't analyze the infix as a a prefix or a suffix that is not at the word edge.

If you're confused, think of a stack of books as a metaphor. Figure 1 shows a stack of five books. We can think of the middle book as the root and the books on top of it and under it as the prefixes and suffixes. The second book from the top is not at the top or the bottom of the pile, but it is also not inside of another book. It is in between two books. Likewise, prefixes and suffixes might be in between two affixes, instead of at a word edge, but they will not be *inside* of another affix.



Figure 1. A stack of five books. Photo by Julie Doner.

In Figure 2, on the other hand, I've added a sixth book, in the middle of the second book from the top. This book is *inside* another book. It is like an infix, which appears inside of another morpheme. Likewise, in (5a), the nominalizing morpheme *-in-* appears inside of the morpheme *to* 'alive' in the word *tino* 'life.' Neither the sequence /t/ nor /o/ on their own are morphemes. Instead, the morpheme *to* is interrupted.



Figure 2. A stack of books with one book inside of another book. Photo by Julie Doner.

In English, infixation only occurs with expletives, as in (6), or in language play, as with Homer Simpson style *-ma-* infixation in (7) or Hip Hop style *-iz-* infixation in (8). All three of these infixes occur before the primary stressed syllable of the word.

(6)	a.	important	im<bloody>portant
	b.	fantastic	fan<fucking>tastic
	c.	Kalamazoo	Kala<goddamn>mazoo

(McCarthy 1982, as cited in Yu 2007: 1)

(7)	a.	saxophone	saxo<ma>phone
	b.	violin	vio<ma>lin
	c.	Michelangelo	Micha<ma>langelo

(Yu 2004, as cited in Yu 2007: 1)

(8)	a.	house	h<iz>ouse
	b.	soldiers	s<iz>oldiers
	c.	ahead	ah<iz>ead

(Viau 2002, as cited in Yu 2007: 1)

When describing infixes, it is important to identify and describe the pattern of where they get inserted, as well as their form and meaning. To identify the pattern, you may need to refer to the edge of the word (either the beginning or end of the word) or to the syllable or prosodic structure of the word. Some examples of where an infix might be inserted include:

- Before/after the first/last consonant of the stem
- Before/after the first/last vowel of the stem
- Before/after the first/last syllable of the stem
- Before/after the primary stress of the stem

When mentioning an infix in isolation, it should be written with a hyphen on each side, as with *-in-, -ma-,* and *-iz-,* to show that it is a bound morpheme that is inserted inside of the stem. When a word containing an infix is segmented, as in examples (5)-(8), the infix should be marked with angle brackets <> on either side. This shows that it is an infix and not a string of suffixes or prefixes.

Infixation is a little bit controversial. Some linguists, such as Kalin (2022), argue that infixes are really prefixes or suffixes that have been displaced.

Circumfixes

Circumfixes are affixes that appear in two pieces, on either side of the stem. As with infixes, circumfixes are rare. If you propose a circumfix, you need to have a compelling reason why it should be analyzed as one circumfix instead of two separate morphemes. The data in (9) illustrates a morphological pattern that can be analyzed as a circumfix.

In Tagalog, an Austronesian language of the Philippines, the circumfix *ka- -an* can be affixed to a noun to mean 'a group of X', as shown in (9). Neither the first nor second part of the circumfix can appear without the other. In other words, there is no word *kaintsik* or *intsikan*. Instead, both morphemes must always appear together.

(9) *Tagalog*

a.	intsik	'Chinese person'	ka-intsik-an	'the Chinese'
b.	pulo	'island'	ka-pulu-an	'archipelago'
c.	tagalog	'Tagalog person'	ka-tagalog-an	'the Tagalogs'

(Lieber 2016: 88)

Circumfixes are written in the form X- -Y, where X is the part of the circumfix that goes at the beginning

of the stem and Y is the part of the circumfix that goes at the end of the stem. This shows that it is a bound morpheme and that it is attached around the stem.

Circumfixes can be considered a special instance of **parasynthesis**, where one meaning is signaled simultaneously by more than one morphological process, since it is encoded simultaneously by both a prefix and a suffix.

Suprafixes

Finally, **suprafixes** are morphemes that are encoded in the **prosodic** structure of a word. Sometimes morphological distinctions in prosody are analyzed as suprafixes and included as a kind of affix, or sometimes they are considered to be a form of **non-concatenative morphology**, which will we discuss shortly in Section 6.4.

Suprafixes can create contrast in meaning through pitch, length, or stress. An example of a tone suprafix is shown in (10) for the language Tsuut'ina, a Dene language of Alberta. In Tsuut'ina, a high tone is used to mark an incomplete action while a low tone is used to marked a complete action.

(10)		**incomplete action**	**complete action**	**gloss**
	a.	-k'ós	-k'òs	'to throw'
	b.	-tšú?	-tšù?	'to seize'

(Cook 1971: 169)

An example of a length suprafix can be found in Estonian, a Uralic language spoken in Estonia, in Northern Europe, as shown in (11). The **genitive case** is used to mark possession on nouns, while the **partitive** case is used to mark partialness, similar to the English word *some*. For some Estonian nouns, the only distinction between genitive case is the length of a syllable. In (11a), the genitive form of vodka *viina*, has a long vowel. In the partitive case, the syllable is lengthened, creating an overlong vowel, *vii:na*. In (11b), the long consonant in the genitive *linna* 'city' becomes overlong in the partitive *linn:a.* Finally, in (11c), the diphthong in the genitive *laulu* 'song' becomes long in the partitive *lau:lu*.

(11)		**genitive case**	**partitive case**	**gloss**
	a.	viina	vii:na	'vodka'
	b.	linna	lin:na	'city'
	c.	laulu	lau:lu	'song'

(Spahr 2016: 152)

Some English words show a stress shift between noun-verb pairs, where the noun has initial stress and the verb has final stress, as shown in (12). This could be considered a kind of suprafix affecting stress.

(12)		verb form	verb meaning	noun form	noun meaning
	a.	*re-CORD*	'to convert sound or text into a permanent form'	*RE-cord*	'a log of past happenings'
	b.	*per-MIT*	'to allow'	*PER-mit*	'a document recording authorization'
	c.	*re-BEL*	'to engage in an act of defiance'	*RE-bel*	'someone who regularly engages in acts of defiance'
	d.	*pro-GRESS*	'to move forward or onward'	*PRO-gress*	'movement forwards or onwards'
	e.	*pre-SENT*	'to give a gift or award publicly'	*PRE-sent*	'a gift'

Tonal and length suprafixes can be written with a hyphen preceding a placeholder V (for vowel) or σ (Greek letter *sigma*, for syllable) and tone or length markers, as shown in (13). There are many different markers used to describe tone, depending on preference and the number of tones in use in the language in question; any of these notations can be used, but be sure to explain your notation if you are not using a well-known system. (For an overview of notation systems for tone, see Section 3.12 of *Essentials of Linguistics, 2nd edition*). Suprafixes are usually marked with a preceding hyphen, similar to suffixes, to show that it is a bound morpheme and because suffixes are more common cross-linguistically than prefixes. Similar to infixes, this notation should be supplemented with a description of which vowel or syllable it applies to. There is not really a good notation to use for suprafixes affecting stress; these patterns should be described in words.

(13)		notation	meaning
	a.	-V˧	a vowel with a mid-high tone
	b.	-σ́	a syllable with a high tone
	c.	-V:	a long vowel

Key takeaways

- Prefixes attach to the left of the stem, suffixes attach to the right of the stem, infixes attach internally to the stem, circumfixes attach to either side of the stem, and suprafixes are encoded in the prosody of the stem.
- Infixes and circumfixes are rare. If you think you have found an infix or a circumfix, double check that

a simpler analysis, using prefixes and suffixes instead, is not possible.
- Affixes are marked with hyphens to show which side of the stem they attach to.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=2846#h5p-39

Consult the following data in answering the check yourself questions:

(14) *Juliesh* (conlang)

a.	solabi	'I walk'
b.	sobok	'I swim'
c.	filabi	'we walk'
d.	fobok	'we swim'
e.	sobilabi	'I walked'
f.	fibibok	'we swam'

References and further resources

For students

Anderson, Catherine, Bronwyn Bjorkman, Derek Denis, Julianne Doner, Margaret Grant, Nathan Sanders, and Ai Taniguchi. 2022. Section 3.12: Tone and intonation. *Essentials of Linguistics, 2nd edition.* Pressbooks. https://ecampusontario.pressbooks.pub/essentialsoflinguistics2/chapter/3-12-tone-and-intonation/

Lieber, Rochelle. 2016. *Introducing morphology.* 2nd edition. Cambridge: Cambridge UP.

Academic sources

Cook, Eung-Do. 1971. Vowels and tones in Sarcee. *Language* 47 (1): 164-179.

🔍 Kalin, Laura. 2022. Infixes really are (underlyingly) prefixes/suffixes: Evidence from allomorphy on the fine timing of infixation. *Language* 98 (4): 641-682.

Spahr, Christopher. 2016. *Contrastive representations in non-segmental phonology*. Doctoral dissertation, University of Toronto.

Yu, Alan. 2007. *A natural history of infixation*. Oxford: Oxford UP.

6.3. INFLECTION AND DERIVATION

Another way we can classify affixes is whether they are **inflectional** or **derivational**. In this section, we will spend just a little bit of time on the difference between inflection and derivation to get the basic idea. We will go much deeper into their differences in Chapter 9.

Inflection

Inflectional morphemes encode the grammatical properties of a word. Some common examples of inflectional morphemes include **plural** markers on nouns, as shown in (1); **tense** and **aspect** markers on verbs, as shown in (2); or **comparative** and **superlative** markers on adjectives, as shown in (3).

(1)		**singular noun**	**plural noun**
	a.	*cat*	*cat-s*
	b.	*chair*	*chair-s*
	c.	*bush*	*bush-es*

(2)		**bare verb**	**past tense**	**past participle**	**present participle**
	a.	*walk*	*walk-ed*	*walk-ed*	*walk-ing*
	b.	*sew*	*sew-ed*	*sew-n*	*sew-ing*
	c.	*sleep*	*slep-t*	*slep-t*	*sleep-ing*

(3)		**bare adjective**	**comparative**	**superlative**
	a.	*red*	*redd-er*	*redd-est*
	b.	*long*	*long-er*	*long-est*
	c.	*happy*	*happi-er*	*happi-est*

Generally speaking, we don't consider inflectional forms of the same stem to be different words, but to be different forms of the same word. The collection of all of the inflectional forms of a root are called the **paradigm** for that word. For example, the row (1a) is the paradigm for the noun *cat*, row (2a) is the paradigm for the verb *walk*, and row (3a) is the paradigm for the adjective *red*.

A word without any inflectional morphemes is often called the **bare form**. In examples (1)-(3), the bare form of each paradigm is in the first column. The bare form may be a root or it may be a complex form containing both a root and derivational morphemes.

We can also talk about an inflected word form as having **morphosyntactic features**. For example, *cat* would have the feature [+singular] and *cats* would have the feature [+plural]. Morphosyntactic features classify words into groups based on similar behaviour. For example, English words with the feature [+plural] share the behaviour of being marked with a plural marker and having the meaning of 'more than one'.

Derivation

In contrast, derivational morphemes create words of new **parts of speech** or words with new meanings. For example, the verb *scare* can be changed into a new word, the adjective *scary*, by adding the derivational morpheme -y. Although *scare* and *scary* share the same root *scare*, they are different words. The verb *scare* refers to an event and belongs to a verbal paradigm and can take verbal inflectional morphology, such as *scared* and *scaring*, while the adjective *scary* refers to a quality and belongs to an adjectival paradigm and can take adjectival morphology, such as *scarier* and *scariest*.

Derivational morphemes don't always change the part of speech of the word, though. Morphemes that change the meaning of the word enough that we want to call it a new word, even if it keeps the same part of speech, are still called derivational morphemes. For example, the prefix *re-*, attaches to verbs to form new verbs that mean to perform the action again, as in (4). Both *scared* and *rescared* are verbs.

(4) a. My sister **scared** me.

b. My sister scared me again. She **rescared** me.

Sometimes derivational morphemes do both; they change the meaning substantially and change the part of speech. For example, adding the suffix *-er* to a verb creates a noun that identifies the person who performed the action, known as an agentive noun, or adding *-able* to a verb creates an adjective meaning that the action is possible to perform.

Derivational word forms based on the same root belong to the same **word family**, but each has their own, separate, inflectional paradigm. For example, the word family of *scare* includes the verb *scare* (5a), the adjective *scary* (5b), the noun *scare* (5c), and the adverb *scarily* (5d). It also includes more complex forms such as the repetitive verb *rescare* (5e), the agentive noun *scarer* (5f), and the adjective *scareable* which expresses possibility (5g). The word family even includes words containing multiple derivational morphemes, such as the word *scariness* in (5h), which consists of three morphemes: the verb root *scare*, the adjective-forming suffix *-y* (spelled with an <i> when it is not at the end of the word), and the noun-forming suffix *-ness*.

(5)	a.	My sister **scared** me.	*verb*
	b.	the **scary** costume	*adjective*
	c.	a big **scare**	*noun*
	d.	The ghost moved **scarily**.	*adverb*
	e.	After I calmed down, my sister **rescared** me.	*repetitive verb*
	f.	the **scarers** at the haunted house	*agentive noun*
	g.	My husband is easily **scareable**.	*adjective of possibility*
	h.	the **scariness** of this costume	*noun derived from the adjective*

While it is often possible to list the complete paradigm for a word, it is not possible to list the complete word family of a root, since we can build infinitely complex words out of one root.

Key takeaways

- Inflectional morphemes encode the grammatical properties of a word.
- The list of the different inflectional forms of a word is called a paradigm.
- We can formally indicate the inflectional properties of a word using morphosyntactic features.
- Derivational morphemes create new words by changing the part of speech of a word, substantially changing its meaning, or both.
- The list of different words derived from the same root is called a word family.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=2872#h5p-44

6.4. REDUPLICATION AND NON-CONCATENATIVE MORPHOLOGY

Reduplication

Reduplication is a special morphological process that involves full or partial copying of the stem. Some examples of full reduplication in Javanese are shown in (4). In all of these examples, reduplication indicates plurality.

(4)	**bare form**	**bare form gloss**	**reduplicated form**	**reduplicated form gloss**
a.	baita	*ship*	baita~baita	*various ships*
b.	səsupe	*ring*	səsupe~səsupe	*various rings*
c.	omaha	*house*	omaha~omaha	*various houses*

(Uhlenbeck 1978: 90, as cited by Booij 2005: 35)

Javanese also has partial reduplication. Whereas the reduplicated forms in (4) repeat the entire stem, the reduplicated forms in (5) only reduplicate the initial consonant. Note that the vowel that has been added in the reduplicated forms is not copied from the stem; in all cases, it is a /ə/, regardless of the vowels of the stem. In all of the examples in (5), reduplication indicates a change from a noun into a verb.

(5)	**bare form**	**bare form gloss**	**reduplicated form**	**reduplicated form gloss**
a.	gəni	*fire*	g~ə-gəni	*to warm oneself by the fire*
b.	jawah	*rain*	j~ə-jawah	*to play in the rain*
c.	tamu	*guest*	t~ə-tamu	*to visit*

(Uhlenbeck 1978: 90, as cited by Booij 2005: 35)

Javanese uses many different reduplicative patterns with a variety of meanings. For a thorough description of many of the ways reduplication is used in Javanese, see Miyake (2011).

According to the Leipzig glossing rules, reduplication should be marked with a tilde (~) connecting the reduplicant to the stem.

Non-concatenative morphology

Affixes, for the most part, join the root on its edges, resulting in a linear string of morphemes, but not all morphological processes result in a linear string of morphemes. Morphological processes that produce affixes in a linear string are called **concatenative**, from the Latin word *catena* 'chain', while morphological processes that do not result in a linear string are called **non-concatenative**. We have already seen some examples of non-concatenative morphology; **suprafixes**, which modify the **prosody** of a word, are one type of non-concatenative morphology.

Root-and-pattern morphology

A well-known example of non-concatenative morphology is **root-and-pattern morphology**, which is common in the Semitic language family. In Semitic languages, many roots consist only of three consonants. These are known as **triconsonantal roots**. Different sets of vowels are interwoven between the consonants in specific patterns, creating different but related word forms. These vowel patterns are sometimes called **transfixes**, because they are inserted across the root. An example from Hebrew is shown below in (1).

(1)	**Transfix pattern**	**Pattern meaning**	**Hebrew verb**	**Gloss**
a.	CaCaC	intransitive verb	gadal	He grew.
b.	CiCeC	to cause X	gidel	He raised.
c.	CuCaC	to be caused to be X	gudal	He was raised.
d.	hiCCiC	to make X	higdil	He enlarged.
e.	huCCaC	to be made to be X	hugdal	He was enlarged.

(adapted from Ussishkin 2000: 659)

Example (1) shows several Hebrew words formed from the triconsonantal root *gdl*. In the first column, we see several different transfixes that can be applied to the root. Each transfix pattern includes three Cs which represent the placement of the root consonants. The transfix pattern also indicates which vowels are interwoven into the word and where, as well as if there are any additional consonants, such as the initial /h/ in examples (d) and (e).

Triconsonantal roots are bound roots, because they cannot appear on their own, without the vowels of a transfix.

Ablaut

Ablaut, also called **internal change**, is when a morphological process is marked by changing the vowel of the stem. Ablaut is used in many Germanic languages, including English. Ablaut is similar to root-and-pattern morphology, except it only affects one vowel.

Some examples of ablaut in the formation of the English plural are shown in (2). As you can see, there are many different vowels that may be affected by ablaut in English. For example, /oʊ/ changes to /aɪ/ in the plural of *mouse* in (2a), and /u/ changes to /i/ in the plural of *goose* in (2b) or *tooth* in (2c), and so forth.

(2)	**singular**	**plural**
a.	mouse	mice
b.	goose	geese
c.	tooth	teeth
d.	woman	women

Ablaut also shows up in English verbs. As shown in (3), there are two forms in English that may have ablaut compared to the **bare form**: the past tense and the past participle. The **past participle** is used in the English perfect (*have sung*, *have come,* etc.) or the passive (*was sung*, *was found*). Some verbs, like *sing* in (3a), have a different vowel for each of the three forms. Others use the same vowel for the bare form and the past participle but a different vowel for the past tense, such as *come* in (3b). Others have the same vowel for the past tense and past participle but a different vowel for the bare form, such as *find* in (3c) or *get* in (3d). Finally, *drive* in (3e) has a different vowel in all three forms—the bare form and past participle may be spelled the same, but they're pronounced with different vowels—but the past participle additionally has the regular past participle suffix.

(3)	**bare form**	**past tense**	**past participle**
a.	sing	sang	sung
b.	come	came	come
c.	find	found	found
d.	get	got	got
e.	drive /dɹɑiv/	drove /dɹoʊv/	drive-n /dɹɪvən/

Key takeaways

- Reduplication is a morphological process whereby the stem is copied. The entire stem can be copied, which is called full reduplication, or just part of the stem, which is called partial reduplication.
- Non-concatenative morphology are morphological processes that do not result in a linear string. Suprafixes, root-and-pattern morphology, and ablaut are all examples of non-concatenative morphology.
- Root-and-pattern morphology is common in Semitic languages and consists of triconsonantal roots that are combined with a transfix. Transfixes indicate which vowels should be used and the arrangement of the consonants.
- Ablaut is when a morphological category is marked by a vowel change.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=2852#h5p-45

Consult the following data in answering the check yourself questions:

(4) *Javanese*

a.	abang	'red'
b.	abang-abang	'some what red, reddish'
c.	biru	'blue'
d.	biru-biruan	'bluish'

(Miyake 2011: 50)

(5) *Morrocan Arabic*

a.	ktəb	'write'
b.	ktab	'book'
c.	ktuba	'books'
d.	katib	'writer'

(Noamane 2018)

(6)
a. sing-song
b. ding-dong
c. chat-chat
d. dilly-dally

References and further resources

For a general audience

🎉 McLendon, Lisa. November 9 2020. Why ablaut reduplication is top-top. *ACES: The society for editing*. https://aceseditors.org/news/2020/why-ablaut-reduplication-is-tip-top

For students

Booij, Geert. 2005. *The grammar of words: An introduction to linguistic morphology.* Oxford: Oxford University Press.

Academic sources

🔍 Miyake, Yoshimi. 2011. Reduplication in Javanese. *Asian and African Languages and Linguistics* 6: 45-59.

Noamane, Ayoub. 2018. Morphological causatives in Moroccan Arabic: Word-based or root-based? *Asinag* 13: 217-240.

Uhlenbeck, E. M. 1978. *Studies in Javanese Morphology*. The Hague: Martinus Nijhoff.

Ussishkin, Adam. 2000. Root-and-pattern morphology without roots or patterns. *Proceedings of the North East Linguistics Society* 30(2).

6.6. ALLOMORPHY

Sometimes a morpheme changes form in different contexts. The different forms of a morpheme are called **allomorphs**. For example, the plural allomorph of English takes many different forms, as shown in (1). These include predictable variation, such as variation between **voiceless** /s/ after voiceless consonants as in *cats* in (1a), **voiced** /z/ after voiced consonants as in *dogs* in (1b), and epenthetic /əz/ after **sibilants** as in *bushes* in (1c). The English plural form also has unpredictable, or irregular, variation, including vowel changes as in *foot* versus *feet* in (1d), irregular suffixes such as *ox* versus *oxen* in (1e), and plural forms borrowed from other languages as with *cactus* versus *cacti* in (1f). Sometimes multiple strategies combine as with *child* versus *children* in (1g), which exhibits both an irregular suffix *-ren* and a vowel change from /aɪ/ to /ɪ/. Finally, sometimes the plural form is the same as the singular form, as with *sheep*. These are all examples of plural allomorphs in English.

(1)		**singular**	**singular pronunciation**	**plural**	**plural pronunciation**
	a.	*cat*	/kæt/	*cat-s*	/kæts/
	b.	*dog*	/dɑg/	*dog-s*	/dɑgz/
	c.	*bush*	/bʊʃ/	*bush-es*	/bʊʃəz/
	d.	*foot*	/fʊt/	*feet*	/fit/
	e.	*ox*	/ɑks/	*ox-en*	/ɑksən/
	f.	*cactus*	/kæktɪs/	*cacti*	/kæktaɪ/
	g.	*child*	/t͡ʃaɪld/	*child-ren*	/t͡ʃɪldɹən/
	h.	*sheep*	/ʃip/	*sheep*	/ʃip/

We can classify allomorphs based on two properties. First, we consider the relationship between the different forms of the morpheme. If the different allomorphs are related to each other through a regular phonological pattern, they are called **phonological allomorphs**. If they are not, they are called **suppletive** allomorphs. Second, we consider the contexts that trigger allomorphy. Allomorphy can be triggered by **phonological**, **morphological**, or **lexical conditioning**.

Phonological allomorphs

When the form of an allomorph is predictable based on phonological context, it is called a **phonological**

allomorph. In other words, phonological allomorphs are in complementary distribution based on a predictable phonological context. It is possible to write a morphophonological rule to describe where each allomorph is used.

For example, the plural allomorphs in (1a)-(1c) are phonological allomorphs. The plural morpheme surfaces as /əz/ after **sibilants**, as /s/ after other voiceless consonants, and as /z/ elsewhere. We can write two morphophonological rules which, taken together, describe this pattern, as in (2).

(2) a. plural -z → -əz / sibilant _

b. plural -z →-s / voiceless C _

We can also write a distribution diagram, as in (3).

(3)		
		-əz / sibilant _
	plural	-s / voiceless C _
		-z / elsewhere

Suppletion

On the other hand, **suppletive** allomorphs are not predictable from phonological context.

Weak suppletion has some phonological similarity to the other allomorphs. Some example of weak suppletion can be identified in English past tense forms, shown in (4).

(4)		**present**	**present pronunciation**	**past**	**past pronunciation**
	a.	*buy*	/baɪ/	*bought*	/bɑt/
	b.	*catch*	/kɛt͡ʃ/	*caught*	/kɑt/
	c.	*teach*	/tit͡ʃ/	*taught*	/tɑt/

The present tense forms *buy, catch,* and *teach* have nothing in common with each other that would explain why they do not take the regular *-ed* past tense marker, which is why these past tense forms are considered suppletive and not phonological. However, each present tense and past tense pair share the same initial consonant—/b/ for *buy* and *bought*, /k/ for *catch* and *caught*, and /t/ for *teach* and *taught*—which is why these are considered weak suppletion.

In contrast, in **strong suppletion**, the allomorphs do not share any phonological similarity. For example, as shown in (5a), the root for the verb *go* and its past tense *went* have no phonological similarity. Likewise, shown in (5b), the root *good* has no phonological similarity with the **comparative** form *better* or the **superlative**

form *best*. Therefore, *good* exhibits strong suppletion. On the other hand, the relationship between *better* and *best* could be considered weak suppletion, because they share the initial /bɛ/.

(5)	a.	**bare root**	*go*	/goʊ/
		past tense	*went*	/wɪnt/
	b.	**root**	*good*	/gʊd/
		comparative	*better*	/bɛɾəɹ/
		superlative	*best*	/bɛst/

Where does suppletion come from?

Historically, suppletion has many different possible sources.

One common source of suppletion is when a **productive**, historical process has died out, leaving only a few remnants in the modern grammar. This is the source of the plural form *feet*. Old English nouns followed many different conjugation patterns, much like Latin. Some, including *foot*, had a regular **ablaut** pattern, where plurality was marked by a vowel change in the stem, between the singular *fot* and the plural *fet* (OED). Eventually, though, the plurals in English went through a process of **regularization**, until today, where the majority of words take an *-s* plural marker. However, some of the more frequent words, including *foot,* kept the ablaut plural form.

Another common source of suppletion is through **analogy**, when a word adopts a similar pattern to another word. For example the word *catch* is borrowed from French *cacher,* meaning *to hunt* (OED), so at first it may be surprising that it has a suppletive past tense that looks like it may have been left behind by an older ablaut pattern of Old English. **Etymologists** think the past tense *caught* came from analogy from other words ending in -t͡ʃ which also have irregular pasts, such as *latch* and *teach* (OED).

Another source of suppletion is when two **paradigms** merge. For example, the past tense of GO, *went*, comes from a different verb, *wend,* meaning to turn or change direction. It has been used as the past tense of GO since the 15th century, replacing the old past tense forms of GO, which were also suppletive (OED).

In general, suppletion is much more common in frequent words, since children need to be exposed to the irregular pattern frequently enough to learn it.

You can look up the historical development of English words in the *Oxford English Dictionary*, which can likely be accessed online for free through your university library. Search for the word you're interested in, and then click the *Etymology* tab to learn about its history.

Conditioning

The choice of which allomorph is used where can depend on different factors: it can be phonologically, morphologically, or lexically conditioned.

When allomorphy is phonologically conditioned, it means the choice of allomorph depends on phonological context. For example, the regular forms of the plural allomorph are phonologically conditioned. The plural morpheme *–s* appears as [əz] after a sibilant, [s] after a voiceless consonant, and [z] elsewhere.

Morphological conditioning, on the other hand, depends on the morphological context, usually the grammatical properties of the morpheme, such as an inflectional category. For example, in Spanish, the verb meaning 'to go' uses the stem *ir* in the infinitive and future, *va-* in the present and imperfective past, and *fu-* in the perfective past. This is morphological conditioning because it depends on morphological features such as present, infinitive, or future.

Finally, lexical conditioning depends on the properties of the stem. For example, the Persian plural marker appears as *–an* for human nouns and *–ha* for non-human nouns, which is an example of lexical conditioning based on semantic properties. Lexical conditioning also includes any form that is not predictable (i.e., must be memorized). The irregular plural *-ren* in *children* is an example of lexical conditioning that is not predictable. There is nothing about the word *child* that predicts that it would have an irregular plural.

Spelling changes vs. allomorphy

One mistake that students sometimes make is misidentifying a spelling change as allomorphy. Allomorphy describes only changes in pronunciation, not changes in spelling.

Sometimes a change in spelling and allomorphy co-occur, as in (6). In (6a), the morpheme *invade* in its bare form ends with a /d/ sound and is spelled with a /d/. In (6b), when *-ion* has been suffixed, the pronunciation of the final segment of the stem changes from /d/ to /ʒ/.

(6)	a.	*invade*	/ɪnveɪd/
	b.	*invas-ion*	/ɪnveɪʒən/

Sometimes, there is allomorphy without a change in spelling, as in (7). In (7a), the morpheme *electric* in its bare form ends with a /k/ sound. However, after *-ity* has been suffixed, as shown in (7b), the /k/ sound changes to an /s/ sound. On the other hand, if *-ian* is suffixed, the /k/ sound changes to an / ʃ / sound. All three sounds /k s ʃ / are spelled with a <c>.

(7)	a.	*electric*	/əlɛktɪk/
	b.	*electric-ity*	/əlɛktrɪsəti/
	c.	*electric-ian*	/əlɛktriʃən/

Finally, sometimes there is a change of spelling without allomorphy, as in (8). In (8a), the morpheme *happy* in its bare form ends with an /i/ sound, spelled with a <y>. When *-ness* is suffixed, the spelling changes from <y> to <i>, but the pronunciation stays the same.

(8)	a.	*happy*	/hæpi/
	b.	*happi-ness*	/hæpinəs/

Remember, as discussed in Section 4.1, that writing as a modality is secondary to spoken language. Written language is conscious, learned by convention, and, as shown in these examples, does not always accurately reflect the subconscious grammatical patterns of language.

Key takeaways

- A morpheme can take different forms in different contexts, which is called allomorphy.
- When the form of an allomorph is predictable based on phonological context, it is called a phonological allomorph.
- If the form of an allomorph is not predictable based on phonological context, it is called a suppletive allomorph. Weak suppletion is when the allomorphs bear some resemblance,

while strong suppletion is when the allomorphs bear no resemblance.

- Allomorphy can also be classified based on what triggers the allomorphy. Phonologically conditioned allomorphs arise due to phonological context, morphologically conditioned allomorphs arise due to the inflectional properties of the stem, and lexically conditioned allomorphs arise due to the lexical properties of the stem, such as its semantic properties or idiosyncratic properties of the stem.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=2875#h5p-43

6.7. CLITICS

A clitic is a morpheme which has some of the properties of an independent word and some properties of an affix. The word clitic comes from the Latin word for leaning and is related to words like *incline*. You can think of a clitic as a morpheme that leans on another word, which is called the host.

Clitics depend on a word that they are adjacent to. An enclitic leans on the word that comes before it, and so looks kind of like a suffix, while a proclitic leans on the word that comes after it, so looks kind of like a prefix.

According to the Leipzig glossing rules, you may indicate that a morpheme is a clitic by using an equal sign (=) instead of a hyphen (-), but this is an optional rule.

There are lots of different kinds of clitics. What unites them all is that they don't behave quite like a word, and they don't behave quite like an affix either. In this section, we will look at some examples of some of these different kinds of clitics.

Bound morphemes which modify an entire phrase

One group of morphemes that can be categorized as clitics are bound morphemes which modify an entire phrase. One example of this kind of clitic is **genitive** *'s* in English. As shown in example (1), genitive *'s* indicates possession. For example, in (1a), *'s* attaches to *Sharon* and indicates that the house belongs to Sharon.

(1) a. Sharon's house
b. [My little sister]'s house
c. [My little sister with the green shirt]'s house
d. [My little sister with the green shirt and the mischievous grin]'s house
e. [My little sister that ran a marathon]'s house

Like an affix, *'s* cannot stand on its own—it is a bound morpheme. However, unlike an affix, it doesn't modify just the word it attaches to; it modifies an entire phrase. This is shown in examples (1c)-(1e). In these three examples, *'s* attaches to the words *shirt*, *grin*, and *marathon*, respectively. However, these phrases do not mean that the house belongs to the shirt, the grin, or the marathon. In all three sentences, the house belongs to the sister, since *sister* is the head of the phrase. In other words, *'s* modifies the entire phrase it attaches to, not just the one word it attaches to.

Bound morphemes which attach to a word they do not modify

Another category of clitics are bound morphemes that attach to a word they do not modify. This is the behaviour of determiners in the language Nisga'a. In Nisga'a, determiners cliticize to the word that appears before them, even if they don't belong to the same phrase. This is shown below in (2). The noun *Mary* takes the **proper noun** determiner *=t*, which encliticizes to the verb *giba* before *Mary*. Likewise, the noun *hlgiikw* 'sister' takes the **common noun** determiner *=hl*, which encliticizes to the word *Mary* before it.

(2) *Nisga'a*

Yukwt		giba**s**			Mary**hl**		hlgiikwt.		
yukw	**=t**	giba	-t	[=t	Mary]	[**=hl**	hlgiikw	-t]	
IPFV	=3	wait	-3	[=PN	Mary]	[=CB	sister	-3]	

'Mary is waiting for her sister.'

(Tarpent 1987: 211-212, as cited in Forbes 2024)

Morphemes that occur in a different location than expected

Some morphemes are classified as clitics simply because they appear in a different location than expected. Object pronouns in Romance languages are often called clitics for this reason. Consider the French example below. The typical word order of French is subject-verb-object, as shown in (3a). The object, which is the entity that the event happens to, in this case *la fille* 'the girl,' appears after the verb when it is a full noun phrase. However, when the object is a pronoun, as in (3b), it appears before the verb instead.

(3) a. *French*

(subject verb object)

Pierre	voit	**la**	**fille.**
Pierre	see.3SG	the	girl

'Pierre sees the girl.'

b. (subject object verb)
Pierre **la** voit.
Pierre her see.3SG
'Pierre sees her.'

Clitic pronouns in Romance languages have some very unique and well-known properties. For example, some Romance languages have a construction called clitic doubling where a noun phrase appears twice—both as a full noun phrase and as a clitic pronoun, as shown in (4). In this sentence, the indirect object is marked twice: once right before the verb as the pronoun *le* and once at the end of the sentence as the full noun phrase *a Juan*.

(4) *Spanish*
Carolina **le** dio un libro **a** **Juan.**
Carolina DAT3SG gave a book DAT Juan
'Carolina gave a book to Juan.'

(Belloro 2007: 6)

Another unique property of clitic pronouns in some Romance languages is that they can undergo a process of clitic climbing. Clitic climbing is when a clitic moves from an embedded clause to the main clause, as shown in example (5). In the Catalan example in (5), the 2nd person singular clitic pronoun *t'* semantically belongs in the embedded clause with the verb *veure* 'see'. However, it shows up in the main clause attached to *he vingut* 'have come,' even though it has no semantic association with the event of coming.

(5) *Catalan*
T'-he vingut a veure.
you- have.1SG come to see.INF
'I have come to see you.'

(Solà 2002: 228)

Another special kind of clitic that appears in an unexpected position are called second-position clitics. These clitics always appear after the first word in their phrase, no matter what the word is. Some examples from Serbo-Croatian are shown in (6). In these examples, there is a string of three second-position clitics, *smo* 'are', *mu* 'him', and *je* 'her.' When there is a cluster of clitics like this, it is known as a clitic group. In sentence (6a), the clitic group appears after the subject pronoun *mi* 'we'. In (6b), the subject is dropped, so the clitic group moves to appear after the verb *predstavili* 'introduced'. Sentence (6c) is a question, so the clitic group appears

after the question word *zašto* 'why'. Regardless of what the first word of the sentence is, the clitic group always appears immediately after it, in second position.

(6) a. *Serbo-Croatian*

Mi	**smo**	**mu**	**je**	predstavili	juče.
we	are	him.dat	her.acc	introduced	yesterday

'We introduced her to him yesterday.'

b.

Predstavili	**smo**	**mu**	**je**	juče.	
introduced	are	him.dat	her.acc	yesterday	

'We introduced her to him yesterday.'

c.

Zašto	**smo**	**mu**	**je**	predstavili	juče?
why	are	him.dat	her.acc	introduced	yesterday

'Why did we introduce her to him yesterday?'

(Bošković 2016: 28)

Words that cannot bear stress

Another reason a word might be considered a clitic is if it is prosodically deficient. Generally, this means that it does not usually bear stress. Many functional words, including English words such as *the*, *a*, and *of* might be considered clitics for these reasons.

Key takeaways

- Clitics are bound morphemes that have some properties of an independent word and some properties of an affix.
- Clitics that attach to the beginning of their host are called proclitics and clitics that attach to the end of their host are called enclitics.

- Some unusual properties of clitics include that they might appear in unexpected positions, they might attach to a word that they do not modify, or they might be a bound morpheme that modifies an entire phrase.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=2858#h5p-48

Consult the following data in answering the check yourself questions:

(7) a. *Comanche*

tɨasi-se	nɨ	tɨhka
again-DM	I	eat

'Again I ate.'

b. *nɨ tɨhka

I	eat

'I eat.'

c. tɨhka nɨ

eat	I

'I eat.'

(McDaniels 2008, as cited in Bošković 2016: 35)

References and further resources

Academic sources

Belloro, Valeria. 2007. Spanish clitic doubling: A study of the syntax-pragmatics interface. PhD thesis, State University of New York at Buffalo.

Bonet, Eulalia. 2019. Clitics and clitic clusters in morphology. *Oxford Research Encyclopedia of Linguistics*. https://oxfordre.com/linguistics/display/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-519

Bošković, Željko. 2016. On second position clitics crosslinguistically. In *Formal studies in Slovenian syntax: In honor of Janez Orešnik*, ed. Franc Lanko Marušič and Rok Žaucer. Amsterdam: John Bemjamins. 23-53.

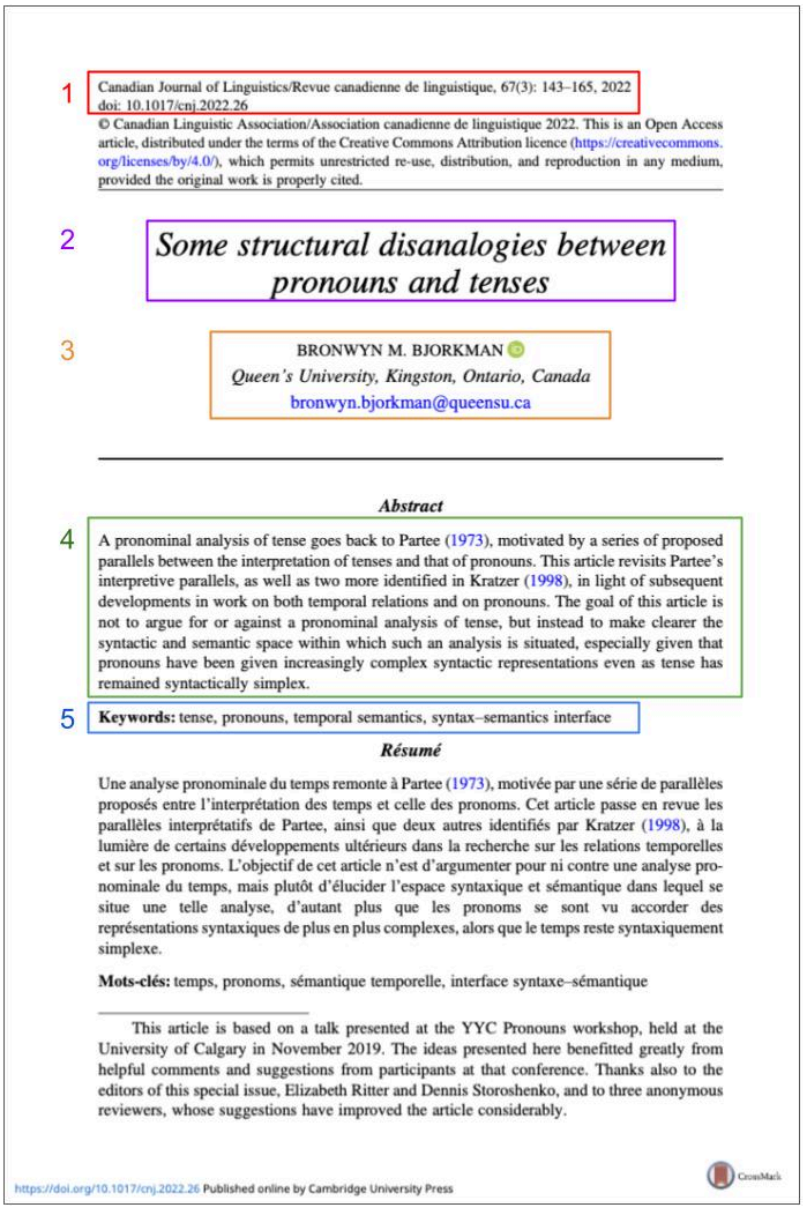
Forbes, Clarissa. 2024. Tsimshianic. In *The Languages and linguistics of Indigenous North America: A comprehensive guide, Volume 2,* ed. Carmen Dagostino, Marianne Mithun, and Keren Rice. Walter de Gruyter. 985-1012.

Solà, Jaume. 2002. Clitic climbing and null subject languages. *Catalan Journal of Linguistics* 1: 225-255.

6.8. BECOMING A LINGUIST: ANATOMY OF AN ACADEMIC ARTICLE

The first page

The first page of an academic article has lots of useful information, as shown in the annotated picture of the first page of Bjorkman (2022) in Figure 1. I've marked five areas where you should look when first navigating an academic article, which I will explain below. This article is published in the *Canadian Journal of Linguistics*. Other journals will have different layouts, but they will generally be similar and have many of the same elements.



1
Canadian Journal of Linguistics/Revue canadienne de linguistique, 67(3): 143–165, 2022
doi: 10.1017/cnj.2022.26
© Canadian Linguistic Association/Association canadienne de linguistique 2022. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

2
Some structural disanalogies between pronouns and tenses

3
BRONWYN M. BJORKMAN
Queen's University, Kingston, Ontario, Canada
bronwyn.bjorkman@queensu.ca

Abstract

4
A pronominal analysis of tense goes back to Partee (1973), motivated by a series of proposed parallels between the interpretation of tenses and that of pronouns. This article revisits Partee's interpretive parallels, as well as two more identified in Kratzer (1998), in light of subsequent developments in work on both temporal relations and on pronouns. The goal of this article is not to argue for or against a pronominal analysis of tense, but instead to make clearer the syntactic and semantic space within which such an analysis is situated, especially given that pronouns have been given increasingly complex syntactic representations even as tense has remained syntactically simplex.

5
Keywords: tense, pronouns, temporal semantics, syntax–semantics interface

Résumé

Une analyse pronominale du temps remonte à Partee (1973), motivée par une série de parallèles proposés entre l'interprétation des temps et celle des pronoms. Cet article passe en revue les parallèles interprétatifs de Partee, ainsi que deux autres identifiés par Kratzer (1998), à la lumière de certains développements ultérieurs dans la recherche sur les relations temporelles et sur les pronoms. L'objectif de cet article n'est d'argumenter pour ni contre une analyse pronominale du temps, mais plutôt d'élucider l'espace syntaxique et sémantique dans lequel se situe une telle analyse, d'autant plus que les pronoms se sont vu accorder des représentations syntaxiques de plus en plus complexes, alors que le temps reste syntaxiquement simplexe.

Mots-clés: temps, pronoms, sémantique temporelle, interface syntaxe–sémantique

This article is based on a talk presented at the YYC Pronouns workshop, held at the University of Calgary in November 2019. The ideas presented here benefitted greatly from helpful comments and suggestions from participants at that conference. Thanks also to the editors of this special issue, Elizabeth Ritter and Dennis Storoshenko, and to three anonymous reviewers, whose suggestions have improved the article considerably.

https://doi.org/10.1017/cnj.2022.26 Published online by Cambridge University Press

Figure 1. The first page of an academic journal article, Bjorkman (2022), published by the *Canadian Journal of Linguistics*. Used under CC BY 4.0 license.

Box 1 (red): The publication information

At the top of the page, as part of the header, the publication information of the article is listed. In this case, it includes the journal title, the volume and issue numbers, the page numbers, the year of publication, and the DOI, which is a unique number assigned to academic articles to help you find them. You will need this information in order to write a citation for the article. In some journals, this could be listed on the bottom of the page or on a cover page before the article begins, instead.

Box 2 (purple): The title

This is the name of the article, and one of the first things you will look at when deciding whether you should read an article.

The title of this article is *Some structural disanalogies between pronouns and tense.* What can you tell about this article based on this title?

- The paper is probably written in English, since the title is in English.
- It is about the structure of **pronouns** and **tenses**, and so is probably related to the fields of morphology, syntax, and/or semantics.
- The word *disanalogies* suggests that it is about how pronouns and tense are different from each other.

Box 3 (orange): The author's information

The author of the article is usually listed underneath the title, alongside other information about the author, such as their affiliation. This article is written by Bronwyn Bjorkman who is affiliated with Queen's University in Kingston, Ontario.

This article also provides the author's email. If you do wish to contact an author, I recommend googling them, in case their contact information has changed since the publication of the paper.

Box 4 (green): The abstract

Underneath the author's information is a paragraph labeled *abstract.* When *abstract* is used as an adjective, it usually means the opposite of *concrete*. But when *abstract* is used as a noun, especially in academic settings, it usually means a summary that you can use to decide whether you want to engage with a bigger work. You can think of the abstract as a movie trailer or the blurb on the back of a book.

Not all journal formats include a header labeling the abstract. Sometimes, the abstract will just be found at the beginning of the paper in a box or a different font. More rarely, a paper may not have an abstract at all.

Here is the text of this article's abstract:

> A pronominal analysis of tense goes back to Partee (1973), motivated by a series of proposed parallels between the interpretation of tenses and that of pronouns. This article revisits Partee's interpretive parallels, as well as two more identified in Kratzer (1998), in light of subsequent

> developments in work on both temporal relations and on pronouns. The goal of this article is not to argue for or against a pronominal analysis of tense, but instead to make clearer the syntactic and semantic space within which such an analysis is situated, especially given that pronouns have been given increasingly complex syntactic representations even as tense has remained syntactically simplex.
>
> -Bjorkman 2022: 143

Here's what you can tell from this abstract:

- Partee (1973) and Kratzer (1998) argue that there are some parallels between tense and pronouns, and they use those parallels to provide an analysis of tense.
  Note: Barbara Partee and Angelika Kratzer are both very well-known and influential semanticists. Their status combined with the age of their articles suggests that their analyses of tense have been very important, perhaps foundational.
- These analyses are about the interpretation of tense and pronouns, which suggests that this article may include some semantic analysis.
- More work has been done on tense and on pronouns since Partee's and Kratzer's articles were published, which is why their analyses are worth revisiting.
- The article does not necessarily contradict Partee and Kratzer.
- One difference between pronouns and tense is that pronouns tend to be analyzed as having complex syntactic structures but tense as having simpler structure.

Because the *CJL* is a bilingual publication, it provides abstracts for all of its papers in both French and English (even though the papers themselves are not bilingual). In this case, the French abstract is below the English one, labeled *Résumé*.

Different kinds of academic materials have different kinds of abstracts, with some differences. For example, the abstract of an academic article, like this one, is only about a paragraph long. It won't usually include references or headings (although in some other fields, it may). It is usually provided right before the article it is about.

The abstract for a conference presentation, in contrast, is often submitted anonymously months before the conference begins, and is used to determine which presentations are included in the conference. The selected abstracts will then be de-anonymized and published on the conference website, so that people can use them to decide which presentations to attend. Conference abstracts are usually one or two pages long, with strict page limits set by the conference organizers. Headings are usually used to conserve space and direct the readers' attention, and an abbreviated reference list may be included at the end.

Box 5 (blue): Keywords

Underneath the abstract, you will often find keywords, although not all journals include keywords for their articles. These keywords give you more clues about what the article is about. The keywords will often indicate which languages are being analyzed, which subfields of linguistics the article can be classified under, and which grammatical properties are being analyzed.

The keywords for this paper are *tense*, *pronouns*, *temporal semantics*, and *syntax–semantics interface*. Here's what you can tell from the keywords:

- We can confirm that the main topics of the paper are tense and pronouns (which we already figured out from the title and abstract).
- The main subfield of this paper is probably semantics, but it is also relevant to syntax.

This paper also includes French keywords underneath the French abstract, under the label *Mots-clés*.

The body of the paper

The body of the paper will probably be divided into numbered sections and subsections. As mentioned in Section 2.6, morphosyntax papers don't usually have a formulaic structure. Instead, the structure will typically be described in an outline at the end of the introduction.

Throughout the paper, you may find tables, graphs, linguistic data, syntax trees, and other formalisms. These will typically be numbered and discussed in the text. It is not considered acceptable to include data and formalisms without providing a description, relevant context, and explanation. If you are not sure how to interpret the data or formalisms, the discussion in the text may help you to understand them better.

The fine print

Footnotes and endnotes

Academic papers tend to have a lot of additional notes. If they are found at the bottom of each page throughout the paper, they are called footnotes. If they are found at the end of the paper, they are called endnotes. Papers will either use one or the other, not both.

When you are writing a paper, don't go out of your way to include footnotes or endnotes. If it is important, it should be included in the main text, not in a note. In published papers, many of the footnotes or endnotes get added through the review process, as responses to the reviewers.

Footnotes and endnotes may contain points of clarification or additional details. A lot of researchers hide their problems and limitations in their notes, so if you are writing a critique of a paper or looking for ideas for further research, reading the notes can be helpful.

In some citation styles, references are included in the notes. As described in Section 2.7, we typically use in-text citations in linguistics instead.

Acknowledgements

There will typically be an acknowledgements section included in the article. This may be found as a footnote on the title or in a special section at the end of the article. If the researchers collected data using elicitation, their language consultants will often be thanked in the acknowledgements section, giving you more information about how the data was collected.

Abbreviations list

If there is glossed linguistic data, an abbreviations list should be included in the article. This may be found in a footnote on the first example or at the end of the article in a special section.

References

At the end of the article, there will be a references list. The references list should include the full bibliographic information of every source cited in the paper. The format of the references list used in the *Canadian Journal of Linguistics* is described in detail in Section 2.7.

Key takeaways

- On the front page of an academic article, you should be able to find the title, the author's information, the bibliographic information of the article, a summary of the article called an *abstract*, and sometimes article keywords.
- The main body of the paper will usually be divided into numbered sections and subsections. The organization of the paper will often be described at the end of the introduction.
- Data and formalisms included throughout the paper will be numbered and should be described in the main text.
- Footnotes and endnotes can be a good place to find problems with the paper or areas of further research.
- Papers will often have acknowledgements and abbreviations list that can provide information to help you interpret the data in the paper.
- Academic articles should always end with a references section that provide the full bibliographic information of all sources cited in the article.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=4099#h5p-50

References and further resources

Sources for examples

Bjorkman, Bronwyn. 2022. Some structural disanalogies between pronouns and tenses. *Canadian Journal of Linguistics 67* (Special Issue 3: Pronouns): 143-165.

6.9. BECOMING A LINGUIST: HOW TO READ AN ACADEMIC ARTICLE

Academic articles vs. textbooks

You are probably used to reading textbooks, but reading academic articles may be a new skill. Here are some important differences between academic articles and textbooks to keep in mind.

The most obvious difference between a textbook and an academic article is the length. While a textbook is a whole book, an academic article is short. In linguistics, they are often between 15 and 50 pages long.

A textbook is usually used for learning or studying, while an academic article is usually used for research. This changes how you read! When you are reading a textbook, every detail might be important, and your professor will probably expect you to understand all or most of the reading. On the other hand, your professor will probably not expect you to understand every part of an academic article. When you are reading an academic article, you shouldn't pay equal attention to all parts of the paper. Think about why you are reading that paper, and focus on those parts. For example, if you are studying the same language as the paper, you will spend more time studying the data. If you are trying to do a similar study on your own, you will spend more time studying the methods. If you care most about the theory the paper is arguing for, then you will spend more time studying the argumentation and conclusions.

A textbook will assume that you know nothing about the topic. The purpose of the textbook is to teach you the topic! An academic article, in contrast, assumes you have background knowledge. This is the bit that makes reading academic articles the most difficult. When you are just starting out, of course you won't have all of the background knowledge! Morphosyntax as a field is big enough that even experienced researchers won't have all of the background knowledge necessary to understand the article. Academic articles, remember, typically present original research, which means that they are presenting new information. The article will contain brand new ideas, so even the most experienced professor will be learning something new from the article. Here are some strategies that can help you navigate your gaps in knowledge:

- Most of the time, you just need to understand the main claim of the article, not all of the details. Work on figuring out that main claim, and don't stress over all of the details that you don't understand.
- Try to figure out which pieces of background knowledge are really necessary to understanding the main goal of the paper, and which are not that important. Don't look up every word you don't know.
- Remember your goal for reading the paper! Do you care more about the data, the methods, or the theory? Spend most of your time and energy trying to understand the parts that are relevant to you.

- Pay attention to the year the paper was written. If you know a little bit about the history of linguistics, the year is a big clue about what kinds of assumptions and frameworks the author was working with.
- Sometimes the paper uses a technical term that is new to you, but maybe you do know it, just by a different name. If it looks familiar but they call it something else, don't doubt yourself! It is good to double check that they mean the same thing, though.

Once you have identified a gap in your knowledge that you really need to fill in order to meet your research goals, how do you go about learning what you need to know? Of course, you can search online, but you may find that online searches might turn up simplified or irrelevant explanations, depending on the term. Here are some ways that you can get more specialized answers, ordered from easiest to most difficult:

- Ask your instructor for advice or ideas about where to go. Your instructor may be able to give you specific advice about individual topics.
- Look up terms in a linguistics glossary or dictionary, such as the one at the end of this textbook or Utrecht's *Lexicon of Linguistics*.
- Find a textbook that introduces and explains the topic.
- Find a cover article that explains the topic. A cover article is a special kind of academic article that, instead of presenting new original research, gives an overview of the current ideas and research on a given research area.
- Find a dissertation or book that addresses the topic, and read the introduction or literature review section of the dissertation or book.

Finally, in morphology and syntax, we are often working with a model of grammar. A textbook will present a complete model. It may be a simplified model, and the textbook author has thought about how all the pieces integrate into the whole. On the other hand, an academic article may present just one piece and go deep into that one piece. The author of the academic article might not have an opinion about how this piece interacts with all of the other pieces in the model.

Reading techniques: The three pass method

Believe it or not, most academics do not read an article from beginning to end straight-through. Instead, they bounce around strategically, making decisions about where they want to focus and how much time and energy

they need to spend on understanding the article. In this section, I will describe the three pass method to reading an academic article which can help you begin to read more strategically. This description is adapted from Keshav (2002) with elements from Tatman (2016).

The first pass: A strategic skim

The first pass through the article should be quick, maybe even only 10-15 minutes. You're only going to read parts of the article, and your main goal of the first pass is to decide if this is the right article for you to spend your time on.

Step one: Read the title, abstract, and keywords carefully

As you're reading, try to figure out what the main goal of the paper is. Decide if you want to spend more time on this article, and if so, which parts are most relevant. Ask yourself why you are reading this paper. Which questions are you trying to answer? Does this article address those questions?

Step two: Scan the introduction and conclusion

Skim through the introduction and conclusion. Try to determine what the main conclusions of the paper are and what kind of evidence was used.

At this point, you may also wish to note any terms you don't know that are used more than once in both the introduction and the conclusion. Consider looking them up now, so that you are better equipped to understand the meat of the paper in the second pass. Remember, you don't need to look up every term, just the ones that seem like they will be important.

Step three: Flip through the paper and read the headings and skim the data

In this step, you are figuring out how the paper is structured. You are trying to determine which sections will be most relevant to the questions you're trying to answer, and how the different sections relate to each other. You can use this information to pace yourself in the second pass, especially if the article is long enough that you won't be able to read it in one sitting. You might also choose to only read portions of the paper in the second pass.

As you skim the paper, pay attention to what kind of data is used. For example, the data may consist of full sentences, individual words, morphological paradigms, or statistics. Which language(s) does the data come from? Is it adult or child language? Also pay attention to what kinds of formalisms and diagrams are used. For

example, the paper may include semantic formula, syntax trees, tables, charts, graphs, or formalisms that are unfamiliar to you. You can use the data and the formalisms used to get a bigger picture of what kind of paper it is: the theoretical framework, the methodology, and the subfields that are relevant.

Step four: Stop and evaluate

At the end of the first pass, take some time to answer the following questions (adapted from Keshav 2007).

1. **Category**: What kind of paper is it? Is it summarizing previous work or presenting a new analysis? What subfield(s) does it fall under? What methodologies are used? Is it descriptive, describing the properties of some language, or theoretical, arguing for a theoretical position, or both?
2. **Context**: What theoretical framework does it adopt? Is it one you are familiar with? What other papers is it related to?
3. **Corrections**: Did you notice any errors or faulty assumptions?
4. **Contributions**: How does this paper advance the field?
5. **Clarity**: Is it well-written?

At this point, decide whether to keep reading. You may choose to stop after the first pass if you don't think the paper will be useful to you after all, either because it wasn't quite the right topic or methodology, because you think it makes faulty assumptions, or because you think you already got the information you need from the first pass.

The second pass: A quick read-through

In the second pass, you will either read the paper through from beginning to end or read just the sections that you picked out as most relevant to you. In the second pass, you will read more carefully but you still aren't trying to understand every detail. Instead, you should focus mainly on trying to understand the main idea of the paper. After the second pass, you should be able to explain the main argument of the paper as well as the main supporting evidence. If you don't understand a paragraph after reading it a couple times, move on. The second pass shouldn't take you much longer than 2-3 hours, depending on the length of the paper.

As you read, formulate questions based on what you don't understand and based on further research goals:

1. What don't you understand? Pick out terms that seem important that you don't know. Are there parts of their argument that leave you confused?
2. What is something you'd like to know more about? If you had to write a research paper about something, what ideas does this paper give you?

There are many reasons why you might still not understand the main ideas of the paper after the second pass. The paper may be in a topic area, framework, or methodology that you aren't familiar with, with lots of unfamiliar terminology and/or formalisms. In other words, you don't have sufficient background. The paper may also be poorly written. Finally, maybe you're just tired. If you don't understand the paper after the second pass, there are several possible responses. You could rest and try again. This may help simply because the second time trying to understand something is easier than the first time. You could do more background reading and research to fill in your gaps of understanding, and then try again. Or you can set it aside and hope that it's not going to be important to your course or your career.

The second pass is enough depth for most papers you read. It is enough for most course assigned readings. However, if you are leading a discussion about the paper, writing a critique or review of the paper, or using it as the main resource for your own research, you should probably move on to the third pass.

The third pass: A deep reading

If you choose to do a third pass, this is when you try to understand every single detail in the paper. Consider every step of the argumentation and every piece of data. Read the footnotes. Look for flaws. Ask yourself, if you were writing this paper, would you have made the same choices and drawn the same conclusions?

Key takeaways

- Reading an academic article requires different strategies than other kinds of reading.
- When reading an academic article, you should keep in mind your purpose for reading it and strategize which parts of the article to spend more time on.
- Do not expect to understand all of the details in an academic article, especially in an area that is new to you.
- Most academics usually skim the title, headings, introduction, and conclusion before they read an article as a whole, to help them decide if they want to read it more carefully and which parts to focus on.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=3267#h5p-51

References and further resources

For students

Fruehwald, Josef. August 22, 2022. Reading a technical paper. *Lin517: Natural Language Processing*. https://jofrhwld.github.io/teaching/courses/2022_lin517/resources/reading/

Kerstens, Johan, Eddy Ruys, and Joost Zwarts. 2001. *Lexicon of Linguistics*. Utrecht University. https://lexicon.hum.uu.nl/

Keshav, Srinivasan. 2007. How to read a paper. *ACM SIGCOMM Computer Communication Review* 37 (3): 83-84.

Tatman, Rachael. January 29 2016. How to read a linguistics article in 8 easy steps. *Making Noise & Hearing Things.* https://makingnoiseandhearingthings.com/2016/01/29/how-to-read-a-linguistics-article-in-8-easy-steps/

6.10. GUIDED READING: IS *N'T* A CLITIC OR AN AFFIX?

In this section, we are going to read through an academic article together, *Cliticization vs. inflection: English N'T* by Arnold Zwicky and Geoffrey Pullum. You should be able to find the paper in your university library.

At the end of this page, there is a glossary of technical terms used in this paper that haven't been covered yet in this textbook.

The first pass

As discussed in Section 6.9, the first thing you should do when reading an academic article is to read the abstract and keywords carefully. The abstract of this paper isn't labeled, but is found in a smaller font right under the authors' affiliations. This paper doesn't have keywords, so we can skip that step.

Questions to ask while you're reading the abstract:

- Why do you think I picked this article to read? How is it related to the course content in general and to this chapter, specifically?
- Which ideas and terms from this chapter show up in the abstract?
- What year was this paper written?

Next, skim the introduction and conclusion. For this step in this paper, I would skim the unlabeled section at the beginning, Section 1: Background, and Section 5: Further Observations. When you're scrolling or flipping through the pages between sections 1 and 5, pay attention to the titles of the headers and the kinds of data. Since this paper has uninformative section headers, you may wish to read the first sentence or two of each section as well as the header. When you get to Section 5, you will probably notice that it is not a conclusion, but rather in-depth analysis and contextualization. At this point, I would start skimming to find the conclusion—it turns out the true conclusion is just the final paragraph of the paper.

Questions to ask while you're skimming the introduction, conclusion, headers, and data:

- What is the main conclusion of this paper?
- Which subfield(s) would you classify this paper into?
- Is this paper more descriptive or theoretical? That is, does it describe a property of a particular language or more generally argue for a certain kind of model of Language?
- How is this paper structured?
- What kinds of data are found in this paper?
- What technical terms terms show up a lot? Are there any that you don't know or that you need a refresher on?
- How is this paper different than the papers that came before? What new ideas did it introduce?
- Can you tell yet why this paper matters? What is its contribution to the field of linguistics?

The structure of Zwicky and Pullum (1983)

Here, I provide a reverse outline of this paper. A reverse outline is when you take a completed paper and you make an outline of it. This can be useful as a reading and study tool, but it can also be a useful to step for evaluating and revising the structure of paper drafts that you have written.

0. *unlabeled* (page 502)

- Context: Sometimes you can be surprised even by well-studied languages.
- Main claim: This paper discusses one surprise: that *n't* is an inflectional affix, not a clitic.

1. Background (pages 502-504)

- Describes and motivates 6 criteria for distinguishing between clitics and inflectional affixes.

2. Criteria A-D (pages 504-505)

- Demonstrates how the first four criteria apply to non-controversial examples of clitics and inflectional affixes.
- These first four criteria are all related by being about how the clitics and affixes combine with their hosts and stems.

3. Criteria E-F (pages 505-506)

- Demonstrates how the last two criteria apply to non-controversial examples of clitics and inflectional affixes.
- These last two criteria are related by being explained if cliticization happens later than other syntactic and morphological operations.

4. The contracted negator *n't* (pages 506-510)

- Applies the six criteria to *n't*.
- Demonstrates that *n't* behaves like an affix and not a clitic for all 6 criteria.

5. Further observations (pages 510-512)

- Uses cross-linguistic evidence to demonstrate that it is not surprising for there to be an inflectional affix marking negation.
- Argues against the alternative explanation that *n't* is a special clitic.

The second pass

Now go back to the beginning of the paper and read it through from beginning to end. In this guided reading section, we will do a deep dive into parts of Section 4, which is both the most important part of the paper and also where most of the more complex ideas are found. I will leave the rest for you to read on your own.

Beginning of section 4, page 506

At this point of the paper, the authors have presented 6 criteria for distinguishing inflectional affixes from clitics and have illustrated how those criteria work for some non-controversial examples of both inflection and clitics, to establish that these criteria are valid. Now, they are using these six criteria to test a more controversial morpheme: negative *n't*, and will show that it behaves like an inflectional affix. Note that one of the authors

of this paper, Zwicky, had previously adopted the opposite analysis—and has changed his mind when he considered the evidence more fully.

This section doesn't go through the criteria in the same order as the rest of the paper. Instead, they start with Criterion E: "Syntactic rules can affect affixed words, but cannot affect clitic groups." Back on page 504, the reason for this criterion is explained. They are working under the assumption that the model of grammar is structured in such a way that all syntactic operations occur before any phonological operations, and that cliticization *also* occurs after all syntactic operations. On the other hand, word formation processes, including inflection, happen *before* syntax.

(1) *Order of operations assumed by Zwicky and Pullum (1983)*

Module of grammar	Morphology	→	Syntax	→	Phonology
Operations	inflection		SAI		cliticization

To test this criterion, they look at Subject-auxiliary inversion (SAI), which is the process involved in yes-no question formation. SAI is unquestionably a syntactic operation, and so it should occur after inflection but before cliticization, as shown in (1). This predicts that inflectional affixes should move with the auxiliary in yes-no questions, as shown in (2). Here, different inflectional forms of DO can undergo SAI: *do, does*, and *did*.

(2)	a.	Do you want a cookie?
	b.	Does he want a cookie?
	c.	Did you want a cookie?

It also predicts that clitics should not move with the auxiliary, which is also what happens, as shown in (3) (Zwicky and Pullums's (8)).

(3)	a.	You could've been there.
	b.	*Could've you been there?

What does *n't* do? It can move, so it behaves like an affix according to criterion E. This is shown in (4) (Zwicky and Pullums's (4)), where it undergoes SAI.

(3) a. You haven't been there.

b. Haven't you been there?

Assumptions about a model of grammar

As shown in (1), this paper assumes a certain order of operations in the grammar. First, words are built using morphological rules in the lexicon. Then, the words are combined into phrases and sentences in the syntax component. Finally, the sentences are interpreted by phonology, where it is decided how the sentence should be pronounced. These assumptions show up in subtle ways throughout the paper, for example in their use of the word *lexicalization*. They don't explicitly state their assumptions, because this was the model of grammar that most generative linguists in the 1980s had adopted.

But this is not longer the case! In the early 1990s, two linguists, Morris Halle and Alex Marantz, proposed a new model of grammar called **Distributed Morphology** with different assumptions about the relationship between morphology and syntax. They noticed that there are many similarities between the rules that form words and the rules that form phrases, and so they argued that having one set of rules that takes care of *both* morphology and syntax is a simpler, and therefore better, model of grammar than one that has two separate components.

Distributed Morphology is now a very popular model of the architecture of grammar, but it has still not been adopted universally. A few syntacticians still assume a separate morphological component. Meanwhile, more models hypothesizing different relationships between morphology and syntax continued to be proposed.

If you would like to know more, the basics of Distributed Morphology are explained in a relatively straightforward way on this FAQ by Rolf Noyer.

Table 1, page 508

Zwicky and Pullum's Table 1 shows the complete list of auxiliaries that can host *n't* in the authors' grammars, providing evidence for criteria A, B, and C.. The table includes the bare form of the auxiliary, the pronunciation of the bare form, the form of the auxiliary with *n't*, and the pronuniciation of the form of the auxiliary with *n't*. Note that the transcriptions here aren't in standard IPA, which is common in older papers like this when typing unusual characters was more difficult.

Table 1: Non-standard transciption used in Zwicky and Pullum (1983)

Zwicky and Pullum's symbol	IPA symbol	Meaning
u	ʊ	high back lax rounded vowel
š	ʃ	voiceless post-alveolar fricative
n̥	n̩	syllabic alveolar nasal

Criteria A states that "clitics can exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems." In other words, clitics can usually attach to anything, but affixes are more picky. The fact that this table is even possible provides evidence that *n't* behaves like an affix—*n't* can only attach to auxiliaries and the full list of possible stems is short enough that it can fit on one page!

Criteria B states that "arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups." This table shows that *n't* behaves like an affix because it cannot attach to auxiliaries like *may* and *am*. There is no phonological reason why *may* should not be able to host *n't*—*mayn't* is a possible word in English—making this an arbitrary gap. The auxiliary *am* is a bit more difficult to pronounce with *n't*, but some dialects do allow it, suggesting that, again, this is an arbitrary gap.

Criteria C states that "morphophonological idiosyncrasies are more characteristic of affixed words than of clitic groups." In other words, there is more lexically conditioned **allomorphy** for affixes than for clitics. This is shown throughout the table, in the many differences in pronunication between the bare form and the affixed form. For example, the vowel in *do* /du/ changes in *don't* /dont/ and the /l/ in *shall* /ʃæl/ is dropped in *shan't* /ʃænt/. There is also not a consistent pattern that can predict whether the negative form is one syllable or two—for *can*, both options are possible!

Criterion D, page 509

Criterion D states that "semantic idiosyncracies are more characteristic of affixed words than of clitic groups," which means that words with affixes are more likely to have special non-predictable meanings than combinations of clitics and hosts. In this section, Zwicky and Pullum show that *not* and *n't* don't have the exact same meanings in all contexts, but the difference is subtle. This difference in meaning shows that *n't* behaves more like an affix than a clitic. This section is one of the trickiest to understand, especially if you haven't studied much semantics before.

The argument in this section depends on the notion of **scope**, which is a word used to describe which parts of a sentence are modified by a semantic operator. In this case, the semantic operator is negation. Compare sentences (4a) and (4b). In (4a), the word *not* only modifies *rotten*. That is, the scope of negation is only the word *rotten*. In (4a), Hentie did eat a fruit, but the fruit was fresh. In (4b), on the other hand, the word *not*

modifies the entire sentence; in other words, *not* takes scope over the entire sentence. In (4b), we do not know whether Hentie ate a fresh fruit. All we know is that the action of eating a rotten fruit did not occur.

(4)	a.	The fruit Hentie ate was not rotten.
	b.	It is not the case that Hentie ate a rotten fruit.

When there are two semantic operators in the same clause, they interact. Which order you compose them changes the meaning of the sentence. For example, in (5), we see the difference between must(not(P)) and not(must(P)). In semantics, P is used as a placeholder for a **proposition**, which is a semantics term meaning a statement that can be true or false.

(5)	**semantic notation**	**meaning**	**English example**
a.	must(not(P))	It must be that case that not P.	You must not come early. You are required to stay away until the appropriate time, because we will be fumigating the room, and it will not be safe to enter.
b.	not(must(P))	It is not the case that P must be.	It is not necessary for you to come early. I don't need your help to set up, but you can come early if you want to.

The difference in scope between *not* and *n't* is illustrated in (6) and (7) (from in-text examples from Zwicky and Pullum 1983: 509). The modal *must not* only has one possible scope meaning. The sentence *you must not go home* can only mean that you are not allowed to go to home. The sentence with *n't* instead of *not* has the same meaning.

(6)		must(not(P))	not(must(P))
a.	You must not go home.	✓	✗
b.	You mustn't go home	✓	✗

The pattern is different with *can*, though. In these sentences, an accent (é) is used to mark stress. With the full-word form in (7a), the sentence is ambiguous. It can mean that you have the option of not going home, but it can also mean that you are not allowed to go home. The sentences with *n't* in (7b) and (7c), on the other hand, only have one possible meaning; they can only mean that you are not allowed to go home, which is the opposite scope order that was allowed with *mustn't*.

(7)		can(not(P))	not(can(P))
a.	You can nót go home.	✓	✓
b.	You cánnot go home	✗	✓
c.	You can't go home.	✗	✓

The relationship between *must not* and *mustn't* is predictable—they share the same pattern as each other. In contrast, the relationship between *can not* on the one hand and *can't* and *cannot* on the other is not predictable. We then have a semantic idiosyncracy in the meaning of a word using *n't*, indicating that *n't* behaves more like an affix than a clitic.

Don't mix up **preposition** with an E and **proposition** with an O. Both are technical terms in linguistics, but are very different things! A preposition is a part of speech that typically indicates location, such as *in*, *on*, *at*, or *under*. A proposition, on the other hand, is a statement that can be true or false, such as *I enjoy eating cake*.

Reading with a writer's eye

As you read this paper, did you think of any ways that you would write it differently that would make it clearer and easier to read? Doing so is a good way to become a better writer.

Here are some changes that I think would make this paper a bit easier to read:

- I'd give the headers more transparent headings. For example, instead of labeling Section 2 "Criteria A-D," I'd label it something like "Relationship between a bound morpheme and its host."
- Throughout Sections 2 and 3, I'd make sure there was at least one example of an inflectional

affix and of a clitic for each criteria, as much as possible.
- I'd repeat the relevant criteria in Section 4 so the reader doesn't need to scroll back to find them.
- At the beginning of Section 4, I'd tell the reader I was discussing criterion E, instead of leaving it for the reader to infer.

What about you? Do you have any suggestions for how to make this paper clearer?

Glossary of terms used in this paper

The following is a list of technical terms used in this paper that have not yet been introduced in this textbook, as well as some advanced words used in scientific discourse that aren't quite technical terms.

Agreement: When the features on one word gets transferred to or shared with another word.

Auxiliary verbs: Verbs like *have*, *be*, and *do* that appear before a main verb.

Conjunctive: Having the meaning of 'and'.

Constituent: A group of words behaving like a syntactic unit.

Copula: A verb used to introduce non-verbal **predicates**, such as *am* in *I am a linguist*.

Feed: When one rule or process must occur before another one, the first rule is said to feed the second. This is a term borrowed from **phonology**.

Finite verbs: Verbs which are inflected for **tense** (as opposed to infinitive verbs, which are not).

Formative: A word you can use in morphology to describe a linguistic element when you don't want to be specific about whether it is a word, a morpheme, or something else.

Government: A particular structural relationship in syntax, largely no longer in use.

Emphatic stress: When you pronounce a word more forcefully for emphasis. When a word has a clitic variant, the clitic usually can't be used if there is emphatic stress on it.

Idiosyncracy*:* A pattern or feature that is unique to an individual. In this paper, the individual is the morphological **stem**. Attachment of an **affix** to a stem is more likely to have exceptions or unusual patterns than attachment of a clitic to a host.

Inchoative/causative verb: An inchoative is a verb expressing that a certain state has begun. A causative is a verb indicating that the subject caused a resulting state. For example, verbs such as *moisten* and *soften* mean 'cause to become moist/soft.'

Lexical items: Items formed in the **lexicon**. That is, they are formed by morphological rules, not syntactic rules.

Lexicalization: The process by which words are added into the **lexicon**. The paper seems to be referring,

here, to how auxiliary + *n't* constructions were previously combined syntactically but were so frequent in combination that they got added to the lexicon. Once they had their own lexical entries, the combinations were able to pick up idiosyncratic meanings and forms.

Monosyllables: Words which contain only one syllable.

Neologism: Newly-formed word.

Node: The point at which several branches come together in a syntax tree.

Parochial: Limited in use and scope.

Readjustment rule: A rule that adjusts the output of syntactic rules to prepare it for the phonology.

Selection: When a word or morpheme is "picky" about what things can attach to it to form a new word or a phrase.

Sentential negation: When the entire sentence is negated, not just a word or a phrase inside the sentence.

Simple clitic: A clitic which has an equivalent full-word form that can appear in the same positions.

Special clitic: A clitic which either has no equivalent full-word form or it has an equivalent full-word form that appears in different positions.

Stipulate: To add an assumption to our model without any motivation other than that it makes the data work out.

Subject-auxiliary inversion (SAI): The word order change involved in the formation of yes-no questions in English, in which the subject and the auxiliary swap positions. For example, the statement "You are awake" becomes the yes-no question "Are you awake?" when the subject *you* swaps position with the auxiliary *are*.

Superlative: An inflectional affix on adjectives to mean 'most'.

Syllabic: Behaving as the nucleus of a syllable (i.e., instead of a vowel).

Syncategoremic: Something that doesn't quite have meaning on its own, but only has meaning when combined with something else.

Transformation: A rule of grammar that rearranges the order of words, used in some grammatical models of the 1960s-1980s. They have generally been replaced by movement rules in many modern models of grammar.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=4108#h5p-52

References and further resources

Academic sources

Noyer, Rolf. n.d. Distributed Morphology: Frequently Asked Questions list. https://www.ling.upenn.edu/~rnoyer/dm/home.html

Zwicky, Arnold, and Geoffrey Pullum. 1983. Cliticization vs. inflection: English N'T. *Language* 59 (3): 502-513.

7. PREDICATES, ARGUMENTS, AND CLAUSES

placeholder

7.1. PREDICATES

A clause or simple sentence typically consists of a **subject** and a **predicate**. The subject is who or what the sentence is about and the predicate is what the sentence is asserting about the subject. We will learn more about subjects in Section 7.2.

In the sentences in (1), the subject is underlined and the predicate is in bold.

(1)	a.	My friend **rides the bus to campus.**
	b.	Cats **like milk.**
	c.	Some professors are **strict.**
	d.	The library is **down the street.**
	e.	Taylor Swift is **a singer-songwriter.**

The word predicate can be used in two main ways. Sometimes it can refer to just the **head** of the predicate (usually a verb), or it can refer to the whole phrase. In (1), the whole phrases are bolded.

Although most predicates are verbs, such as *ride* in (1a) and *like* in (1b), there are also non-verbal predicates. Non-verbal predicates include adjectival predicates, as in (1c), prepositional predicates, as in (1d), or nominal predicates, as in (1e). In English, non-verbal predicates always appear with the verb BE, which is known as the **copula**. The copula does the grammatical job of bearing tense and agreement marking, since adjectives, prepositions, and nouns can't. Although there are multiple kinds of copular constructions with slightly different meanings, you can think of the copula roughly as an equal sign, indicating that the subject has the property denoted by the non-verbal predicate.

Although in English non-verbal predicates always have a copula, some other languages have what is known as a zero copula. In these languages, there may be no verb at all when there is a non-verbal predicate. Some examples of nominal predicates with zero copula from Sinhalese (Indo-European language family, spoken in Sri Lanka), Tubu (Saharan language family, spoken in Niger and Chad), and Pitjantjatjara (Pama-Nyungan family, spoken in Australia) are shown below.

(2) a. Sinhalese

unnaehee	*hungak*	*presidde*	*kene-k*
he	very	famous	person-nom

'He is/ was a very famous person.'

(Gair 1970: 145, as cited in Stassen 2013)

b. Tubu

sígən	*līfi*
3sg.emph	orphan

'He is/was an orphan.'

(Lukas 1953: 170, as cited in Stassen 2013)

c. Pitjantjatjara

wait	*nglayayala*
man	doctor

'The man is/ was a doctor.'

(Doughlas 1959: 55, as cited in Stassen 2013)

Some other languages have zero copula sometimes but an overt copula in other contexts. For example, Maltese, an Afro-Asiatic language spoken in Malta, has a zero copula in the present tense and an overt copula in the past tense.

(3) Maltese

a. *Albert* *tabib*

Albert doctor

'Albert is a doctor.'

b. *Albert* *kien* *tabib*

Albert be.3sg.m.pst doctor

'Albert was a doctor.'

(Stassen 2013)

According to WALS, just over half of the world's languages always use an overt copula (Stassen 2013).

Key takeaways

- Sentences consist of a subject and a predicate.
- The predicate is the part of the sentence that asserts something about the subject.
- Most predicates are verbs, but there are also adjectival, nominal, and prepositional predicates.
- Non-verbal predicates are introduced by a copula in many languages.
- Some languages use a zero copula, either all of the time or limited to some contexts.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=4619#h5p-55

References and further resources

Academic sources

Stassen, Leon. 2013. Zero copula for predicate nominals. In *World Atlas of Language Structures Online*, ed. Matthew Dryer and Martin Haspelmath. https://wals.info/chapter/120

7.2. ARGUMENTS

Most verbs can't stand on their own. Some verbs describe the properties of one entity, while other verbs describe the relationship between several entities. Because of this, verbs often require certain phrases be included with them in a sentence. The requirement is called **selection**, and the phrases that are required are called **arguments**.

Transitivity

We classify verbs based on how many arguments they require, which is called transitivity. An **intransitive** verb has one argument, typically the subject. A **transitive** verb has two arguments, typically a subject and direct object. A **ditransitive** verb has three arguments, typically a subject, direct object, and indirect object.

Table 1. Kinds of transitivity

Kind of verb	Number of arguments	Typical arguments	Example
Intransitive	1	Subject	*The cat slept.*
Transitive	2	Subject and direct object	*The cat chased a mouse.*
Ditransitive	3	Subject, direct object, and indirect object	*The cat left me a dead mouse.*

Subjects

The **subject** is typically (but not always) the one performing the action. Subjects in English can be identified because they go before the verb and because they trigger agreement on the verb in the present tense. If the subject is singular, the verb will be singular, as shown in (1a) where *the cat* is singular, so we get the singular form of the verb *is*. If the subject is plural, the verb will be plural. This is shown in (1b), where the subject *the cats* is plural and so we get the plural form of the verb *are*.

(1)	a.	The cat is purring.
	b.	The cats are purring.

If you're not sure what the subject is in English, you can switch a sentence to present tense and look for subject-verb agreement.

The many facets of subjecthood

In a prototypical sentence, a subject has many properties:

- The subject appears in a particular position. In English, the subject comes before the verb.
- The subject agrees with the verb.
- The subject is what the sentence is about.
- The subject is the one performing the action.
- The subject is morphologically marked as being the subject.

However, a subject doesn't have all of these properties in all sentences. For example, in (2), the sentence is about lunch, but the subject is *we*.

(2) As for lunch, we can eat some sandwiches.

In **passive** constructions, the entity that performs the action is deleted or included in a *by*-phrase. The subject of a passive is the entity that undergoes the action. This is shown in (3). An **active** sentence is in (3a), with the one performing the action, *we*, in subject position. In the passive sentence in (3b), however, the one performing the action is included in an optional *by*-phrase. The entity that undergoes the action, s*ome sandwiches*, is in subject position.

(3) a. We ate some sandwiches.
b. Some sandwiches were eaten (by us).

Finally, sometimes the subject has a special form. This happens for some pronouns in English, as we already saw in Section 2.1. The first person plural pronoun appears as *we* in subject position but *us* in other positions, as shown in (4).

(4) a. **We** gave you a gift.
b. You gave **us** a gift.

However, the subject of a **non-finite** clause does not appear in the special form for subjects. If we turn the clause in (4a) into a non-finite clause, as in (5), we get *us* not *we*.

(5)	a.	[For **us** to give you a gift] would be kind.
	b.	*[For **we** to give you a gift] would be kind.
	c.	They want [**us** to give you a gift].
	d.	*They want [**we** to give you a gift].

When we discuss the subject, we are usually referring to the structural position—in English, this is the position before the verb. If we want to refer to a phrase that has some of these other properties, we have other words for that. For example, what the sentence is about is called the *topic*, the one performing the action is called the *agent theta role*, and the morphological form associated with subjects is called ***nominative case.*** We will learn more about theta roles and case later in this textbook.

Objects

In English, objects usually appear after the verb. There are two kinds of object: **direct objects** and **indirect objects**. If there is only one object, it is a direct object, as shown in (6). Direct objects typically undergo the action described by the verb.

(6) The cat drank some milk.

In (6), the verb is *drank*. The subject is *the cat* and the direct object is *some milk*.

Ditransitive verbs have two objects. Ditransitive verbs usually describe a change of possession. The direct object undergoes the action. That is, it is the entity that switches possession. The indirect object is usually the entity that has possession of the object at the end of the action.

The two objects in ditransitive verbs can appear in either order, but the two orders have some structural differences. The sentences in (7), which have the same meaning, illustrate this difference.

(7)	a.	My sister gave me her cat.	*double object construction*
	b.	My sister gave her cat to me.	*prepositional ditransitive*

The sentence in (7a) is called the **double object construction**. In (7a), both objects are noun phrases. The indirect object, who receives the entity being transferred, is *me* and appears first. The direct object, which is the entity being transferred, is *her cat* and appears second.

The sentence in (7b) is called the **prepositional ditransitive**. In prepositional ditransitives, the direct

object *her cat* appears first. It is a noun phrase. The indirect object *to me* comes second. It is a prepositional phrase.

If you're not sure which object is the direct object and which is the indirect object, rephrase the sentence so that it uses a preposition. The one that comes with a preposition is the indirect object.

The differences between the two ditransitive constructions are summarized in Table 2.

Table 2. Ditransitive constructions

Construction	**First object**	**Second object**
Double object construction	NP indirect object	NP direct object
Prepositional ditransitive	NP direct object	PP indirect object

Adjuncts

Many sentences also include many phrases that aren't selected for by the verb or by any other word in the sentence. These are optional and add extra information. They are called **adjuncts**. Some examples of adjuncts are shown in (8). In these examples, the adjuncts are in parentheses. This shows that they are optional, since the sentence is still grammatical if we leave them out.

(8) a. The cat slept (on the porch).

b. The (black) cat slept.

c. The cat slept (peacefully).

d. The cat slept (because it was tired).

Key takeaways

- A verb has requirements for phrases it must appear with, which is called selection. The phrases it selects are called arguments.
- We classify verbs based on how many arguments they have, which is called transitivity. Intransitive verbs take one argument, transitive verbs take two arguments, and ditransitive

verbs take three arguments.
- Subjects appear before the verb and trigger subject-verb agreement.
- Objects appear after the verb. Direct objects undergo the event and indirect objects are the endpoint of the event.
- Ditransitives allow the objects to appear in either order. If the indirect object comes first and both objects are NPs, it is called a double object construction. If the indirect object comes second and is a PP, then it is a prepositional ditransitive.
- Phrases that aren't selected for by the verb are called adjuncts.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://pressbooks.openedmb.ca/wordandsentencestructures/?p=4622#h5p-56

7.3. CLAUSES

Consider the following sentences.

(1)	a.	I think [that spaghetti is delicious].
	b.	We should have spaghetti for supper [because it is delicious].
	c.	[Spaghetti is delicious] and [garlic bread is delicious, too].
	d.	[That spaghetti is delicious] is obvious.
	e.	We should have spaghetti [which is delicious] for supper.
	f.	[Because spaghetti is delicious], we should have it for supper.
	g.	I want [to have spaghetti for supper].
	h.	[Having spaghetti for supper] is a good idea.
	i.	We will eat some spaghetti [when it is ready].

All of the sentences in (1) have two clauses. Look at the parts of the sentences that are in square brackets. Each of them has its own verb and arguments and can stand on its own as a sentence (if you take out the *that* and *because*), even though it is also part of a larger sentence. These are all clauses.

A clause consists of a predicate and its arguments. All of the sentences in (1) have two clauses, but a sentence can have more than two. The sentence in (2) has six clauses! Sentence (2) rambles a bit and is repetitive, but it is grammatical.

(2)	[I **think** [that we should **have** spaghetti, [which **is delicious**], for supper]] and [we should **have** garlic bread, too,] [because I really **want** [to **have** spaghetti and garlic bread]].

The easiest way to find all of the clauses in a sentence is to find the predicates. There will normally be one clause for each predicate. In (2), I have bolded the predicates of each clause. Then, find the arguments and adjuncts that belong to each predicate. Sometimes a clause has a special word to introduce the clause, like *that* and *because*. These are called complementizers and belong to the clause, too. There may also be a *wh*-word that belongs with the clause.

Serial verb constructions

Some languages allow a construction called a **serial verb** construction, in which multiple verbs form a single predicate. We can tell that they all belong to the same predicate because they share the same arguments—the same subject and the same object. Below are some examples of serial verb constructions from Tabas, an Austronesian language spoken in Indonesia; Dâw, a Makú language spoken in Northwest Amazonia; and Igbo, a Niger-Congo language spoken in Nigeria.

(3) a. Tabas (Bowden 2001: 297, as cited in Aikhenvald and Dixon 2006: 2)
n=babas welik n-mot do
3SG=bite pig 3SG=die REAL
'It bit the pig dead.'

b. Dâw (Aikhenvald and Dixon 2006: 2)
yõ:h bə:-hãm-y **ɔ** w
medicine spill-go-happen.straight.away
'The medicine spilt straight away.'

c. Igbo (Lord 1975: 27, as cited in Aikhenvald and Dixon 2006: 2)
ó tì-wà-rà étéré à
he hit-split.open-TENSE plate the
'He shattered the plate.'

Because there are multiple verbs in a serial verb construction, we might think that there are multiple clauses, too. However, since all of the verbs in a serial verb construction still belong to the same predicate, we can still say that there is one clause for each predicate.

Serial verb constructions have different properties in different languages. For example, the verbs in a serial verb construction might all be together in one word, as in (3b) and (3c), or might be separated into multiple words, as in (3a). They might all be next to each other, as in (3b) and (3c), or there might be words in between the different verbs, as in (3a). Inflection might appear on all of the verbs in the construction, as in (3a) or just on one of the verbs, as in (3c).

There is more than one kind of clause. Clauses are generally classified based on their relationship to other parts of the sentence. Some clauses also have special structure and can be classified based on their internal structure. Over the next few sections, we will learn how to classify clauses based on these two properties.

Main and embedded clauses

If you look at the examples of clauses in (1), you may notice that some clauses are inside of other clauses. The ones that are inside are called embedded clauses. You may also see them referred to as subordinate or dependent clauses. The clauses that they are inside of are called main clauses. They can also be called independent, matrix, or root clauses.

Let's look again at example (1a), repeated below in (3). In this sentence, the embedded clause is *that spaghetti is delicious*. The main clause is *I think that spaghetti is delicious.* The embedded clause is the object of the main clause verb *think*, which makes the entire embedded clause part of the main clause.

(3)		I think [that spaghetti is delicious].

Argument clauses

Some clauses are arguments of a verb. As we already saw, the embedded clause in (3) is the object of the main clause verb. An embedded clause can also be in subject position. This is the case in (1d) and (1h), repeated below in (4).

(4)	a.	[That spaghetti is delicious] is obvious.
	b.	[Having spaghetti for supper] is a good idea.

In (4a), the embedded clause *that spaghetti is delicious* is the subject of the predicate *is obvious*. In (4b), the embedded clause *having spaghetti for dinner* is the subject of the predicate *is a good idea*.

Adjunct clauses

Clauses can also be adjuncts. Remember that adjuncts are optional modifiers. Adjunct clauses can modify the entire main clause, a verb, or a noun. Some examples of adjunct clauses are found in (1b), (1e), (1f), and (1i), and are repeated below in (5). In (5a) and (5b), the embedded clause *because it is delicious* modifies the entire main clause. In (5c), the embedded clause *which is delicious* modifies the noun *spaghetti*. In (5c), the embedded clause *when it is ready* modifies the verb *eat*. All of these embedded clauses are adjuncts; they are not required to form a complete sentence.

(5) a. We should have spaghetti for supper [because it is delicious].

b. [Because spaghetti is delicious], we should have it for supper.

c. We should have spaghetti [which is delicious] for supper.

d. We will eat some spaghetti [when it is ready].

Kinds of sentences

We can also classify sentences based on how many clauses it has and how the clauses are combined. If a sentence only has one clause, it is called a simple sentence. The single clause in a simple sentence will always be a main clause. Some examples of simple sentences are shown in (6).

(6)	a.	Spaghetti is delicious.
	b.	We should have spaghetti for supper.
	c.	A spaghetti dinner is obvious.

If a sentence has two or more conjoined main clauses, it is called a compound sentence. Example (1c), repeated below as (7), is an example of a compound sentence.

(7) [Spaghetti is delicious] and [garlic bread is delicious, too].

Finally, if a sentence has one or more embedded clauses, it is called a complex sentence. A complex sentence will also always have at least one main clause, as well. All of the sentences in (1) *except* the compound sentence in (1c) are complex sentences. The complex sentences from (1) are repeated below in (8).

(8) a. I think [that spaghetti is delicious].

b. We should have spaghetti for supper [because it is delicious].

c. [That spaghetti is delicious] is obvious.

d. We should have spaghetti [which is delicious] for supper.

e. [Because spaghetti is delicious], we should have it for supper.

f. I want [to have spaghetti for supper].

g. [Having spaghetti for supper] is a good idea.

h. We will eat some spaghetti [when it is ready].

Key takeaways

- A clause consists of a predicate, all of the arguments and adjuncts that modify the predicate, and sometimes a complementizer.
- A main clause can stand on its own, while an embedded clause is inside of a main clause.
- An embedded clause can be in subject, object, or adjunct position.
- A simple sentence has one main clause, a compound sentence has two or more main clauses, and a complex sentence has one or more embedded clauses.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=4625#h5p-57

References and further resources

Academic sources

Aikhenvald, Alexandra, and R. M. W. Dixon. 2006. *Serial verb constructions: A cross-linguistic typology*. Oxford: Oxford UP.

7.4. STATEMENTS, QUESTIONS, AND COMMANDS

Besides classifying clauses based on their structural position, we can also classify them based on their function. We will discuss three kinds of function here: declaratives, interrogatives, and imperatives. Note that the main clause and the embedded clause of the same sentence might have different functions.

Declaratives

Declarative clauses make a statement. Most clauses are declarative.

(1)	a.	I like chocolate.
	b.	My father was going to the store yesterday.
	c.	The plane will land in Frankfurt tomorrow.

In English, embedded declarative clauses can be introduced with the complementizer *that* or no complementizer at all.

(2) a. It's not unusual [that I like chocolate].
b. Jennie told you [my father was going to the store yesterday].
c. I expect [that the plane will land in Frankfurt tomorrow].

Interrogatives

Interrogatives clauses make a question. There are two main kinds: polar questions and content questions.

Polar questions

Polar questions, also called yes-no questions, are questions that can be answered with *yes*, *no*, or *maybe*. In

English, main clause polar questions are formed by subject-auxiliary inversion. That is, the auxiliary verb and the subject switch places compared to the equivalent declarative clause.

(3) a. Do I like chocolate?
b. Was my father going to the store yesterday?
c. Will the plane land in Frankfurt tomorrow?

English embedded polar questions do not have subject-auxiliary inversion and are introduced by the complementizers *if* or *whether*.

(4) a. You wonder [whether I like chocolate].
b. Jennie asked [if father was going to the store yesterday].
c. I'm not sure [whether the plane will land in Frankfurt tomorrow].

According to WALS, only about 1% of languages form polar questions by changing the order of words, like English does. The most common way of forming polar questions is to use a question particle, which is the strategy used by 61% of the languages in the sample on WALS. An example of a question particle in Maybrat, a language isolate of Indonesia, is shown in (5). In Maybrat, the question particle *a* appears at the end of the clause.

(5) Maybrat (Dol 1999: 200, as cited in Dryer 2013)
ana m-amo Kumurkek a
3PL 3-go Kumurkek Q
'Are they going to Kumurkek?'

Other common strategies for marking polar questions in the world's languages are to use intonation (18%) or verbal morphology (17%) (Dryer 2013). An example of a language that uses verbal morphology is Hunzib, a Nakh-Daghestanian language spoken in the North Caucasus region of Russia, near the border with Georgia, as shown in (6).

(6) eǎ'e-čó-**y**
go-PRES.1/2-**Q**
'Are you going?'

Content questions

Content questions, also called *wh*-questions, are questions that cannot be answered with *yes*, *no*, or *maybe*. In English, main clause content questions are formed by subject-auxiliary inversion and by moving a question word to the beginning of the clause. The question words in (7) are in bold. Question words are also often called *wh*-words.

(7) a. **Who** likes chocolate?
b. **What** do I like?
c. **Where** was my father going yesterday?
d. **Why** was my father going to the store yesterday?
e. **Whose** father was going to the store yesterday?
f. **When** will the plane land in Frankfurt?
g. **How** will the plane land in Frankfurt tomorrow?
h. **Which** airport will the plane land at tomorrow?

Embedded content questions in English do not have subject-auxiliary inversion, but the question word still moves to the beginning of the clause.

(8) a. I asked [**who** likes chocolate].
b. They wondered [**what** I like].
c. We wanted to know [**where** my father was going yesterday].
d. I questioned [**why** my father was going to the store yesterday].
e. They asked [**whose** father was going to the store yesterday].
f. We were wondering [**when** the plane will land in Frankfurt].
g. My kid is asking [**how** the plane will land in Frankfurt tomorrow].
h. They need to know [**which** airport the plane will land at tomorrow].

Many other languages also move the question word to the beginning of the clause to form content questions, but in other languages, the question word stays in its argument position. This is called *wh*-in situ. Mandarin is one example of a *wh*-in situ language, as shown in (9). The declarative sentence in (9a) has the same word order as the interrogative sentence in (9b), unlike the English equivalents.

(9)	a.	Yuehan xihuan Mali.
		John like Mary
		'John likes Mary.'
	b.	Yuehan xihuan shei?
		John like who
		'Who does John like?'

(Dong 2009: 11)

Imperatives

Imperatives make a command. In English, imperatives have a silent subject and the verb appears in its bare form.

(10)	a.	Eat chocolate!
	b.	Go to the store!
	c.	Take a plane to Frankfurt tomorrow!

Mixing functions

A complex sentence can have main and embedded clauses with different functions, as shown in Table 1.

Table 1. Combinations of main and embedded clause functions

	Main declarative	**Main polar question**	**Main content question**	**Main imperative**
Embedded declarative	Ben said [that Ava is hungry].	Did Ben say [that Ava was hungry]?	When did Ben say [that Ava was hungry]?	Tell them [that Ava is hungry]!
Embedded polar question	Ben wondered [if Ava was hungry].	Did Ben ask [if Ava was hungry]?	When did Ben wonder [if Ava was hungry]?	Ask [if Ava is hungry]!
Embedded content question	Ben asked [what Ava wants for lunch].	Did Ben wonder [what Ava wants for lunch]?	Why did Ben ask [what Ava wants for lunch]?	Ask [what Ava wants for lunch]!

Key takeaways

- Declarative clauses make a statement, interrogative clauses ask a question, and imperative clauses give a command.
- Interrogative clauses come in two types: polar and content questions.
- A complex sentence can have main and embedded clauses with different functions.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=4712#h5p-58

References and further resources

Academic sources

Dryer, Matthew. Polar questions. In *World Atlas of Language Structures Online*, ed. Matthew Dryer and Martin Haspelmath. https://wals.info/chapter/116

Hongyuan, Dong. 2009. Issues in the semantics of Mandarin questions. PhD thesis, Cornell University.

7.5. RELATIVE CLAUSES

Relative clauses are a kind of clause that modifies nouns. Some examples are shown in (1).

(1)	a.	The witch trapped **Rapunzel**, [who __ had very long hair], in a tower.
	b.	**The princess** [__ trapped in the tower] had very long hair.
	c.	**The princess** [that the witch had kidnapped __] was very lonely.
	d.	**The tower** [where the princess was trapped __] was isolated in the woods.

Relative clauses are always adjuncts to a noun phrase. In (1), the nouns that are modified by the relative clauses are bolded. Any noun phrase can be modified by a relative clause.

Relative clauses also always have a gap. In (1), the gaps are marked with underscores. The gap can be in subject position, as in (1a) and (1b), in object position, as in (1c), or in adjunct position, as in (1d), but the gap will always have the same meaning as the noun that the relative clause modifies. We can illustrate this by rephrasing the complex sentences in (1) as two simple sentences, as shown in (2).

(2)	a.	The witch trapped **Rapunzel** in a tower. **Rapunzel** had very long hair.
	b.	**The princess** had very long hair. **The princess** was trapped in the tower.
	c.	**The princess** was very lonely. The witch had kidnapped **the princess**.
	d.	**The tower** was isolated in the woods. The princess was trapped **in the tower**.

Relative clauses may use complementizer *that*, as in (1c), a question word such as *who* in (1a) or *where* in (1d), or neither, as in (1b). If it has neither a complementizer or a question word, it is called a reduced relative clause. Interestingly, it is not possible to use both *that* and a question word. This is shown in (3). The relative clause can be introduced by *who* as in (3a) or *that* as in (3b), but cannot be introduced by both at the same time, in either order, as shown in (3c)-(3d).

(3) a. The witch trapped **the princess**, [who __ had very long hair], in a tower.

b. The witch trapped **the princess** [that __ had very long hair] in a tower.

c. *The witch trapped **the princess** [who that __ had very long hair] in a tower.

d. *The witch trapped **the princess** [that who __ had very long hair] in a tower.

Relative clauses have very similar structure as embedded content questions. In fact, relative clauses that use question words look identical to embedded content questions—the only difference is their position in the main clause. Relative clauses modify nouns, but embedded content questions are in the argument position of a verb.

(4) a. The witch trapped **the princess**, [who __ had very long hair], in a tower.

b. I wondered [who __ had very long hair].

(5) a. **The tower** [where the princess was trapped __] was isolated in the woods.

b. I asked [where the princess was trapped __].

Factive clauses

Not all clauses that modify noun phrases are relative clauses. Another kind of clause, called a factive clause, also modifies noun phrases. Unlike relative clauses, factive clauses do not have a gap. Some examples of factive clauses are shown in (6).

(6) a. The fact [that Rapunzel was trapped] makes me sad.

b. I wonder when Rapunzel got the idea [that she could escape].

Factive clauses have the same structure as embedded declarative clauses. The only difference is that they modify a noun, instead of being in the argument position of a verb.

Key takeaways

- Relative clauses are adjuncts to a noun phrase and have a gap.
- Relative clauses have similar structure to embedded content questions.
- Factive clauses also modify noun phrases but do not have a gap.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=4704#h5p-59

CHAPTER 7 PRACTICE EXERCISES

Data analysis

Exercise 1. Predicates

Intermediate [7.1, 7.2, 7.3]

Determine how many clauses are in each of the following sentences. Then identify the predicates in each clause. For each predicate, determine if it is verbal, nominal, adjectival, or prepositional. For verbal predicates, determine its transitivity.

a. The members of the other team were the winners.

b. The hockey player shot the puck.

c. Every player skated.

d. Most players were on the ice.

e. The losing team was sad.

Exercise 2. Verb arguments and adjuncts

Beginner-Intermediate [7.2]

Identify the subjects, direct objects, indirect objects, and adjuncts to the verbs in the following sentences.

a. The player passed the puck to his teammate at the last second.

b. The players skated quickly towards the net.

c. The skates were sharp.

d. The coach celebrated because her team won the game.

Exercise 3. Simple, complex, and compound sentences

Beginner [7.3]

Determine whether the following sentences are simple, complex, compound, or both complex and compound.

a. It was very cold in the hockey arena.

b. The goalie blocked the puck and his team cheered.

c. You can become a good skater if you practice a lot.

d. The winning team won because they had better teamwork.

e. The people in the audience enjoy hockey but they dislike how cold the arena is.

f. The team that lost was sad.

Exercise 4. Embedded clauses

Intermediate [7.3]

Put square brackets around the embedded clauses in the following sentences. Classify each embedded clause as a subject, object, or adjunct clause.

a. I knew that my favourite team won.

b. If the home team scored more than three goals, a local restaurant would give out free pizza.

c. That the teams went to a shootout didn't surprise me.

d. Some people play hockey because it's fun and other people play hockey because it's good exercise.

Exercise 5. Declaratives, questions, and imperatives

Beginner [7.4]

Determine whether each of the clauses in the following sentences are declaratives, polar questions, content questions, or imperatives.

a. Pass me the puck!

b. You should pass me the puck.

c. Why didn't you pass me the puck?

d. Did you pass the puck to someone else?

e. Do you know why you passed the puck to someone else?

f. Why did you ask if I passed the puck to someone else?

g. The referee wondered if the puck went into the net.

h. How did the referee decide that the puck went in?

Exercise 6. Relative clauses

Advanced [7.5]

Identify the relative clauses in the following sentences, if any. For each relative clause, circle the noun phrase that it modifies and put an underscore marking the gap.

a. We had never heard of the team that lost the game.

b. The team that had better teamwork won.

c. The player whom the team trained scored a goal.

d. The goalie whose team had the puck for most of the game didn't move a lot.

e. She was the coach that the team was looking for.

f. The idea that they might win the tournament motivated the team to practice.

g. The player with whom the opposing goalie had a rivalry was always happy to score.

h. The team with better players does not always win.

Exercise 7. Clauses without tense

Advanced [7.6]

Identify the *to*-infinitives, gerund clauses, and small clauses in the following sentences. Note that sentences may contain clauses of more than one type.

a. We called him *Puck McDuck*.

b. The players needed to practice more.

c. Playing ball hockey is a way for people to play hockey in warmer weather.

d. We consider the coach the best in the city.

e. She encourages newcomers to improve by skating around the rink.

Communication and study skills

Exercise 8. Polar questions

Advanced [7.4]

Read WALS chapters 116 and and 92 about polar questions and write a paragraph describing the ways that polar questions can vary across languages.

Research and application

Exercise 9. Relative clause positions

Advanced [7.4]

Some languages put relative clauses before the noun they modify and other languages put relative clauses after the noun they modify, as shown in map 90A on WALS. Look at what other word order patterns are described by maps on WALS by going to the Chapters section then filtering for "Word Order" in the Area column. Make a hypothesis about whether another word order pattern might be correlated with the word order between nouns and relative clauses. Then cross-reference the two maps. Do you notice any patterns?

To cross-reference two maps on WALS, follow these steps:

1. Open one of the two maps.
2. Above the map, it should say, "You may combine this feature with another one. Start typing the feature name or number in the field below." Below that, there should be a box with the map you're looking at listed. Type the number of the second map you want to

look at in the box, below the name of the first map.
3. Click submit.

Here is an example of map 83A cross-referenced with map 90A.

Language Journal

At the end of Chapter 1, we introduced the Language Journal. You should have picked a language to build a profile for as you work through this textbook. You can add to your profile with these questions.

Exercise 10. Overt and zero copulas

Intermediate [7.1]

Does your language use overt or zero copulas? Give examples to back up your answer.

Exercise 11. Serial verb constructions

Advanced [7.3]

Does your language use serial verb constructions? Give an example to back up your answer.

Exercise 12. Polar and content questions

Intermediate [7.4]

How does your language form polar questions and content questions? Give examples to back up your answer.

Exercise 13. Relative clauses

Advanced [7.5]

What do relative clauses look like in your language? Give examples to back up your answer.

Exercise 14. Non-finite clauses

Advanced [7.5]

What do non-finite clauses look like in your language? Give examples to back up your answer.

8. PARTS OF SPEECH

[Placeholder.]

8.1. DETERMINING PART OF SPEECH

The **part of speech** of a word, also called its syntactic or lexical category, is a classification of its behaviour. Some examples of parts of speech include nouns, verbs, adjectives, adverbs, prepositions, and so on. You likely learned a little bit about these categories in school, with the definitions described in Table 1. These traditional definitions are also sometimes called *semantic* definitions, since they classify part of speech based on the meaning of words. But, as you will see shortly, the traditional semantic definitions have some problems. In linguistics, we don't use the traditional definitions to determine part of speech. Instead, we use distributional criteria, as will be discussed below.

Table 1: Traditional semantic definitions of parts of speech

Part of speech	Traditional definition	Examples
Noun	A person, place, or thing	child, school, toy, Canada
Verb	Actions or states	eat, sleep, grow, know
Adjective	Describes quality, quantity, or extent	rich, far, green, numerous
Adverbs	Describes manner, quality, or degree	quickly, very, carefully, partially
Prepositions	Location or origin	in, from, on, at

Problems with the traditional approach

The same word can belong to multiple different categories

One big problem with the traditional definitions of Table 1 is that the same word can belong to multiple different categories. For example, in (1a), the word *work* is a verb; in (1b), it is a noun; and in (1c), it is an adjective.

(1)	a.	We **work** at a joke factory.
	b.	Have you finished your **work** yet?
	c.	My **work** clothes are filthy.

All three instances of the word *work* in (1) have related meaning—they are all associated with some kind of

labour. However, they have very different functions and behaviours in each sentence. Therefore, when we are determining the part of speech of a word, it is these functions and behaviours that we depend on to classify words.

Abstract nouns and stative verbs

Not all words are easy to classify according to the traditional semantic definitions, either. What part of speech is *idea*? You might know it is a noun, but how do you know that? It's definitely not a person or a place. It is debatable whether it's a thing. Or what about the noun *destruction*? That's an action, so we might think it's a verb! *The destruction of the city by Godzilla* describes an event—it's clearly not a person, a place, or a thing. However, the behaviour and function of *destruction* is that of a noun.

Likewise, not all verbs are actions, and some verbs stretch the notion of *state*. The sentence *Tasha owns a car* contains the verbs *owns.* Is owning an action or a state? Only under a very loose definition. What about the verb *seems* in *Abigail seems tired*. Is *seem* an action or a state?

Words can change category

Another piece of evidence that part of speech should not be determined by semantic definitions is that a word can change syntactic category over time without changing its meaning. One example of this is the word *fun*. For most speakers, *fun* is a noun. Compare the sentences in (2) and (3) and notice how *fun* has the same behaviour as other nouns, such as *water*. It does not share the same behaviour as adjectives, such as *hard*.

(2)	a.	That was a lot of **fun**.
	b.	That was a lot of **water**.
	c.	*That was a lot of **hard**.
(3)	a.	I had more **fun** than her.
	b.	I had more **water** than her.
	c.	*I had more **hard** than her.

For younger speakers and in informal speech, though, *fun* can also behave like an adjective. Like other adjectives, *fun* can take comparative and superlative inflection for these speakers, forming the words *funner* and *funnest*. For these younger speakers, *fun* can behave the same way as adjectives like *hard,* as shown in (4). Nouns, on the other hand, cannot appear in this kind of structure, as shown for *water* in (4b).

(4) a. Swimming is **funner** than running.
b. *Swimming is **waterer** than running.
c. Swimming is **harder** than running.

Different categories in different languages

Another piece of evidence that we cannot use meaning to define syntactic category comes when we compare languages. For instance, if you've studied a Romance language, you've learned that certain concepts which are expressed in English using adjectives are expressed in Romance using verbs and nouns. For example, if you want to express hunger in English, you would say *I am hungry*, using the adjective *hungry*. In contrast, in many Romance languages and in Swahili, you would say the equivalent of *I have hunger*, using a noun, as shown in (5).

(3) a. Yo tengo hambre
1SG have.1SG hunger
'I'm hungry.' (Literally: 'I have hunger.') Spanish
b. J'-ai faim
1SG-have.1SG hunger
'I'm hungry' (Literally: 'I have hunger.') French
c. nina njaa
have.1SG hunger
'I'm hungry' (Literally: 'I have hunger.') Swahili

If we (as English speakers) were to use meaning to define word category, we might be tempted to call words like *hambre, faim*, and *njaa* adjectives meaning 'hungry.' But this would be wrong, because they actually behave as nouns, as the literal translations in (5) suggest. Meaning doesn't help us to define categories within a language, and it doesn't help us define across categories *across* languages.

People can identify the category of made-up words

Perhaps the most fascinating piece of evidence that we cannot use semantic criteria to classify part of speech comes from the fact that people can reliably identify the category of made-up words, even when they don't know their meanings.

Consider below an excerpt from the poem *Jabberwocky* from *Through the Looking-Glass* by Lewis Carroll. These lines contain five made-up words, which are bolded below.

(6) And as in **uffish** thought he stood,
The **Jabberwock**, with eyes of flame,

Came **whiffling** through the **tulgey** wood,
And **burbled** as it came!

Even though we don't know what the words mean, we can tell that *uffish* and *tulgey* are likely adjectives, a *Jabberwock* is likely a noun, and *whiffling* and *burbled* are likely nouns. We cannot be using semantic criteria, since we don't even know what these words mean, so we must be using some other kind of evidence to determine the part of speech of these words.

Distributional criteria

If we cannot use semantic criteria, then what *do* we use? We use three criteria for determining part of speech:

- Syntactic distribution, which is the words that it can appear adjacent to.
- **Inflectional** morphology, which is how the word can be inflected.
- **Derivational** morphology, which are the derivational morphemes that it contains.

Let's look again at the stanza from the Jabberwock to see these criteria in action.

(7) And as in **uffish** thought he stood,
The **Jabberwock**, with eyes of flame,
Came **whiffling** through the **tulgey** wood,
And **burbled** as it came!

We can tell that *Jabberwock* is a noun because it appears after the word *the*, which is an example of syntactic distribution. For an example of inflectional morphology, consider the word *burbled*. It takes the past tense marker *-ed*, which is a property of verbs. Finally, *uffish* uses the suffix *-ish*, which is a derivational morpheme that appears on adjectives.

At this point you may be wondering if this is circular reasoning. We have decided to call things *nouns* because they behave like nouns. And we have decided how nouns behave by looking at nouns. The observation here is that there are groups of words that have similar behaviour. The names of the categories are essentially arbitrary: we could call them *blickershnitzels*, and the observation would still be true and useful. *Table, dog, teacher, brick, honesty,* and so on, all behave in a similar fashion in some respects, and we give this group of words the name *noun*. *Lick, sit, own, jump, describe, believe,* and so on, also all behave in a similar fashion in some respects, and we give this group of words the name *verb*. When we say that something is a noun and has category N, all we're saying is that that thing acts like the other noun-words. Likewise, when we say that something is a verb, we're simply saying that that thing acts like the other verb-words.

In the next few sections, we will go through the common syntactic, inflectional, and derivational properties of the different parts of speech. You can use these different behaviours to classify words by part of speech. However, as someone who is able to use words in different languages, you already know these properties, at

least subconsciously. We just need to learn to be consciously aware of these patterns. We can also look at data from a language we don't speak to find patterns of behaviour of different words and identify parts of speech in these other languages.

Of course, different languages have different word order, as well as different morphological patterns, and so the criteria won't be exactly the same across languages. Some linguists question whether all languages have the same categories at all! But there does seem to be some common patterns across languages. Many of the criteria we use in one language are parallel to similar criteria in other languages.

Categorizing phrases

Individual words are not the only things that have part of speech. Groups of words that behave as units—called phrases—also have distributional patterns that we can use to categorize them. Plus, there's a pattern. Phrases have a head—the main word of the phrase—and the part of the speech of the head of the phrase will predict the distributional patterns of the phrase. For example, a phrase headed by a noun is called an NP (short for noun phrase). Any NP can be the subject of a sentence, as shown in (8), even if we add more words to the phrase, or if we switch the head noun for a a different head noun.

(8) a. [the apple] is red
b. [the sweet apple] is red
c. [the apple on the ground] is red
d. [the apple that fell off the tree] is red
e. [some paint] is red
f. [my bicycle] is red

Likewise, any phrase headed by a verb is a VP, any phrase headed by a preposition is a PP, and so forth.

Key takeaways

- We classify words based on their behaviour and function. These classifications are called part of speech or syntactic category.
- Traditional definitions of parts of speech depend on semantic criteria, but there are several difficulties with determining part of speech based on meaning.

- Linguists use syntactic distribution, inflection, and derivation to classify words into part of speech.
- Phrases can also be classified by the part of speech of their head and share the distribution of other phrases with the same kind of head.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=4391#h5p-53

References and further resources

Attribution

Portions of this section are adapted from the following CC BY NC SA source:

Gluckman, John. n.d. Chapter 1: Syntactic categories. *The science of syntax*. https://pressbooks.pub/syntax/chapter/chapter-2-parts-of-speech/

8.2. NOUNS

We can identify nouns based on their inflectional morphology, derivational morphology, and syntactic distribution.

In terms of inflectional morphology, nouns may inflect for person, number, gender, and/or case. In English, nouns inflect for number, with singular and plural forms. English pronouns also have person and gender distinctions, in addition to number.

There are many different kinds of derivational morphemes that modify nouns. Some examples of derivational morphemes in English are shown below in Table 1.

Table 1. English derivational morphemes

	suffix	**examples**
a.	-ment	*employment, basement*
b.	-ness	*friendliness, emptiness*
c.	-ity	*sincerity*
d.	-ty	*certainty*
e.	-(t)ion	*devotion*
f.	-ation	*expectation*
g.	-ist	*specialist, linguist*
h.	-ant	*attendant,*
i	-ery	*shrubbery, mockery*
j.	-ee	*employee, invitee*
k.	-ship	*hardship, friendship*
l.	-aire	*billionaire,*
m.	-acy	*advocacy*
n.	-let	*piglet*
o.	-ling	*underling*
p.	-hood	*neighbourhood,*
q.	-ism	*socialism*
r.	-ing	*fencing*

(adapted from Carnie 2021: 47)

In terms of syntactic distribution, noun phrases typically show up in argument positions, both the arguments of verbs and the arguments of prepositions, as shown in (1).

(1) a. **The apple** is red. *subject*

b. I ate **the apple**. *direct object*

c. I gave **the apple** a shine. *indirect object*

d. There was a worm in **the apple**. *object of preposition*

Nouns can be modified by determiners and adjectives, and so their syntactic position can often be described with reference to determiners and adjectives. In English, nouns can appear after either determiners or adjectives, or appear on their own, as shown in (2).

(2) a. I ate the apple. *after a determiner*

b. I ate the red apple. *after an adjective*

c. I ate apples. *on its own*

Inflectional values on nouns

The most common kinds of inflection found on nouns include case, person, number, and gender. The latter three of these (person, number, and gender) tend to have close interactions and the group is often referred to as φ-features, named after the Greek letter phi.

Case

Case is used to mark the role a noun phrase plays in a sentence. Some languages do not mark case at all, some mark case on on only some nouns, some mark case on all nouns, and some mark case on determiners. In English, we have case distinctions only on pronouns. We use the nominative case form in the subject position of finite clauses, the genitive case for possessives, and the accusative case form on the objects of active voice clauses. In English, the accusative case also acts as the default and shows up in additional contexts, as well.

(3) a. I saw them.

1SG.NOM (no gloss for saw) 3PL.ACC

b. They saw me.

3PL.NOM (no gloss for saw) 1SG.ACC

The pronoun paradigm for English is shown in Table 1. In addition to case, English pronouns also have person, number, and gender distinctions, which we will discuss later in this section.

Table 2. English personal pronouns

Features			Morphological forms		
Person	**Number**	**Gender**	**nominative case**	**accusative case**	**genitive case**
1st	**singular**	**—**	I	me	mine
2nd	**singular**	**—**	you	you	yours
3rd	**singular**	**masculine**	he	him	his
3rd	**singular**	**feminine**	she	her	hers
3rd	**singular**	**inanimate**	it	it	its
3rd	**singular**	**non-binary or unspecified**	they	them	theirs
1st	**plural**	**—**	we	us	ours
2nd	**plural**	**—**	you	you	yours
3rd	**plural**	**—**	they	them	theirs

In English, we use the accusative case for both direct and indirect objects, but in many other languages, they use the accusative case only for direct objects, and have another case, the dative case, for indirect objects. German, as shown in (4), is one example of a language that uses dative case. The example in (4) also illustrates how in some languages, including German, case is marked on determiners rather than on the noun itself.

(4) Sie hat dem Mann das Buch geschenkt.

she.NOM has the.M.DAT man the.N.ACC book given

'She has given the man the book."

German (Beermann 2001, as cited in Alexiadou et al. 2010: 2).

Some languages have extensive case marking systems that they use in addition to or instead of prepositions. Hungarian is an extreme example of this, with 21 cases, as shown in (5).

(5) *Hungarian*

hajó	nominative
hajó-t	accusative
hajó-ban	inessive
hajó-ból	elative
hajó-ba	illative
hajó-n	superessive
hajó-ról	delative
hajó-ra	sublative
hajó-nál	adessive
hajó-tól	ablative
hajó-hoz	allative
hajó-ig	terminative
hajó-nak	dative
hajó-val	instrumental-comitative
hajó-képp	formal
hajó-ul	essive
hajó-ként	essive-formal(-similitive)
hajó-vá	translative-factitive
hajó-ért	causal-final
hajó-nként	distributive
hajó-stul	sociative

(Tompa 1968: 206-209, as cited in Iggesen 2013)

Languages such as English, German, and Hungarian have what is called nominative-absolutive case alignment, which means that we use the same case marker for the subjects of transitive and intransitive verbs, and a different case marker for the objects of transitive verbs. This is shown in Figure 1.

(6) a. **She** saw **her**.

b. **She** slept.

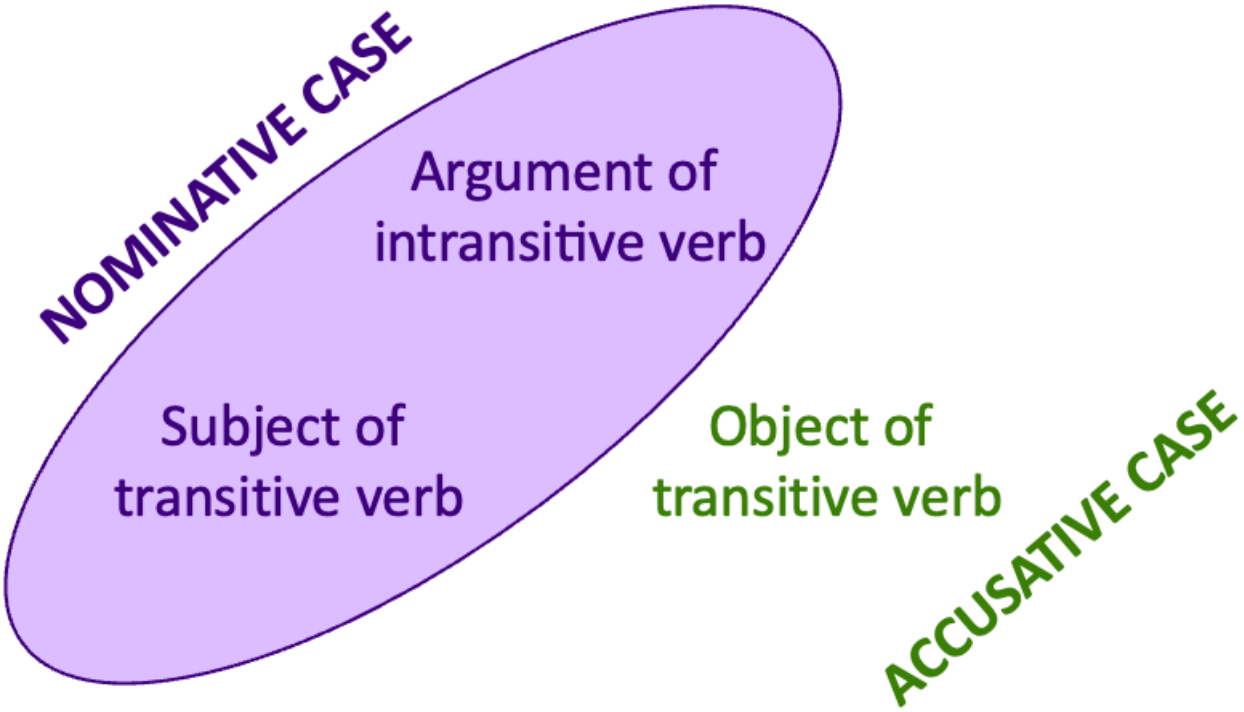


Figure 1: A schematic representation of nominative-accusative case alignment.

On the other hand, other languages mark the sole argument of intransitive verbs with the same case as the object of transitive verbs, which is called absolutive case. The case that is used to mark only subjects of transitive verbs is called ergative case. Ngarinyman, a Pama-Nyungan language of Australia, is an example of an ergative-absolutive language. For example, sentence (7a) is intransitive, and so the sole argument *garu* 'child' appears in absolutive case. In (7b), when the same verb is used transitively, the subject, *yawardarlu* 'horse' appears in ergative case. The object in (7b) appears only as agreement on the verb.

(7) *Ngarinyman*

a.	Bard	wan	-inya	**garu**.
	fall.down	change.location	-pst	**child.abs**
	'The kid fell down.'			

b.	**Yawarda**	**-rlu**	bardba	**-yi**	yuwa	-ni.
	horse	**-erg**	fall.down	**-1sg.obj**	put	-pst
	'The horse threw me off.'					

(Denniss 2023: 142)

We can see another transitive-intransitive pair in (8). Sentence (8a) is intransitive, and the sole argument *warlagu* 'dog' is again marked in absolutive case. In the transitive version, the subject *warlagu* 'dog' is marked with ergative case while the object, *janga* 'woman' is marked in absolutive.

(8) *Ngarinyman*

a.	Rungab	garr	-any	**warlagu**.
	bark	be.at	-pres	**dog.abs**
	'The dog is barking.'			

b.	**Warlagu**	**-rlu**	rungab	nya	-nga	**jan.ga**.
	dog	**-erg**	bark	perceive	-pres	**woman.abs**
	'The dog barked at the woman.'					

(Denniss 2023: 145-146)

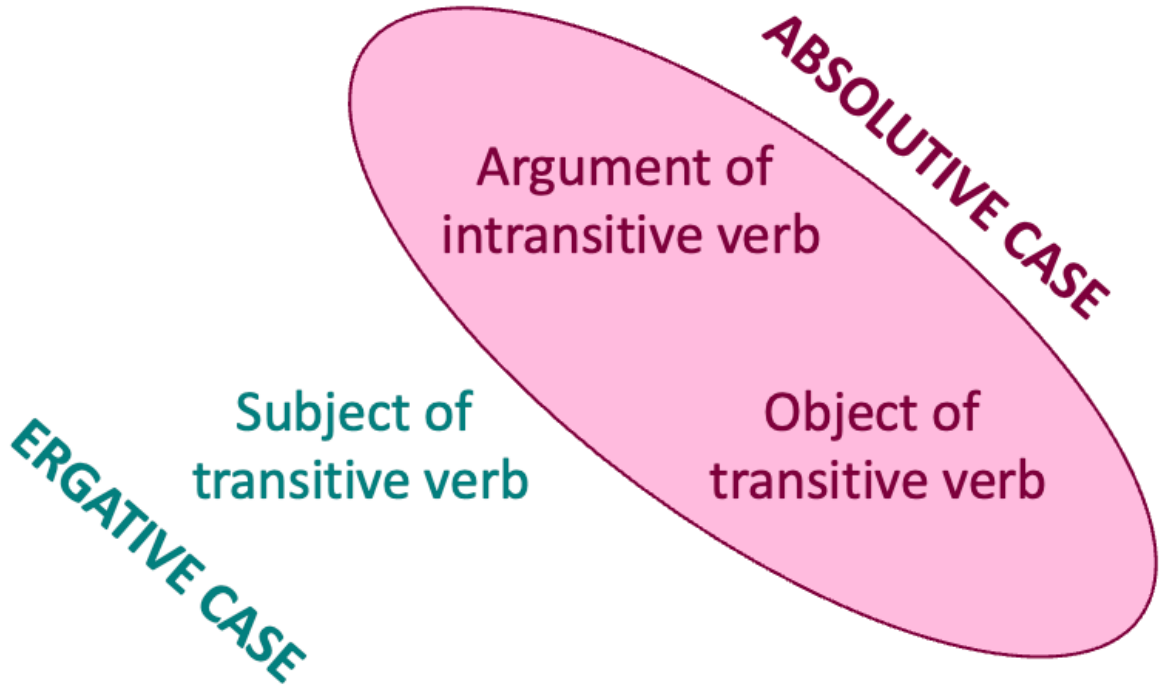


Figure 2: A schematic representation of ergative-absolutive case alignment.

Many languages use a combination of nominative-accusative and ergative-absolutive case. This is called split ergativity. The contexts in which each case system is used can be determined by tense, person, or other factors.

The names and meanings of some common cases are listed in Table 3.

Table 3. The meanings of some common cases.

Case	Meaning
nominative	subject of a finite verb
accusative	object of an active verb
ergative	subject of a transitive verb
absolutive	argument of an intransitive verb or object of a transitive verb
dative	indirect object of a verb
genitive	possessor of another noun
ablative	movement away from
vocative	addressee
locative	location
instrumental	a tool or means used to perform the action

Person

Nouns in most languages are categorized as first, second, or third person. First person nouns refer to the person speaking or signing. Second person nouns refer to the person being addressed. Third person nouns are everyone and everything else.

First person plural pronouns typically mean the person speaking or signing and people associated with them, rather than meaning that multiple people are speaking or signing (although, in rare cases, it can mean that, too). On the other hand, second person plural pronouns can either mean that you are addressing a group *or* that you are referring to the person you are addressing and the people who are associated with them.

Some languages make a distinction between inclusive and exclusive first person plural pronouns. Inclusive first person pronouns include the addressee in their reference (meaning 'us and you'), while exclusive ones do not (meaning 'us and not you'). Niuean, an Austronesian language spoken on the island of Niue, is an example of a language with inclusive and exclusive pronouns. Niuean's pronoun paradigm is shown in Table 4.

Table 4. Niuean personal pronouns (Seiter 1980: 49)

	singular	dual	plural
1st exclusive	au	maua	mautolu
1st inclusive		taua	tautolu
2nd	koe	mua	mutolu
3rd	ia	laua	lautolu

Some languages also have an obviative, which is sometimes referred to as the fourth person. In languages that have an obviative, only one third person noun (usually the most topical one) will be marked as proximate while the remaining 3rd person nouns will be marked with the obviative, as shown below in (9). In the Oji-Cree example in (9a), the subject is in the first person, so the third person object *awaasihs* 'the child (proximate)' is marked with the proximate, not the obviative. But in (9b), there is a third person subject. Therefore, the third person object *awaasihsan* 'the child (obviative)' is marked in the obviative with the suffix *-an*. Oji-Cree is an Algonquian language spoken in Manitoba and Northern Ontario.

(9) Oji-Cree (Oxford 2017: 2)

a. Niwaapamaa awaasihs.

ni-waapan-aa awaasihs-Ø

1-see-DIR- child-PX.SG

'I see the child.'

b. Owaapamaan awaasihsan

O-waapam-aa-n awaasihs-an.

3-see-DIR-OBV child-OBV

'S/he sees the child.'

Number

Number indicates how many instances of the noun there are. In English, we have two number categories: singular and plural, although singular is unmarked. In languages like English which have both a singular and plural, the singular means that there is exactly one while the plural means that there is more than one. The plural marker in English is-*s*, although there are also a number of nouns which take irregular plurals, such as *children*, *feet*, and *cacti*.

(10) a. cat *singular*

b. cat-s *plural*

Some languages also have a dual marker, which indicates that there is exactly two instances of the noun. In languages with a dual marker, singular means exactly one, dual means exactly two, and plural means three or more instances of the noun, as shown in (11).

(11) a. taalib *student.M.SG* 'student'

b. taalib-een *student.M.DU* 'two students'

c. taalib-iin *student.M.PL* 'students'

Gulf Arabic (Holes 1990: 148, as cited in Aikhenvald 2018: 2)

A few languages have a trial, which marks exactly three, and in which case the plural would be used for four or more instances of the noun, as shown in (12).

(12) a. ngayi 'I (singular)'

b. ngarrgu 'we two' (dual)

c. ngarrgunime 'we three' (trial)

d. ngagurr 'we (many)' (plural)

Ngan'gityemerri (Reid 1990: 118-118, as cited in Aikhenvald 2018: 2)

Even fewer languages go higher than three, but languages with a quadruple or a quintuple form do exist. One is American Sign Language.

Some languages also have a paucal, as shown in (13), which indicates that there are only a few.

(13) a. áine ŋára
woman that:3SG
'that woman' (singular)

b. áine ŋara-díaru
woman that-dual
'those two women' (dual)

c. áine ŋara-díato
woman that-paucal
'those few women' (paucal)

d. áine ŋára-di
woman that-PL
'those women' (plural)

Manam (Lichtenberk 1983: 67, as cited in Aikhenvald 2018: 3)

Some nouns arbitrarily take a plural marker no matter how many of the item there are, such as *pants*, *scissors*, and *(eye)glasses*. Often, these are items that typically come in two roughly equal or symmetrical parts, but not all items which consist of a pair require the plural marking, showing that it is an arbitrary grammatical rule that must be encoded in the lexicon. For example, *bicycle*, *eyes*, *chopsticks*, or *socks* all also typically come in pairs or two parts but have both singular and plural forms.

Gender

In grammar, gender refers to noun classes. In languages with extensive gender systems, all nouns belong to a noun class which is part of the lexical entry of that noun. Other words in the phrase will agree with the gender of the noun. For example, Russian nouns trigger agreement on the verb, as shown in (14). With a masculine noun such as *žurnal* 'magazine' in (14a), the verb appears in the masculine bare form. With a feminine noun such as *kniga* 'book' in (14b), the verb appears with the feminine suffix *-a*. Finally, with a neuter noun such as *pis'mo* 'letter' in (14b), the verb appears with the neuter suffix *-o*.

(14) a. Žurnal ležal na stole.
magazine lay.M on table.
'The magazine laid on the table.'

b. Kniga ležal-a na stole.

book lay-F on table

'The book lay on the table.'

c. Pis'mo ležal-o na stole.

letter lay-N on table.

'The letter lay on the table'

Russian (Corbett 2013)

In some languages, the noun class is marked explicitly with a morpheme on the noun, but in other languages, it only shows up on other words that agree with the noun, as with the Russian example in (14).

In SiSwati, a Bantu language spoken in Eswatini and South Africa, the noun class markers are indicated both on the noun and the words that agree with it, such as adjectives. This is illustrated below in (15) for class 2 marker *ba-* in (15a), class 3 marker *li-* in (15b) and class 7 marker *si-* in (15c). Note that, unlike in many Indo-European languages, the noun classes in Bantu languages like SiSwati are not based on gender. Instead all of the human nouns belong to class 1 (when singular) and class 2 (when plural).

(15) a. ba-fati la-ba-k^{h}ulu

Cl2.PL-woman ADJ-Cl2.PL-big

'big women'

b. li-ɬombe le-li-k^{h}ulu

Cl3.SG-shoulder ADJ-CL3.SG-big

'big shoulder'

c. si-lwane le-si-k^{h}ulu

Cl7.SG-animal ADJ-Cl7.SG-big

'big animal'

In some languages, nouns belonging to different classes have distinct inflectional paradigms. This occurs in Russian, for example, as shown in Table 5. Notice how each gender has different suffixes for each case category.

Table 5. Inflectional paradigms of different gender singular nouns in Russian

Inflectional class	Gender	Nominative	Accusative	Genitive	Dative	Instrumental	Locative	Gloss
I	**masculine**	žurnal	žurnal	žurnala	žurnalu	žurnalom	žurnale	'magazine'
II	**feminine**	gazeta	gazetu	gazety	gazete	gazetoj	gazete	'newspaper'
III	**feminine**	kost'	kost'	kosti	kosti	kost'ju	kosti	'bone'
IV	**neuter**	pis'mo	pis'mo	pis'ma	pis'mu	pis'mom	pis'me	'letter'

The different meanings of *sex* and *gender*

Many people confuse the terms *sex* and *gender*. To make matters even more confusing, linguists use *gender* in more than one way. Let's define these terms precisely so that we can be clear about what we mean.

Sex

Definition: A set of biological attributes in humans and animals, including hormone level and expression, reproductive anatomy, chromosomes, and gene expression (CIHR 2023).

Categories: Sex is usually divided into male and female, but there is variation in how these different biological attributes are expressed, including some who have attributes of both sexes or neither (Fausto-Sterling 2000).

Property of: People and animals.

(Personal) gender

Definition: The socially constructed roles, behaviours, and expressions of a person based on how people perceive themselves and interact with others (CIHR 2023).

Categories: Some common gender categories include woman, man, girl, boy, or non-binary person (CIHR 2023).

Property of: People and maybe animals.

Semantic gender

Definition: Nouns that refer to people or animals sometimes carry semantic gender, which means that the personal gender of the person it refers to (or the sex of an animal) is encoded in the meaning of the word. This includes words like *sister* and *brother*, for which the semantic gender is part of the root word, but also words like *duke* and *duchess*, which include semantic gender in a separate morpheme. An increasing number of non-binary forms are being coined, such as *nibling* for the child of a sibling.

Categories: Depending on the context, these may be referred to as male, female, and unspecified, or as masculine, feminine, and epicene.

Property of: Words, mostly nouns. Morphemes encoding primarily semantic gender are usually derivational morphemes.

Grammatical gender

Definition: Grammatical gender refers to noun classes in grammar. Unlike semantic gender, these noun classes may trigger agreement in other words or determine which inflectional paradigm a word uses.

Categories: In some languages, including many Indo-European languages, grammatical gender has distinctions between masculine and feminine or between masculine, feminine, and neuter. Other languages, including many Algonquian languages, have distinctions between animate and inanimate nouns. Languages in the Bantu family have approximately 16 noun classes (half singular and half plural), and all humans belong to the same noun class, as we saw for SiSwati above in example (15).

Property of: Words, mostly nouns. Morphemes encoding grammatical gender are usually inflectional morphemes.

In languages which use grammatical gender, many nouns may have a grammatical gender which corresponds to the semantic gender of the entity to which it refers, but there are also nouns which do not have semantic gender and are assigned grammatical gender arbitrarily. For example, *milk* does not have semantic gender, and it is arbitrarily assigned masculine gender in French and feminine gender in Spanish, as shown in (16).

(16) a. le lait
DEF.SG.M milk
'the milk'
French
b. la leche
DEF.SG.F milk
'the milk'
Spanish

There are are even nouns that have a semantic gender that doesn't match its grammatical gender. One rather well-known example of this is the German word *mädchen* 'girl' which is semantically female but grammatically neuter.

(17) das mädchen

NOM.DEF.SG.N girl

'girl'

German

The distinction between semantic and grammatical properties can also be made for other inflectional categories, including number (as we saw earlier for *pants* and *scissors*) and animacy. For example, in Meskwaki, an Algonquian language from the Iowa region, most plants use inanimate gender in everyday use, but may be marked with animate gender when they are being spoken to or in religious contexts. This is illustrated below for strawberry in (18).

(18) a. ahte:himin-i

strawberry-SG.INAN

'strawberry (inanimate)'

b. ahte:himin-a

strawberry-SG.AN

'strawberry (animate), when addressed'

Meskwaki (Dahlstrom 1995: 58)

On the other hand, the noun *raspberry* is always animate—even in everyday contexts—and cannot appear with inanimate gender.

(19) a. wi:tawi:h-a

raspberry-SG.AN

'raspberry'

b. *wi:tawi:h-i

raspberry-SG.INAN

Meskwaki (Dahlstrom 1995: 58)

Singular *you* and *they*

In English, the pronoun *you* is used for both the singular and plural, but this was not always the case. *You* started out as a plural pronoun, and the pronouns *thou* (nominative), *thee* (accusative), and *thy* (genitive) were the singular pronouns. Here are some examples of the archaic singular pronoun from the Shakespeare play *Hamlet*.

(11) a. **Thou** seest the heavens (Act II Scene IV)

b. I'll give **thee** a wind. (Act I Scene III)

c. ...and everyone did bear **thy** praises in his kingdom's great defence (Act I Scene III)

In Early Modern English, at the time of Shakespeare, *you* could be used as a plural pronoun or it could be used as a singular pronoun in more formal contexts and to show respect to the addressee. Over time, the polite *you* form became more and more common. Some prescriptivists didn't like this change and complained about it, as illustrated by the quote below.

> There has never been a greater perversion of language than the using the pronoun *you*, in the place of *thou*. [...] Using *you* for *thou* causes us to depart from some of the fundamental rules of grammar. For example: we are taught that "a verb should agree with its nominative number and person."
>
> -Thomas 1857: 114-115

Now, in modern English, *you* is firmly established as a singular pronoun, replacing *thou/thee/thy* entirely. Some dialects have even coined new plural pronouns such as *y'all* or *youse* to re-establish the distinction between second person singular and plural.

Using the plural pronoun to show respect to single addressees is also used in other languages. For example the plural second person pronoun *vous* is used as a polite form to address single addressees in French, instead of the singular second person pronoun *tu*, which, depending on the dialect, is used only for close friends and family.

The pronoun *they* also originated as a plural pronoun but has been used as a singular pronoun in certain contexts since at least the 1400s (Balhorn 2004). Since the other third person singular pronouns specify gender or animacy, singular *they* is often used when the gender is unknown or irrelevant, as in (12).

(12) a. (*Seeing an unidentified distant figure*) **They**'re waving at us.

b. (*After answering a phone call*) **They** had the wrong number.

(Bjorkman 2017: 1)

Singular *they* was also used when it referred back to a noun modified by a **quantifer**, even if the gender was specified, as shown through the historical examples in (13).

(13) a. Shakespeare (*A Comedy of Errors*, 1623)

There's not a man I meet but doth salute me
As if I were **their** well-acquainted friend

b. Austen (*Pride and Prejudice*, 1813)

Both sisters were uncomfortable enough. Each felt for the other, and of course for **themselves**.

(Bjorkman 2017: 3)

More recently, *they* has begun to be used as a singular pronoun for specific people whose gender is known. Some modern English speakers accept neither the sentences in (14) or (15). Some accept the sentences in (14) but not (15), and some accept all of the sentences in both (14) and (15). Sentences that are grammatical for some people but not all are marked with a percentage sign (%). Subscript letters, called co-indexes, are included beside nouns to show that they refer to the same entity.

(14) a. %The professor$_i$ said they$_i$ cancelled the exam.

b. %Our eldest child$_i$ broke their$_i$ leg.

c. %I'll let my cousin$_i$ introduce themselves$_i$.

(Bjorkman 2017: 2)

(15) a. %Janet$_i$ said they$_i$ cancelled the exam.

b. %Thomas$_i$ broke their$_i$ leg.

c. %I'll let my sister$_i$ introduce themselves$_i$.

(Bjorkman 2017: 2)

Several different researchers have suggested that the different judgment patterns for sentences like those in (14)-(15) are the result of a few small differences in people's grammar. Bjorkman (2017) argues that some people have a grammar that requires referential nouns to be specified for gender, even if it is not morphologically marked, while others have a grammar that allows nouns to be unspecified for gender. Konnelly and Cowper (2020) argue that there are three distinct grammars for *they* in competition, representing different stages of a change in progress, as described in (15).

(15) a. Stage 1 singular *they*: (quantified antecedent, or antecedent of unknown/irrelevant gender), as shown in (12)-(13)

b. Stage 2 singular *they*: antecedent of known gender, but ungendered description//name, as in (14)

c. Stage 3 singular *they*: antecedent of any gender, no restriction on description/name, as in (15)

Camilliere et al. (2021) did a sociolinguistic study that shows that people's linguistic behaviour does cluster into the three distinct patterns proposed by Konnelly and Cowper. How people rate singular *they* in different contexts is correlated with social attitudes toward transgender people and prescriptivist language, as well as age and social distance between the speaker and the referent of the pronoun.

Some people feel like gendered pronouns like *she* or *he* do not fit them and may request that others refer to them by using *they*. We can show respect for them by using the pronouns that they request. However, sometimes even people who want to use singular *they* to refer to specific people may struggle to because their grammar is more restrictive, and they may find the change in usage difficult to acquire at first. The good news, though, is that learning and internalizing a new grammar is entirely possible with practice! Kirby Conrod, a linguist who uses *they* pronouns, gives some tips on how to do that in their blog.

Classifiers

Classifiers are similar to gender, in that nouns are lexically assigned to different groups, often based on some semantic criteria. However, unlike gender, classifiers do not trigger agreement on other words and there may be hundreds of classifiers in use in a given language.

Mandarin is one example of a language that uses noun classifiers. In Mandarin, nouns must always be marked with a classifier when they are combined with a numeral, as shown in (20). The classifiers can indicate either individual objects, as in (20a), or containers, as in (20b) and (20c).

(20) a. liang ben shu
two CL book
'two boxes'
b. liang xiang shu
two CL:box book
'two boxes of books'
c. liang ping jiu
two CL:bottle wine
'two bottles of wine'

Mandarin (Zhang 2007: 45)

Classifiers in Mandarin often encode the shape of the object. For example, the classifier *tiao* indicates long and rope-like objects. Words such as *she* 'snake', *tui* 'leg', *kuzi* 'pair of pants', *he* 'river', and *bandeng* 'bench' all appear with the classifier *tiao* (Zhang 2007: 44). Classifier *ke* is used for small, solid, and kernel-like objects, such as *huangdou* 'soybean', *ya* 'tooth', *dingzi* 'nail', and *zidan* 'bullet' (Zhang 2007: 46).

Classifiers can also appear on the verb in some languages, to indicate the semantic properties of one of its arguments. This is how handshape classifiers are often used in American Sign Language. It also occurs in spoken language. For example, in the Papuan language Waris, the classifier *put-* 'round objects' indicates that the object argument, a coconut, is round.

(21) sa ka-m put-ra-ho-o

coconut 1SG-to VCL:ROUND-get-BENEFACT-IMPERATIVE

'Give me a coconut' (lit. coconut to-me round.one-give)

Waris (Aikhenvald 2000: 3)

Derived nouns

There are some derivational morphemes that are common on nouns cross-linguistically, which are summarized in Table 6. But there are also many, many possible derivational meanings that might only show up in a single language. Some derivational morphemes change the part of speech of their stem (as in Table 7 and Table 8, which list derivational morphemes that change verbs and adjectives into nouns), whereas others keep the same part of speech (as in Table 9).

Table 7. Deverbal nouns (V → N) (adapted from Haspelmath and Sims 2010: 87)

agent noun	refers to the subject of the verb	English	*drink*	→	*drink-er*
		Arabic	*ħamala* 'carry'	→	*ħammaal* 'carrier'
patient noun	refers to the object of the verb	English	*invite*	→	*invit-ee*
instrument noun	refers to an instrument used to perform the verb	English	*blend*	→	*blend-er*
		Spanish	*picar* 'mince'	→	*pica-dora* 'meat grinder'
action noun	refers to the action of the verb	English	*destroy*	→	*destruction*
		Arabic	*otkry-t'* 'discover'	→	*otkry-tie* 'discovery'

Table 8. Deadjectival nouns (A → N) (adapted from Haspelmath and Sims 2010: 87)

quality noun	refers to the quality of having the property of the adjective	English	*bold*	→	*bold-ness*
		Japanese	*atarasi-i* 'new'	→	*atarasi-sa* 'newness'
person noun	refers to an entity with the property of the adjective	English	*short*	→	*short-y*
		Russian	*umn-yj* 'smart, clever'	→	*umn-ik* 'clever guy'

Table 9. Denominal nouns (N → N) (adapted from Haspelmath and Sims 2010: 87)

diminutive noun	refers to a smaller, younger, cuter, or more intimate version of the noun	English	*basin*	→	*basin-ette*
		Spanish	*gat-o* 'cat'	→	*gat-it-o* 'little cat'
augmentative noun	refers to a larger version of the noun	English	*church*	→	*megachurch*
		Russian	*borod-a* 'beard'	→	*borod-išča* 'huge beard'
status noun	refers to the having the status or rank of the noun	English	*child*	→	*child-hood*
				→	
inhabitant noun	refers to one who inhabits the noun	English	*Canada*	→	*Canad-ian*
		Arabic	*Miṣr* 'Egypt'	→	*miṣr-iyyu* 'Egyptian'
female noun	refers to a noun having feminine semantic gender	English	*duke*	→	*duchess*
		German	*König* 'king'	→	*König-in* 'queen'

Key takeaways

- Cross-linguistically, nouns may inflect for person, number, gender, and or case.
    - Person marks whether the noun is a conversation participant.
    - Number marks how many of the noun there are.
    - Gender marks the inflectional noun class of the word.
    - Case marks the role of the noun in the sentence.
- Nouns typically show up in argument position of a verb or the object of a preposition.
- Nouns may be modified by determiners or adjectives.
- There are many derivational nouns that occur on nouns, including some that turn adjectives and verbs into nouns.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=4394#h5p-60

References and further resources

For a general audience

Canadian Institutes of Health Research. 2023. What is gender? What is sex? https://cihr-irsc.gc.ca/e/48642.html

Conrod, Kirby. 2020. Pronouns 101: Introduction to your loved one's new pronouns. *Medium*. https://kconrod.medium.com/pronouns-101-introduction-to-your-loved-ones-new-pronouns-3fef080266d0

Fausto-Sterling, Anne. 2000. *Sexing the Body*. New York: Basic Books.

McCulloch, Gretchen, and Kirby Conrod. 2020. The grammar of singular *they*: Interview with Kirby Conrod. *Lingthusiasm* episode 43. https://lingthusiasm.com/post/615600862742609920/lingthusiasm-episode-43-the-grammar-of-singular

Thomas, P.B. 1857. Thou and you. *The R.I. Schoolmaster* 3 (4): 114-115.

For students

Haspelmath, Martin and Andrea Sims. 2010. Understanding Morphology. London: Hodder Education.

Academic sources

Aikhenvald, Alexandra. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

Aikhenvald, Alexandra. 2018. Number systems in grammar – position paper. *Langugae and Culture Research Centre 2018 Workshop.* https://www.jcu.edu.au/__data/assets/pdf_file/0004/789700/Number-systems-in-grammar-Position-paper.pdf

Balhorn, Mark. 2004. The rise of epicene *they*. *Journal of English Linguistics* 32 (2): 79-166.

Bjorkman, Bronwyn. 2017. Singular *they* and the syntactic representation of gender in English. *Glossa* 2 (1): 80.

Camilliere, Sadie, Amanda Izes, Olivia Leventhal, and Daniel J. Grodner. 2021. *They* is changing: Pragmatic and grammatical factors that license singular *they*. *Proceedings of the Annual Meeting of the Cognitive Science Society* 43: 1542-1548. https://escholarship.org/uc/item/3tc9s9b0

Corbett, Greville. 2013. Number of genders. In *WALS,* ed. Matthew Dryer and Martin Haspelmath. Zenodo. https://wals.info/chapter/30

Dahlstrom, Amy. 1995. Motivation vs. predictability in Algonquian gender. In *Papers of the 26th Algonquian Conference,* ed. David Pentland. Winnipeg: University of Manitoba. 52-66.

Denniss, Jessica. 2023. *Non-conflating roots: The syntax of light verbs and complex predicates in Ngarinyman*. PhD thesis, University of Toronto.

Iggesen, Oliver. 2013. Number of cases. In *WALS,* ed. Matthew Dryer and Martin Haspelmath. Zenodo. https://wals.info/chapter/49

Konnelly, Lex, & Elizabeth Cowper. 2020. Gender diversity and morphosyntax: An account of singular *they*. *Glossa* 5 (1): 40.

Oxford, Will. 2017. Proximate DP, obviative KP: Balancing the morphosyntax and pragmatics of obviation. *Proceedings of the 2017 annual conference of the Canadian Linguistic Association.*

Seiter, William. 1980. *Issues in Niuean syntax*. New York: Garland Press.

Thomas, P.B. 1857. Thou and you. *The R. I. Schoolmaster* 3 (4): 114-115. https://www.jstor.org/stable/44786667

Zhang, Hong. 2007. Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics* 16 (1): 43-59.

8.4. ADJECTIVES AND ADVERBS

As with the other parts of speech, adjectives and adverbs can be identified based on their morphological and syntactic distribution.

Adjectives

Inflection on adjectives

Many adjectives inflect into comparative and superlative forms. The comparative means to a greater degree, while the superlative means to the utmost degree. Some examples of adjective paradigms are shown in Table 1. In English, shorter words tend to take the comparative suffix *-er* and the superlative suffix *-est*, while longer words use *more* and *most* periphrastically.

Table 1. The inflectional paradigms of some English adjectives

base form	**comparative**	**superlative**
happy	happi-er	happiest
black	black-er	black-est
good	better	best
quiet	quiet-er	quiet-est
intelligent	more intelligent	most intelligent
beautiful	more beautiful	most beautiful

Derivation of adjectives

Some common derivational morphemes that apply to adjectives are listed in (1)

(1)	a.	-ing	*the dancing cat*
	b.	-ive	*indicative*
	c.	-able	*readable*
	d.	-al	*traditional*
	e.	-ate	*intimate*
	f.	-ish	*childish*
	g.	-some	*tiresome*
	h.	-(i)an	*reptilian*
	i.	-ful	*wishful*
	j.	-less	*selfless*
	k.	-ly	*friendly*
	l.	un-	*unhappy*

(adapted from Carnie 2021: 48)

Cross-linguistically, derivational morphemes that form adjectives commonly come from verbs, nouns, or other adjectives. Two common types of deverbal adjectives (V –> A), shown with examples in Table 2, are adjectives that describe the verbal action as possible, called the facilitative, or that attribute the verbal action as typical of a person, called the agentive.

Table 2. Deverbal adjectives (V → A) (adapted from Haspelmath and Sims 2010: 89)

facilitative adjective	English	*break*	→	*break-able*
	Basque	*jan* 'eat'	→	*jan-garri* 'edible'
agentive adjective	English		→	
	Spanish	*habla-r* 'talk'	→	*habla-dor* 'talkative'

Some common kinds of denominal adjectives (N –> A) are shown in Table 3. Relational adjectives are adjectives that describe relatedness to the noun. Proprietive adjectives describe having the property of the noun. Privative adjectives describe the lack of the noun. Finally, material adjectives describe the material something is made of.

Table 3. Denominal adjectives (N → A) (adapted from Haspelmath and Sims 2010: 89)

relational (= 'related to N') adjective	English		→	
	Russian	*korol'* 'king'	→	*korol-evskij* 'royal'
proprietive (= 'having N') adjective	English	*wonder*	→	*wonder-ful*
	Ponapean	*pihl* 'water'	→	*pil-en* 'watery'
privative (= 'lacking N') adjective	English	*peanut*	→	*peanut-free*
	Russian	*vod-a* 'water'	→	*bez-vod-nyj* 'waterless'
material adjective	English	*wood*	→	*wood-en*
	German	*Kupfer* 'copper'	→	*kupfer-n* 'made of copper'

Some common kinds of deadjectival adjectives (A –> A), shown in Table 4, include attenuative adjectives, which create a weaker form of the adjective; intensive adjectives, which create a stronger form of the adjective, and negative adjectives, which create a negative form of the adjective.

Table 4. Deadjectival adjectives (A → A) (adapted from Haspelmath and Sims 2010: 89)

attenuative adjective	English		→	
	Tzutujil	*kaq* 'red'	→	*kaq-koj* 'reddish'
intensive adjective	English		→	
	Turkish	*yeni* 'new'	→	*yep-yeni* 'brand new'
negative adjective	English		→	
	German	*schön* 'beautiful'	→	*un-schön* 'ugly'

Syntactic distribution of adjectives

Adjectives typically modify nouns, and so their distribution can often be described with respect to nouns and other things that modify nouns. In English, adjectives come before nouns, but after determiners.

(2) a. the happy child

b. a blue book

c. that long drive

Adjectives can also appear in copular constructions, following the verb BE. Watch out, though nouns, verbs, and prepositions can also appear following BE!

(3) a. The child will be happy.

b. The book is blue.

c. That drive was long.

Adjectives can be modified by adverbs, so their distribution can also be described with respect to adverbs. Adjectives in English typically follow the adverbs that describe them.

(4) a. very quick

b. so happy

c. more readable

Adverbs

Inflection on adverbs

Adverbs don't take much inflection, although they can occasionally be used comparatively with the word *more*, as in (5).

(5) She ran the race **more quickly** than I did.

Derivation of adverbs

Many adverbs in English use the suffix *-ly*.

Syntactic distribution of adverbs

Adverbs can occur in many different locations, including at the beginning or end of the clause, as in (6).

(6) a. **Unfortunately**, I dropped my ice cream at the park.

b. I dropped my ice cream at the park, **unfortunately**.

Adverbs can also modify adjectives, as well as other adverbs. For example, in (7a), the adverb *very* modifies the adjective *happy*. In (7b), the adverb *so* modifies the adverb *very*.

(7) a. **very** happy

b. **so** very happy

Adverbs can also modify verbs, usually appearing at the beginning or end of the verb phrase.

(8) a. I **quickly** went to the store.

b. I went to the store **quickly**.

In fact, adverbs can appear in so many positions that it may be easier to describe where they can't appear. They cannot appear in between determiners and nouns, as in (9a), and they cannot appear after copula BE, as in (9b).

(9) a. *the **quickly** dog

b. *This dog is **quickly**.

You may have noticed something about this distribution—it is the *opposite* of the distribution of adjectives. In other words, the distribution of adjectives and adverbs are in **complementary distribution**.

One category or two?

You may be familiar with the idea of complementary distribution from phonology, where a complementary distribution pattern can be used to argue that two segments are allophones of the same phoneme. When two kinds of things are in complementary distribution, this can be used as evidence that they are actually variants of a single category.

Adjectives and adverbs also have many things in common:

- They both describe as their basic functions, although adjectives typically describe nouns while adverbs describe clauses, verbs, and other adjectives.
- They both can take the derivational morpheme *-ly* (compare the adjective *friend-ly* to the adverb *quick-ly*).
- They both can appear in the comparative form.

Because of these reasons, it has been proposed that they are two subcategories of one bigger category, A. However, most morphosyntacticians continue to distinguish them, and so we will, as well.

Key takeaways

- Adjectives may inflect into comparative and superlative forms, and adverbs may have a comparative form.
- There are many derivational morphemes that form adjectives from verbs, nouns, and other adjectives.
- Adverbs can be formed out of adjectives with the derivational morpheme *-ly.*
- In English, adjectives appear in between determiners and nouns or after the copula BE. Adverbs can appear anywhere *except* where adjectives can appear.

Check yourself!

8.5. FUNCTIONAL PARTS OF SPEECH

THIS IS A DRAFT

Functional and lexical parts of speech

So far we've only looked at lexical parts of speech. **Lexical** words are generally easier to define and encode the meaning of the sentence. There are also **functional** parts of speech. Functional parts of speech are much harder to define and have grammatical functions. Functional parts of speech include determiners, auxiliaries, modals, conjunctions, and complementizers. Adpositions are also a kind of functional part of speech, but they are more on the edge. For example, some prepositions, like *of*, are very functional and difficult to define, while other prepositions, such as *under*, are much easier to define.

Open and closed parts of speech

We can also distinguish between **open** and **closed** parts of speech. Open parts of speech can easily have new words added to them, while it is generally harder to add words to a closed parts of speech. Most lexical parts of speech are open and most functional parts of speech are also closed.

Frindle

Frindle is a children's book by Andrew Clements about a fifth-grade boy who decides to start calling pens by a new name, *frindle*, and gets his friends and classmates to start using the new word, too.

Based on what you know about open and closed parts of speech, do you think it's realistic that *frindle* would catch on? Why or why not?

Identifying functional parts of speech

Adpositions

If you recall from Section 1.2, some languages use **prepositions**, which appear before nouns, and some languages use **postpositions**, which appear after nouns. Both use the symbol P. The word ***adposition*** can be used if you don't want to distinguish between prepositions and postpositions.

The function of adpositions is to introduce a noun phrase, indicating its relation to the rest of the sentence.

Some examples of English prepositions include *before*, *after*, *to*, *from*, *of*, *under*, *over*, *through*, *near*, *off*, *on*, *for*.

Prepositions in English often appear before nouns and determiners, as shown in (1).

(1)	a.	**down** the stairs
	b.	**by** Tuesday
	c.	**under** the sea
	d.	**on** a boat
	e.	**near** Winnipeg
	f.	**from** my bag

English also has **intransitive** prepositions, which do not have objects, as shown in (2). Intransitive verbs usually appear either immediately after the verb or immediately after the object.

(2)	a.	I picked the bag **up**.
	b.	I picked **up** the bag.
	c.	The machine broke **down**.
	d.	The dean called the meeting **off**.
	e.	The dean called **off** the meeting.
	f.	She eased **up on** the brakes.

Determiners

The syntactic position of determiners can typically be described with reference to nouns. In English, determiners appear before nouns. There are several subclasses to the category of determiners, as listed in Table 1.

Table 1. Subclasses of the determiner category

Articles	*the, an, a*
Demonstratives	*this, that, these, those*
Quantifiers	*every, some, many, most, few, all, each, any...*
Numerals	*one, two, three...*
Possessive determiners	*my, your, his, her, its, our, their*
Some *wh*-words	*which, whose*

Determiners may agree with any of the inflectional features of the noun. Nominal inflectional features may also appear on the determiner instead of the noun.

Generally, the semantics of determiners is to pick out which specific entity or entities are being referred to. Whereas *cat* on its own could be referring to any cats in the world, or even the idea of cats in general, phrases like *the cat* indicates that there is a single contextually-relevant cat that is being referred to, *a cat* indicates that there is an individual relevant cat but doesn't specifically tell you which one, and so on.

Cross-linguistically, then, determiners also encode values such as definiteness and specificity. A definite determiner, like English *the*, indicates that the reference of the noun is identifiable in the context, while an indefinite determiner, like English *a(n)*, indicates that the reference of the noun is not identifiable. Specificity indicates whether the referent is a unique entity.

(3)	a.	I want to eat that cookie.	definite, specific
	b.	I want to eat the biggest cookie you can make.	definite, non-specific
	c.	My sister gave a cookie to her friend.	indefinite, specific
	d.	Let's go to the bakery and see if there's a cookie that looks good.	indefinite, non-specific

In some languages, definiteness and specificity might inflect on the noun instead of on the determiner.

Pronouns

Pronouns are an interesting part of speech. On the one hand, they have the same distribution as noun phrases, so they look like they might be a subcategory of nouns. On the other hand, unlike nouns, pronouns are a functional part of speech: they usually can be defined using only inflectional features, such as person, number, and gender. Like other functional parts of speech, pronouns are also a closed class. It is not easy to coin new pronouns.

Because of this, some people argue that maybe pronouns are a subclass of determiners rather than a subclass of nouns. Some evidence for this approach comes from (4). Although pronouns usually stand on their own,

in some contexts a pronoun can co-occur with a noun, as in (4a) *us students*. However, there are not contexts where a pronoun can co-occur with a determiner, as in (4b).

(4) a. Us students have gathered together to create a formal petition for less homework.

b. *The us/we would like less homework.

Neo-pronouns

psychological studies showing generic *he* does not get interpreted as gender-neutral cited in Bjorkman 2017 (Martyna 1978; MacKay & Fulkerson 1979; Gastil 1990; Foertsch & Gernsbacher 1997; Miller & James 2009).

Long history of trying to coin a replacement (citations) Newly coined pronouns are called *neopronouns*.

However, because pronouns belong to a closed class, neopronouns don't often catch on easily. It is more common for existing pronouns to be used generically, even when it results in awkward or clunky phrases, such as, generic *you*, coordinated *she or he*, or the very formal *one*.

Similarly, some people do not feel like gendered pronouns such as *she* or *he* describe them well. Some of these people choose to use pronouns like *they* or *it*, but others coin new pronouns that they feel fit them better. For example, in this blog, neopronoun Vagrant Gautam user explains why and how xe chose which pronouns feel the best to xem.

Many different English gender-neutral neopronouns have been coined and are in use. Some are shown below in Table 2.

Table 2. Examples of some gender-neutral neopronouns in English

nominative	accusative	genitive	reflexive
xe	xem	xyrs	xemself
hi	hir	hirs	hirself
ze	zir	zirs	zirself
ey	em	eirs	emself
fae	faer	faers	faerself

One reason functional categories are hard to coin is closed class words are more likely to have irregular inflection, so people don't know how to predict their inflection and the kinds of agreement they might trigger. For example, if you look at only one column in Table 2, do you feel like you could predict the forms in the other columns?

Neopronouns are being coined in other languages, too. For example, *iel* in French is a blend of the masculine pronoun *il* and the feminine pronoun *elle*. To read more about neopronouns in other languages, check out this article.

examples of neo-pronouns in other languages: *iel* in French

We might think that it would be better to just switch to using gender-neutral pronouns for everyone all the time.

(I saw a story that Finnish speakers were borrowing English pronoun pins to advertise their identity – fact check this)

Gender-neutral grammar is not correlated with gender ideology in culture (examples, citations)

Question words

Question words are a kind of pro-form, which means that they replace phrases of another part of speech. We can figure out what part of speech a question word is by looking at what it replaces. We can do that by answering the question and determining the part of speech of the answer. Some question words are ambiguous, and can be answered by phrases from more than one part of speech.

Table 3. Some possible parts of speech of question words in English

	Question word	Part of speech	Example
a.	who	N	Who ate the cookie? The cookie monster.
b.	what	N	What did the cookie monster eat? A cookie.
		D	What TV show did you watch? This one.
c.	where	P	Where did you watch it? On the couch.
		N	Where did you watch it? Here.
d.	when	N	When did you watch it? Last week.
		Adv	When will you watch it? Soon.
e.	why	C	Why did you watch it? Because I was bored.
		N	Why did you watch it? No reason.
f.	how	P	How did you watch it? On the TV.
g.	which	D	Which show did you watch? This one.
h.	whose	D	Whose cookie was eaten? Elmo's.

Tense

Tense markers in English include auxiliaries, modals, and the non-finite marker *to*, as listed in Table 4.

Table 4. Subclasses of the tense category

auxiliaries	HAVE, BE, and DO
modals	*will, would, should, shall, can, could, may, might, must, ought*
non-finite	*to*

In English, elements in the tense category appear in between the subject and the verb.
Each of the three tense subclasses have different properties.

- You can have more than one auxiliary, but only one modal and one *to* (and you can't have a modal and a *to* together, either)
    - Except for some dialects, like Appalachian English, where they do allow double modals
- auxiliaries and modals invert in question formation, but not *to*
- auxiliaries inflect for person, number, and tense like a verb, but modals and *to* do not

Auxiliaries, especially, seem to have many properties in common with verbs, so some people categorize auxiliaries as subclasses of verbs. But they also have some properties of T! If you continue studying syntax, you will discover why!
Distinguish between auxiliary and main verb uses of HAVE BE and DO

Conjunctions

Conjunctions (Conj) connect two units of the same type. English only has four conjunctions: *and*, *or*, *nor*, *neither...nor*, and *either...or*.

Complementizers

Complementizers (C or Comp) put one clause inside another. Some of the complementizers in English are *that*, *for*, *if*, and *whether*.

Negation

Negation (Neg): *not*

Particles

Free morphemes which encode only inflectional information are often called particles. Calling something a particle is a way to avoid classifying the word into a part of speech, often because it is difficult to identify its part of speech. In some cases, there may be only one or two particles that exhibit the same behaviour, and so there is really little basis for assigning it its own part of speech category.

Key takeaways

-

Check yourself!

References and further resources

For a general audience

Abraham, Amelia. 2020. Gender neutral pronouns are important: Not everyone identifies with 'he' or 'she'. *Vogue India*. https://www.vogue.in/culture-and-living/content/gender-neutral-pronouns-are-important-not-everyone-identifies-with-he-or-she

Conrod, Kirby and Vagrant Gautam. 2021. Guest Lecture in pronouns: Vagrant. *Medium*. https://kconrod.medium.com/guest-lecture-in-pronouns-vasundhara-1c61d3023e0c

8.6. COMMON PITFALLS

In this section, we will consider some common mistakes students make when classifying words according to part of speech.

Overlapping distribution

Some syntactic categories can appear in the same place as another. For example, both adjectives and nouns can appear after a determiner, as shown in (1).

(1)	a.	the	**happy**	woman
		D	Adj	N
	b.	the	**woman**	
		D	N	

This test is not actually ambiguous; you just have to remember that adjectives appear in between a determiner and a noun, while nouns will not typically be followed by another noun.

Ambiguous affixes

Some affixes are ambiguous. When an affix is ambiguous, we have to use additionally criteria to conclusively determine the part of speech of a word.

For example, *-ly* can form both adjectives and adverbs. In (2a), *-ly* attaches to the noun *friend* to form the adjective *friendly*. In (2b), on the other hand, *-ly* attaches to the adjective *quick* to form the adverb *quickly*.

(2)	a.	The friendly cat purred at me.
	b.	The cat ran quickly.

There are also two prefixes *un–* that can get mixed up. One *un–* is called the negative *un-* and it takes adjectives and turns them into negative adjectives, such as *unhappy*, *unable*, or *unease*. The other *un–* is called the reversive *un–* which means to undo an action, such as in *unlock*, *unbend*, and *unclog*. It's important not to

get these two homophonous morphemes mixed up because negative *un-* turns adjectives into adjectives and reversive *un–* turns verbs into verbs.

The **present participle**, which is formed by attaching *-ing* to a verb stem, can be used as a **progressive** verb, as a noun, or as an adjective. In some contexts, it is even ambiguous. We can tell that *annoying* in (3a) is a verb because it takes an object, *me*. We can tell that *annoying* in (3b) is an adjective because it appears in between a determiner and a noun. However, (3c) is ambiguous. The word *annoying* could be an adjective in a **copular** construction or a progressive verb.

(3)	a.	This rash has been **annoying** me.	progressive verb
	b.	This **annoying** rash won't go away.	adjective
	c.	This rash is **annoying**.	ambiguous

When present participles are used as nouns, as in (4a), the subject of the verb appears in genitive case and the object of the verb requires a preposition *of*. When a present participle is used as a non-finite verb, as in (4b), the subject will be in accusative case and the object does not need a preposition. If there is no subject or object, though, as in (4c), it is ambiguous between a verb and a noun.

(4)	a.	His **winning** of the race surprised me.	noun
	b.	Him **winning** the race surprised me.	non-finite verb
	c.	**Winning** surprised me.	ambiguous

The **past participle**, which is typically formed by attached *-en* to a verb stem, is also ambiguous. It can also be used as an adjective, as well as in forming **passive voice** and **perfect aspect**. In (5a), we can tell that *stolen* is a verb in the passive voice because of the *by*-phrase *by the thief*. In (5b), we can tell that *stolen* is a verb in the perfect aspect because of the auxiliary verb *had*. In (5c), we can tell that *stolen* is an adjective because it appears in between a determiner and a noun. In (5d), *stolen* is ambiguous; it could be an adjective in a **copular** construction or a passive verb.

(5)	a.	The diamond was **stolen** by the thief.	passive voice verb
	b.	The thief had **stolen** the diamond.	perfect aspect verb
	c.	The **stolen** diamond was recovered.	adjective
	d.	The diamond was **stolen**.	ambiguous

Homonyms

Some words have homonyms of a different category. For example, the word *bear* meaning the kind of animal is a noun in most contexts while the word *bear* meaning to carry is a verb in most contexts. For lexical words like *bear* and *bear*, it is fairly easy to keep them straight, but it is more difficult when one or more of the two words belongs to a functional category.

For example, there are two words *to* with different functions, and they belong to different parts of speech: it can be a preposition or it can be a **non-finite** tense marker. In (5a), we can tell that *to* is a preposition, because it is followed by a noun phrase. In (5b), we can tell that *to* is a non-finite marker because it is followed by a verb.

(5)	a.	I went **to** the store.	noun
	b.	I want **to** go home.	verb

Nouns that look like prepositions

There are some nouns that can indicate location or origin, which some students mistake for prepositions. These nouns may even have equivalent prepositions with the same meaning. For example, *back* in (1a) and *top* in (2a) are both nouns. We can tell that they are nouns because they cannot introduce the following noun phrase directly; instead, they need the preposition *of*.

(1)	a.	to the **back** of the door	noun
	b.	**behind** the door	preposition
(2)	a.	on **top** of the image	noun
	b.	**above** the image	preposition

Key takeaways

- Some criteria for determining part of speech can be ambiguous. Be aware of these ambiguities and use two or more criteria to be sure.
- The present and past participle forms can be used either as verbs, as nouns, or as adjectives.

Some instances are ambiguous.
- Some words have homonyms in different categories. Be aware of these, especially the difference between preposition and non-finite *to*.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=4406#h5p-54

8.7. CASE STUDY: FRENCH *POSSESSIVE ADJECTIVES* ARE NOT ADJECTIVES

Connor Mark

One common difference between descriptive grammars written by linguists and pedagogical grammars is how they use technical terminology. For instance, in French textbooks, the term *adjectifs possessifs* 'possessive adjectives' is commonly used. This might suggest to students that these words are actually adjectives. In reality, they are determiners. Hence, we will be calling them *possessive determiners* to be consistent with our final conclusion. We can analyze how adjectives, determiners, and possessive determiners are used in sentences to demonstrate this.

Before showing why possessive determiners are not adjectives, it might be useful to explain why someone might think that they are adjectives in the first place. First, adjectives in French have to agree in number (singular vs. plural in French) with the noun that they modify. When they fail to agree, the result is an ungrammatical phrase. As shown in (1), since the noun *personnage* 'character' is singular, adjective *principal* 'main' in its singular form can modify it. However, if we try using the plural form, *principaux* 'main', the result is an ungrammatical phrase.

(1)	a.	*French*		
		le	personnage	principal
		the.SG	character.SG	main.SG
		'the main character'		

	b.	*le	personnage	principaux
		the.SG	character.SG	main.PL

The same is true of possessive determiners. They must also agree in number with the noun that they modify. In (2), if we use singular possessive determiner *mon* 'my' with the singular noun *stylo* 'pen', the result is grammatical. Again, however, if we try using plural possessive determiner *mes* 'my' with the singular noun, we get an ungrammatical phrase.

(2) a. *French*

mon	stylo	
my.SG	pen.SG	
'my pen'		

b. *mes stylo

my.PL	pen.SG

Additionally, both adjectives and possessive determiners need to agree in noun class (masculine vs. feminine in French) as well. As shown in (3)[1], the feminine noun *catin* 'doll' must go with feminine possessive determiner *ma* 'my' and feminine adjective *blonde* 'blonde'. If either is replaced with its masculine variant, *mon* 'my' and *blond* 'blond', then the phrase is ungrammatical.

(3) a. *French*

ma	catin	blonde
my.F	doll.F	blonde.F
'my blonde doll'		

b. *mon catin blonde

my.M	doll.F/td>	blonde.F

c. *ma catin blond

my.F	doll.F	blond.M

d. *ma catin blond

my.M	doll.F	blond.M

In terms of agreement, then, adjectives and possessive determiners are quite similar and so it is not unreasonable for someone to group the two together in some way. However, other determiners, such as articles, must also agree in number and noun class with their corresponding noun.

1. These examples were based on a passage from Manitoban author Gabrielle Roy's *La route d'Altamont*.

For number agreement, in (4), singular noun *cheval* 'horse' must go with singular determiner *le* 'the' while plural noun *chevaux* 'horses' must go with plural determiner *les* 'the'.

(4) a. *French*

le	cheval
the.SG	horse.SG

'the horse'

b. *les cheval
the.PL horse.SG

c. *le chevaux
the.SG horse.PL

d. les chevaux
the.PL horse.PL
'the horses'

For noun class agreement, in (5), masculine noun *livre* 'book' can co-occur with masculine determiner *le* 'the' but not feminine determiner *la* 'the'.

(5) a. *French*

le	livre
the.M	book.M

'my book'

b. *la livre
the.F book.M

At this stage, adjectives, determiners, and possessive determiners are all behaving the same way. Intuitively, however, speakers of French know that determiners and adjectives do not belong to the same category. As such, possessive determiners can only logically be adjectives, determiners, or some third category.

One way to test the category of possessive determiners is to look at ways in which adjectives and determiners

behave differently and then see if possessive determiners act more like adjectives, more like determiners, or in a way different from each of the two.

The first test concerns the relative word order of adjectives, determiners, and possessive determiners against cardinal numbers. As a reminder, cardinal numbers are used for counting things (*one*, *two*, *three*, etc.). This can be contrasted with ordinal numbers for example, which order things (*first*, *second*, *third*, etc.). In French, determiners such as *les* and possessive determiners such as *mes* go before cardinal numbers, as shown in (6) and (7) respectively.

(6) *French*

les	deux	femmes
the	two	women

'the two women'

(7) *French*

mes	deux	frères
my	two	brothers

'my two brothers'

Conversely, adjectives come after cardinal numbers. As shown in (8), the adjectives *jaunes* 'yellow' and *jeunes* 'young' come after the cardinal number *trois* 'three'. This is true whether the adjective comes before or after the noun that it modifies.

(8) a. *French*

les	trois	livres	jaunes
the	three	books	yellow

'the three yellow books'

b.

les	trois	jeunes	garçons
the	three	young	boys

'the three young boys'

The second test that can be used is the deletion test. In a sentence, removing an attributive adjective such as *petit* 'little' does not make a sentence ungrammatical, as shown in (9).

(9) a. *French*
Elle aime le petit chien.
she likes the little dog
'She likes the little dog.'

b. Elle aime le chien.
she likes the dog
'She likes the dog.'

However, sentences do become ungrammatical if we remove a determiner (10) or a possessive determiner (11).

(10) a. *French*
Il veut un chat.
he wants a cat
'He wants a cat.'

b. *Il veut chat.
he wants cat

(11) a. *French*
Je vois mes amis.
I see my friends
'I see my friends.'

b. *Je vois amis.
I see friends

Finally, we can use the substitution test. Typically, you can replace words of one syntactic category with a word of the same category but not of another category. In French, if we try to substitute an adjective like *bon* 'good' for another adjective like *grand* 'big', the result is grammatical (12a-b). Conversely, if we try to replace an adjective with a determiner (12c) or a possessive determiner (12b), the result is ungrammatical.

(12) a. *French*

On	a	un	bon	livre.
we	have	a	good	book

'We have a good book.'

b.

On	a	un	grand	livre.
we	have	a	big	book

'We have a big book.'

c.

*On	a	un	le	livre.
we	have	a	the	book

d.

*On	a	un	mon	livre.
we	have	a	my	book

Additionally, if we have a sentence with a determiner in it, it can be freely substituted by another determiner (13a-b) or a possessive determiner (13c) but not an adjective (13d).

(13) a. *French*

Elle	mange	la	poutine.	
she	eats	the	poutine	

'She's eating the poutine.'

b.

Elle	mange	une	poutine.	
she	eats	a	poutine	

'She's eating a poutine.'

c.

Elle	mange	ma	poutine.	
she	eats	my	poutine	

'She's eating my poutine.'

d.

*Elle	mange	grande	poutine.	
she	eats	big	poutine	

Note that the examples directly above also show that possessive determiners can be freely substituted by determiners but not adjectives.

Based on the three tests above—the relative word order compared to cardinal numbers, the deletion test, and the substitution test—it can be seen that determiners and possessive determiners behave in one way whereas adjectives behave in a different way. This suggests that possessive determiners are in fact determiners and not adjectives. This section has shown only three tests that provide such evidence, but there are in fact more.

Hopefully this section has shown you that pedagogical grammars often have analyses that do not fit well into modern morphosyntactic analysis and that even provide more difficult explanations in some cases. In terms of possessive determiners, if they were adjectives, there is no natural explanation for why they always go before nouns, come after cardinal numbers, cannot co-occur with other determiners, can be substituted by determiners and not adjectives, and cannot be deleted when "modifying" a noun. Under the analysis that possessive determiners are determiners, learners know that they can simply put the possessive determiners in place of where they would put a determiner. The only reason why these problems seem to seldom come up among anglophone learners is that these constructions are quite similar to what we have in English.

Key takeaways

- So-called "possessive adjectives" in French actually pattern with determiners.
- Evidence that they are possessive determiners include agreement patterns, word order, obligatoriness, and substitution.
- Pedagogical grammars and traditional language-specific terminology doesn't always match up with linguistic analysis.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=3625#h5p-47

Consult the following data in answering the check yourself questions:

(14) a. *French*

Nous	buvons	nos	boissons.
we	drink	our	drinks

'We're drinking our drinks.'

b. *Nous buvons boissons.

we	drink	drinks

c. Nous buvons les boissons.

we	drink	the	drinks

'We're drinking the drinks.'

References and further resources

Sources for examples

Roy, Gabrielle. 1966. La route d'Altamont. Montréal: Editions HMH.

CHAPTER 8 PRACTICE EXERCISES

Data analysis

Exercise 1. Identifying parts of speech

Beginner-Intermediate [8.1-8.6]

Label the parts of speech of all of the words in the following sentences taken from *A Tale of Two Cities* by Charles Dickens.

Labels you should use: noun (N), pronoun (PN), determiner (Det), verb (V), tense (T), preposition (P), adjective (Adj), adverb (Adv), conjunction (Conj), complementizer (Comp), and negation (Neg).

a. There were a king with a large jaw and a queen with a plain face, on the throne of England.

b. With drooping heads and tremulous tails, they mashed their way through the thick mud, floundering and stumbling between whiles, as if they were falling to pieces at the larger joints

c. There was a steaming mist in all the hollows, and it had roamed in its forlornness up the hill, like an evil spirit, seeking rest and finding none.

d. Two other passengers, besides the one, were plodding up the hill by the side of the mail.

e. In those days, travellers were very shy of being confidential on a short notice, for anybody on the road might be a robber or in league with robbers.

f. Our booked passenger showed in a moment that it was his name.

g. The guard, the coachman, and the two other passengers eyed him distrustfully.

h. Take that message back, and they will know that I received this, as well as if I wrote.

i. I shall not go to bed till night; but I want a bedroom, and a barber.

j. He had a healthy colour in his cheeks, and his face, though lined, bore few traces of anxiety.

Exercise 2. Identifying parts of speech

Intermediate [8.1-8.5]

Consider the Spanish examples below. Based on this data, form hypotheses about criteria that can be used to identify parts of speech in Spanish. If it's helpful, use the chart included below to organize your answers. Examples (1b)-(1d) are taken from *Como agua para chocolate* by Laura Esquival.

Note: There may not be enough data to fill out all of the cells in the chart, but see what you can find!

Glossing abbreviations: CONT = continuous (ongoing action, DEM = demonstrative (i.e., *this* or *that*), DET = determiner, DIM = diminutive (a small thing), F = feminine, M = masculine, PL = plural, PRES = present, PROG = progressive, REFL = reflexive, SG = singular.

(1)	a.	La	viej-it-a	está	espera-ndo	el	autobús.	
		DET.F	old-DIM-F	be.PRES.3SG	wait.PROG	DET.M	bus	
		'The old lady is waiting for the bus.'						
	b.	Éste	dorm-ía	apacible-mente	junto	a	su	herman-a.
		DEM.M	sleep-PAST.3sg	peaceful-ly	together	to	his	sibling-F
		'This one slept peacefully next to his sister.'						
	c.	Tita	la	escuchaba	con	gran	interés.	
		Tita	3SG.F	listen.PAST.CONT.3SG	with	big	interest	
		'Tita was listening to her with great interest.'						

d.	Giró	la	cabeza	y	sus	ojos	se	encontraron	con	los	de	Pedro.
	turn.PAST.3SG	DET.F	head	and	her.PL	eye.PL	REFL	met.PAST.3PL	with	DET.PL	of	Pedro
	'She turned her head and her eyes met Pedro's.'											

Part of speech	**Syntactic distribution**	**Inflectional morphology**	**Derivational morphology**
Noun			
Verb			
Adjective			
Adverb			
Preposition			
Determiner			

Exercise 3.

One way that we can classify verbs into subcategories is through a property called *lexical aspect*.

Exercise 4. Language evolution

Intermediate [8.1, 8.5]

Consider the following sentences. What is the part of speech of *because* in the sentences in (2)?

(2) a. I ate my lunch early because I was hungry.

b. Because of the changes in communication style since the advent of the Internet, new types of slang are being used.

Now consider the sentences in (3), which represent newer, more playful use of the word *because*. What is the part of speech of *because* in the sentences in (3).

(3) a. I have secrets because reasons.

b. Language is changing because Internet.

See also the book Because Internet *by Gretchen McCulloch on the ways that Internet has been changing language!*

Communication and study skills

Exercise 5. Predicate-argument flexibility

Intermediate [8.9]

The following examples, from a language called Gitksan, illustrate a property of several languages called *predicate-argument flexibility*. Explain how these examples demonstrate why we should not use semantic criteria to identify the part of speech of a word. Gitksan is spoken in British Columbia and belongs to the Tsimshianic language family.

Abbreviations

AFF = affirmative particle, CN = common noun connective, SX = subject extraction marker.

(3)	a.	hanaq̓=ɬ	sim̓o:gid(-id)=ist
		woman=CN	chief(-SX)=AFF
		‘The chief is a woman.’	
	b.	sim̓o:git=ɬ	hanaq̓=ast
		chief=CN	woman=AFF
		‘The woman is a chief.’	
(4)	a.	hanaq̓=ɬ	c̓aw-ad=ast
		woman=CN	smart-SX=AFF
		‘The smart one is a woman.’	
	b.	c̓aχʷ=ɬ	hanaq̓=ast
		smart=CN	woman=AFF
		‘The woman is smart.’	
(5)	a.	w̓itxʷ=ɬ	hanaq̓=ast
		arrive=CN	woman=AFF
		‘The woman arrived.’	
	b.	hanaq̓=ɬ	w̓itxʷ-id=ist
		woman=CN	arrive-SX=AFF
		‘The one who arrived is a woman.’	

[Gitksan; Davis et al. 2014: 196-197]

Exercise 6. Word creation

Intermediate [8.1, 8.5]

Invent a new word in a language you speak and compose a dictionary entry for it by answering questions (a)-(f) and then reflect on the properties of your word by answering questions (g)-(h).

a. How do you spell your new word?

b. What language does your new word belong to?

c. How do you pronounce your word? (Use IPA if you know it)

d. What is the part of speech of your new word?

e. What is the definition of your new word?

f. Provide an example sentence containing your new word. If your new word is not an English word, gloss your example.

g. Identify morphological or syntactic characteristics that support the part of speech you identified in question (d). Explain how it supports your classification. You can make reference to your example sentence in (f).

h. How hard would it be to get your word to catch on? Explain, making reference to the part of speech of your invented word.

Note: This assignment is inspired by *Frindle*, a book by Andrew Clements about a kid who invents a new noun, *frindle*, which means 'pen,' and convinces all his friends to use the word, too. The new word catches on and even makes it into the dictionary a few years later.

Exercise 7.

Part 1. Modals and auxiliaries are each subcategories of the T category. Write a short 1-2 page essay based on the data below explaining (a) why they should both be considered part of the T category and (b) but why they should be split into two subcategories.

multiple auxiliaries, not multiple modals

inflected auxiliaries, not inflected modals

Part 2. *Must* and *have to* have almost the same meaning, but *must* is a modal. Is *have to* also a modal? Explain whether *have to* has the same properties as modals. Construct examples using *have to* to support your argument. Hint: You can use the data from Part 1 as a starting part.

Exercise 7. Slang analysis

Intermediate [8.1-8.6]

Think of a slang word or a word that has only recently entered English or another language you speak. What part of speech is it? Construct examples that back up your answer. Is it an open or closed part of speech?

Language journal

At the end of Chapter 1, we introduced the Language Journal. You should have picked a language to build a profile for as you work through this textbook. You can add to your profile

with these questions.

Exercise 8. Identifying parts of speech

Intermediate [8.1-8.6]

Find examples of nouns, verbs, adjectives, and adverbs in your language. What syntactic and morphological criteria can you use to identify them?

References and further resources

Sources for examples

Davis, Henry, Carrie Gillon, and Lisa Matthewson. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language* 90(4): e180–e226.

PART II. MORPHOLOGY

Chapters 9-15 focus on morphology.

9. WORDS

Words are important units in both the fields of morphology and syntax. In this chapter, we will learn all about words, including how to form new words, how to define words, and how different languages package the content of their words in different ways.

Learning objectives

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

- Classify words by their formation process, using an etymological dictionary when necessary.
- Use technical terms such as *lexeme, word form*, *word token*, *citation form*, *word family*, and *paradigm* appropriately in context.
- Classify languages based on their degree of synthesis, fusion, and exponence.
- Explain the problems in defining *word*.

9.1. WORD FORMATION PROCESSES

Sometimes, new words are added to a language, either due to **neologism** [niˈɑləd͡ʒɪzm̩] (the creation of a new word within the language itself) or due to **borrowing** (when one language, the recipient language, adapts words or other grammatical features from another language, the donor language).

Taboo avoidance is a special kind of neologism in which a word is replaced (either by a modified pronunciation or by a newly invented word) to avoid an obscene or taboo meaning, such as *gosh* for *God*. Taboo meanings are often centred on sexuality, bodily waste, and religion.

In this section, we will be using < and > symbols from historical linguistics to show the historical direction of change. For example, *X* > *Y* means that the older form X was replaced by the newer form Y, while *Y* < *X* means the opposite, that the newer form Y originated from the older form X.

Deriving

One of the most common ways to form new words is by adding new **morphemes**. There are two main kinds of morphemes, **inflectional** morphemes and **derivational** morphemes. We will learn more about the difference between them in Chapter 9, but for now, you can think about the two terms this way: inflectional morphemes add grammatical information to a word, such as plurality or a tense marker, while derivational morphemes create a new word.

Some examples of English derivational processes are shown in Table 1.

Table 1. Examples of derivation in English

	Stem	Derivational morpheme	Derived word
a.	new	-ness	newness
b.	play	-ful	playful
c.	playful	-ness	playfulness
d.	friend	-ly	friendly
e.	friendly	un-	unfriendly
f.	friend	-ship	friendship

As shown by Table 1 examples (c) and (e), you can add more than one derivational morpheme to the same word.

When a grammatical pattern can be extended in predictable manners, it is called **productivity**. Some derivational morphemes are productive (like *-ish*) and others are not (like *-ness*), as shown in Table 2. While *-ish* can be used with all six stems, *-ness* can only be used with examples (a) to (c). We will discuss productivity more in Chapter 13.

Table 2. The productivity of *-ish* and *-ness*

	Adjective stem	with *-ish*	with *-ness*
a.	new	newish	newness
b.	blue	bluish	blueness
c.	slow	slowish	slowness
d.	long	longish	*longness
e.	strong	strongish	*strongness
f.	popular	popularish	*popularness

Conversion is a special kind of derivation, in which you change the part of speech of a word without changing the form of the word, for example changing the verb *walk* to the noun *walk*.

Borrowing

Another way that a language might acquire new words is through contact with another language.

In borrowing, the borrowed words are called **loanwords** and are often **nativized**, which means that the pronunciation, and sometimes the morphology, is changed to match the regular patterns of the recipient language. For example, the Innu word for *pancake* is *tekanep*, taken from French *des crêpes* /dekʀɛp/. The determiner *des* is incorporated into the word, the /d/ is devoiced, /ʀ/ becomes /n/, and a vowel is added to break up the consonant cluster.

Borrowing can happen for a variety of reasons, including prestige (e.g., numerous French and Latin loanwords in academic English) and need (e.g., animal words such as *moose, caribou, chipmunk, skunk, opossum*, and *raccoon* have been borrowed into English from Algonquian languages). The direction of borrowing can often be inferred based on knowledge of phonology, sound change, morphology, cognates, real world geography, ecology, cultural differences, history, and so forth. Borrowing often occurs within particular semantic domains (food, religion, politics, and so forth), especially when very different cultures come into contact.

The same source word may take different paths and be borrowed multiple times into the same language.

This may be because two languages are in contact over long periods of time and borrow the same word at multiple points in its history, with different pronunciations as the source language undergoes change. It also happens if the same word is borrowed from different dialects or varieties. For example, *captain, chef*, and *chief* all originate from *caput*, the Latin word for 'head', via French. *Captain* and *chief* were both borrowed in about the 14th century from the Old French words *capitaine* and *chief*. *Capitaine* came from the Latin *capitaneus*, a derived form of the word *caput* to mean 'leader' rather than 'head'. Over the next few centuries, the French word *chief* became *chef*. It was then borrowed again into English in the 18th century as part of the phrase *chef de cuisine* 'head of the kitchen'.

A **calque** or **loan translation** is a specific type of borrowing that includes a direct translation of an idiomatic loanword, rather than a phonological nativization, for example English *black market* < German Schwarzmarkt, literally 'black market'.

Root creation

Sometimes a word is invented without basis on any previously existing words, which is called **root creation**. This happens most often in product naming and fiction. Some examples include *google* 'a search engine brand' or *lightsaber* 'a laser sword from the series '*Star Wars*'. Sometimes, a product name expands to include all similar objects, regardless of brand. This is called **genericization**. Some examples include *kleenex, thermos,* or *popsicle*, which now refer to any facial tissue, double-walled food storage container, or frozen treat on a stick, regardless of brand. However, all three started as brand names originally. The word *google* is also well on its way in genericization. Do you call it *googling* even if you're using a different search engine?

Shortening

Shortening an existing word is called **clipping** or **truncation**, like with *math* < *mathematics*, *phone* < *telephone*, or *flu* < *influenza*.

Sometimes people reanalyze a monomorphemic word to be polymorphemic, and then remove the extra morphemes. This is called **backformation**. For example, the noun *editor* existed before the verb *edit*. However, people misinterpreted the /-ər/ ending of *editor* to be the same as the derivational morpheme *-er* 'one who does V', and assumed that there must be a verb *to edit* to go with the noun *editor*. It is impossible to tell whether a word is a backformation without looking up which form came first in an etymological dictionary. Other examples of backformation include *haze* < *hazy*, *diagnose* < *diagnosis*, or Old English *pise* 'pea' (plural *pisan*) > Modern English *pea* (plural *peas*), with the singular formed by stripping of the supposed plural *-s* in the pise, even though *pise* was originally singular and had no plural morpheme.

Combining

Sometimes new words are coined by combining existing words. If you combine two roots and keep the whole of both roots, you've made a **compound word**, such as *rainbow*, *blackboard*, or *greenhouse*. In contrast, if one or both words is not fully preserved in the resulting form, this is called a **blend** or **portmanteau**. Examples of blends include *motor hotel* > *motel*, *smoke* + *fog* > *smog*, or *breakfast* + *lunch* > *brunch*.

If you only keep the first letter (or two) of each word in a phrase, it's an **acronym** if you pronounce it like a word, as with *laser* < *light amplification by stimulated emission of radiation* or *radar* < *radio detecting and ranging*. On the other hand, if each letter in the abbreviation is pronounced as a letter, it is known as an **initialism** or **alphabetism**. Some examples include FBI > Federal Bureau of Investigation or UK > United Kingdom.

Did you know?

The Internet slang initialism *LOL* is claimed to have originated in Canada! Wayne Pearson, an IT consultant from Calgary, claims to have coined it as a teenager on a chatroom for Calgarians in the 1980s.

When a word that already exists is claimed to be an acronym, this is known as a **backronym**. There has been a recent trend of backronyms going viral on social media, masquerading as the true origin of the word. For example, a false idea now circulating says that *news* stands for *notable events weather and sports*, when the term really comes from the word *new* from the early 15th century. It is best to check an etymological dictionary before clicking "share" on such posts.

Etymological dictionaries

People often guess the history of a word incorrectly, and **folk etymologies** which are not historically accurate often circulate in communities. Sometimes folk etymologies can even drive language change! For example, *penthouse* comes from the Old French word *apentis* meaning 'attached building.' Due to similarity in meaning and pronounciation with the word *house*, its spelling changed over time to become the modern *penthouse*.

If you want to know the history of a word, you can look it up in an etymological dictionary. The best etymological dictionary for English is the *Oxford English Dictionary* (OED). The OED is paywalled, but you can likely access it online through your university's library.

Another good online etymological dictionary to explore is the Online Etymology Dictionary by Douglas Harper. Although it is not technically an academic source, this is a free and reliable etymological dictionary of English.

There are many more etymological dictionaries and other resources out there that cover a wide variety of languages! Check your university library to see what you can find!

Key takeaways

- New words are added to language in a variety of ways. Neologisms can come from other languages, which are called *loanwords*, they can be formed within the language itself, or they can be coined in a process called *root creation*.
- Word formation processes that involve shortening an existing word include *clipping* and *backformation*.
- Word formation processes that involve combining existing words including *compounding*, *blending*, *initialisms*, and *acronyms*.
- The symbols < and > are used to show the direction of change in the study of historical linguistics. These symbols point at the newer form.
- You can look up the origins of a word in an etymological dictionary, such as the *Oxford English Dictionary* (OED).

Check your understanding!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1629#h5p-33

References and further resources

Attribution

↪️ This section was adapted from lecture notes prepared by Nathan Sanders. Used with permission.

For a general audience

🧠 CrashCourse. 2020. Language Change and Historical Linguistics: Crash Course Linguistics #13. https://www.youtube.com/watch?v=dNkMC92kFLA.

🎉 Hudes, Sammy. 2015, October 7. What's it like to coin the term LOL? *Calgary Herald*. https://calgaryherald.com/news/local-news/whats-it-like-to-coin-the-term-lol.

Reference materials

📑 Harper, Douglas. 2024. *Online Etymology Dictionary*. https://www.etymonline.com.

📑 Oxford University Press. 2023. *Oxford English Dictionary*. https://www.oed.com.

9.2. DIFFERENT MEANINGS OF *WORD*

At the beginning of this textbook, we defined *morphology* as the structure of words. Words are also important units manipulated by the syntax. But what is a word? It turns out that the definition of *word* is trickier than it may at first seem, especially when we consider the different kinds of words we see in different languages.

One reason wordhood is difficult to define is that, even within the same language, we use *word* to mean lots of different things.

Consider the following situations. What is meant by *word* in each of them?

(1)	a.	You are playing a word game with friends or family in which you have to rearrange different letter tiles or cards to create a word. You rearrange them to create the word *furnitures.* You argue about whether it is a word.
	b.	You are assigned a paper for a class that is supposed to be 500 words long.
	c.	In a grammar class, you learn that the passive voice must be formed with the word BE; for example, *was seen, is seeing, has been seen.*
	d.	It is estimated that the English language includes approximately 1 million words (Merriam-Webster).

To sort out the differences between these different uses of *word*, linguists use the technical terms *lexeme*, *word token*, and *word form*.

A **lexeme** is a word in the abstract sense, including all of the different inflectional forms that a word can take. In example (1c), what the teacher means is that the passive voice is constructed using the lexeme BE, but it doesn't have to be *be* itself, it can be one of its other forms. In (1d), they are estimating that English includes approximately 1 million lexemes. Lexemes are usually rendered in all caps.

A **word form** is a a particular form of a lexeme. In (1a), you are arguing about whether the lexeme FURNITURE can appear with plural inflection, or, in other words, whether FURNITURE has a plural word form. In (1c), your teacher has taught you that the passive voice must have one of the word forms of the lexeme BE, including *was*, *is*, and *been*.

Technically, a lexeme is a set of all the inflected word forms associated with the root word. However, it is cumbersome to list all of the words in the set each time we refer to it. The word form we use to refer to a lexeme is typically called the **citation form**. It is the citation form that will typically be listed in the dictionary. Languages differ as to which word form is used as the citation form, depending on convention and what makes sense given the morphological properties of the language. For example, in English, we usually use the non-finite form of verbs and the singular form of nouns as the citation form.

The collection of all the word forms in a lexeme is called the **paradigm** for that lexeme. In other words, a

paradigm is the set of all the inflected forms of a word. On the other hand, words that are related by derivation each belong to a different lexeme. A set of words related by derivation is called a **word family**.

A **word token** is a particular occurrence of a word. In example (1b), the teacher wants you to write a paper that has 500 word tokens. Your paper will likely use the word *the* several times, but you will count each instance as a separate word (unlike if you were counting lexemes).

Let's look at an example together. Below, you will see the entry for *buy* from the *Canadian Oxford Dictionary* (Barber 2005).

buy

▶ *verb* (**buys, buy*ing** *past* and *past participle* **bought**)

1. [*transitive*] purchase; obtain in exchange for money, etc.
    - serve to obtain: *money can't buy happiness.*
2. [*transitive*] procure (the loyalty, etc.) of a person by bribery, promises, etc.
    - win over (a person) in this way.
3. [*transitive*] get by sacrifice, great effort, etc.: *dearly bought* || *bought with our sweat.*
4. [*transitive*] *informal* accept, believe in, approve of: *the police bought our story.*
5. [*intransitive*] be a buyer for a store, etc.: *buys for the furniture chain.*

▶ *noun* *informal* a purchase: *that car is a good buy.*

The entry for *buy* includes at least two **homonymous** lexemes, the verb BUY and the noun BUY. These two words are related and belong to the same word family, but must be considered different lexemes because their relationship is derivational rather than inflectional, as the noun is formed by **conversion** from the verb. We might also wonder if the five different definitions of the verb *buy* should also correspond to different homophonous lexemes, or whether they should all be considered different uses of the same lexeme. I will save the finer points of this debate for your semantics class, but we can tell that this **lexicographer**, at least, decided they were one lexeme (or else they would have listed buy_1 verb and buy_2 verb, etc.), likely because the senses listed here all seem to be related to the same core sense by means of extension and metaphor. The citation form for both lexemes is *buy*.

The entry for the verb BUY has three word forms listed in addition to its citation form at the beginning

of the entry: *buys*, *buying*, and *bought*. The entry for the noun BUY does not have any additional word forms listed, likely because the only other word form, the plural *buys*, is completely regular.

Overall, this dictionary entry includes 10 word tokens of different forms of *buy*. There is the headword, which begins the dictionary entry. The second, third, and fourth word tokens are the examples of the different word forms, *buys*, *buying*, and *bought*. The remaining six tokens are found in the examples, such as in *money can't* ***buy*** *happiness*.

Key takeaways

- The word *word* has lots of different meanings. There are technical terms used to distinguish these meanings.
- A *lexeme* is the abstract sense of a word, including all of its different inflectional forms.
- A *word form* is a particular form of a lexeme.
- A *word token* is a particular occurrence of a word.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1621#h5p-32

References and further resources

For a general audience

🧠 Merriam-Webster. How many words are there in English? https://www.merriam-webster.com/help/faq-how-many-english-words.

Sources for examples

Barber, Katherine. 2005. *The Canadian Oxford Dictionary.* 2nd ed. Oxford: Oxford University Press.

9.3. PACKAGING WORDS AND MORPHEMES

Languages differ based on how they package meaning into a single morpheme, and how they package morphemes into a word. Traditionally, languages have been classified into four morphological types based on the structure of the word: **isolating**, **agglutinative**, **fusional**, and **polysynthetic**.

Traditional morphological types

Isolating

Isolating languages, also known as analytic languages, have an almost one-to-one morpheme-to-word ratio. Words in isolating languages are often unaffixed bare roots. Bound morphemes (morphemes that cannot stand alone) are very infrequent.

Yoruba, as shown below in (1), is an example of an isolating language. Each word in these examples contains only a single morpheme. Tense markers, such as the past tense marker *ti* in (1a) and the future marker *ma* in (1c), are independent words.

(1)	a.	ọkunrin	ti	jo.	[Yoruba][1]
		boy	PAST	dance	
		'The boy danced.'			
	b.	ọkunrin	Ø	jo.	
		boy	PRES	dance	
		'The boy is dancing.'			
	c.	ọkunrin	ma	jo.	
		boy	FUTURE	dance	
		'The boy will dance.'			

1. The Yoruba data in these examples was provided by Tolani Akinlade.

Agglutinative

In agglutinative languages, there is a much higher morpheme-to-word ratio. In other words, there are many morphemes in a single word. Usually, there is little interaction between each morpheme and each morpheme contains a single piece of grammatical information; in other words, there is more-or-less one-to-one matching between morphemes and meaning. However, there is still usually only one root morpheme in most words (with the exception of compound words).

Turkish, as shown in Table 1, is an example of an agglutinative language.

Table 1. The paradigm for *ev* 'house' in Turkish (adapted from Lieber 2016: 148-149)

	singular	**plural**
nominative	ev	ev-ler
definite-accusative	ev-i	ev-ler-i
genitive	ev-in	ev-ler-in
dative	ev-e	ev-ler-e
locative	ev-de	ev-ler-de
ablative	ev-den	ev-ler-den

In Turkish, nouns inflect for number (singular or plural) and **case** (nominative, definite-accusative, etc.). Case marks for the role that the noun plays in the sentence (e.g., subject, object, object of preposition, etc.). Each morpheme in Table 1 has exactly one meaning. For example, *-ler* means plural, *-i* means definite-accusative, and so on. To form a word in Turkish, multiple morphemes are strung together.

Fusional

Fusional languages, on the other hand, have morphemes that express multiple pieces of grammatical information simultaneously. There is almost no one-to-one match between pieces of meaning and number of morphemes. There also tend to be multiple morphemes in a single word.

Latin is an example of a fusional language, as shown in Table 2.

Table 2. The paradigm for *puella* 'girl' in Latin (adapted from Lieber 2016: 149)

	singular	**plural**
nominative	puell-a	puell-ae
gentive	puell-ae	puell-ārum
dative	puell-ae	puell-īs
accusative	puell-am	puell-ās
ablative	puell-ā	puell-īs

Latin nouns, like Turkish, also inflect for case and number. However, unlike Turkish, we cannot separate the parts of the word that indicate number and the parts that indicate case. The morpheme *-am*, for example, simultaneously encodes both accusative and singular, while the morpheme *-ārum* simultaneously encodes both genitive and plural. We cannot separate either *-am* nor *-ārum* into two parts; they are each one morpheme with two meanings. There is no morpheme that repeats down either the singular or plural columns, nor any morphemes that repeat across each case row.

Polysynthetic

Finally, polysynthetic languages often have a single word representing meanings that are associated with whole sentences in other languages. These words contain long strings of roots and affixes. Often the object, and sometimes the subject, are inflected onto the verb.

Inuktitut is an example of a polysynthetic language, as shown in (2).

(2)	a.	ayagciqsugnarqnillruuq						
		ayag-	ciq-	yugnarqe-	ni-	llru-	u-	q
		go-	FUT-	probably-	claim-	PAST-	INDIC.INTR-	3SG
		'He said he would probably go.'						
	b.	uqa-	limaar-	vi-	liu(ng)-	inna-	nngit-	-tunga
		speak-	all.of-	NOM-	make-	always-	NEG-	DEC.1SG
		'I was not always making libraries.'						

(Inuktitut; Compton and Pittman 2010: 2168)

Both of the Inuktitut words in (2) contain multiple morphemes. The example in (2a) contains root words that are associated with more than one part of speech, including the verbs *ayag-* 'go' and *ni-* 'claim' and the adverbial *yugnarqe-* 'probably'. The example in (2b) contains inflectional morphemes that are associated with

nouns, like the nominative case marker *vi-* as well as inflectional morphemes that are associated with verbs, like the declarative first-person singular marker *-tunga*. Each of these words are translated into a whole English sentence. Example (2a) even has two clauses in the translation!

In speech, we don't usually pause in between words, so why do we think that the examples in (2) are each single words? There are a few reasons. First, speakers have intuitions that they are single words. Second, the morphemes are mostly bound morphemes that cannot stand on their own. Third, many phonological processes cross morpheme boundaries but do not cross word boundaries, giving us a big clue about where word boundaries are.

Polysynthetic languages are sometimes confused with agglutinative languages. Unlike agglutinative languages, a polysynthetic language will often have more than one root. Compound words, you may recall, also have more than one root. However, unlike compound words, words in polysynthetic languages contain inflectional morphemes from more than one part of speech in a single word. Another potential difference between polysynthesis and compounding is that polysynthesis is productive, while compounding varies in productivity across languages.

Synthesis, fusion, and exponence

The four morphological types above really conflate three properties: **synthesis**, **fusion**, and **exponence**. Not all languages fit perfectly into one of the four traditional categories, so it is sometimes useful to consider each of these properties independently. Some languages may even differ depending on the kind of morpheme. For example, Chapter 21 of the *World Atlas of Language Structures* (WALS) surveys how different languages package both noun and verb inflection. If you cross-reference the two maps associated with Chapter 21, you'll find that some languages, including Chukchi, Finnish, German, West Greelandic, Nenets, Russian, and Yaqui, combine case with number in a single morpheme, similar to the Latin examples in Table 2, but do not combine their tense markers with another category. Conversely, you'll find that other languages, including Aymara, Brahui, Georgian, Hindi, Spanish, and Tibetan combine their tense markers with agreement in a single morpheme but do not combine case with another category (Bickel and Nickels 2013b). It may be enough to classify both of these sets of languages as fusional in some situations, but in other situations it may be useful to make more nuanced distinctions.

The four traditional categories are discrete categories, meaning a language either fits into the category or does not. The three categories of synthesis, fusion, and exponence, on the other hand, are scales, which means that a language might fall on the high end of the scale, on the low end of the scale, or somewhere in the middle.

Synthesis

Synthesis is the measure of how many morphemes may combine into a single word. A language with a high

degree of synthesis will have many morphemes contained into a single word, while a language with a low degree of synthesis will have few morphemes combined into a single word. Languages with a low degree of synthesis are often called **analytic** languages.

Fusion

Fusion is the measure of how phonologically separable morphemes are from their hosts. In a language with a low degree of fusion, it is easy to identify the boundaries between different morphemes. Isolating languages can be classified as low fusion languages because most of the morphemes are separated by a word boundary, which is a very clear phonological boundary. Agglutinative languages also have a low degree of fusion. On the other hand, a language with a high degree of fusion will have morpheme boundaries that are less clear. English exhibits a higher degree of fusion in some irregular verb inflections, such as *catch ~ caught*, *fall ~ fell*, and *bend ~ bent*. In these forms, there is no part of the word that we can clearly identify as either a present tense or past tense morpheme. Semitic languages such as Hebrew and Arabic are well-known for using vowel changes as one of their main morphological processes (instead of just in irregular verbs like English), and so would be classified as having an even higher degree of fusion than English.

Note that this meaning of fusion is very different from the category *fusional* discussed above, which is more closely related to the concept of exponence.

Exponence

Exponence is the measure of how many meanings can be encoded into a single morpheme. Wari', a Chapacura-Wanham language spoken in Brazil, has high exponence on the word *na*, which encodes the person and number of the subject, as well as several categories of verb inflection (realis, non-future, and active), for a total of five grammatical categories expressed in a single morpheme. When multiple meanings are encoded in a single morpheme in this way, it is known as **cumulative exponence**.

(3)	Toc	na	com.
	drink.SG	3SG.REAL.NONFUT.ACTIVE	water
	'He is drinking water.'		

(Wari'; Everett and Kern 1997: 339, as cited in Bickel and Nickels 2013b)

Low exponence, on the other hand, will only have one meaning encoded in each morpheme.

Redefining the traditional categories

Let us now revisit the four traditional morphological categories and redefine them in terms of their synthesis, fusion, and exponence.

Isolating languages have one or close to one morpheme per word, which means they have low synthesis. The morphemes, belonging as they do to independent words, are phonologically separable, and therefore isolating languages have low fusion as well. Finally, most morphemes in isolating languages encode only one meaning, and so they also have low exponence.

Agglutinative languages have many morphemes per word, and therefore have high synthesis. However, the morphemes are still phonologically separable and encode one meaning per morpheme; in this way, agglutinative languages have low fusion and low exponence.

Fusional languages are characterized by having many meanings encoded into a single morpheme, which means they have high exponence. The definition of fusional languages doesn't specify their degree of synthesis or fusion, but fusional languages will tend to have fewer morphemes per word than agglutinative languages since each morpheme is carrying more information.

Finally, polysynthetic languages are highly synthetic. The definition of polysynthesis does not specify the degree of fusion or exponence, but polysynthetic languages tend to have a small degree of fusion, with phonological processes crossing morpheme boundaries, and medium amounts of exponence.

Table 3. Defining the four traditional morphological categories in terms of synthesis, fusion, and exponence

	Synthesis	**Fusion**	**Exponence**
Isolating languages	low	low	low
Agglutinative languages	high	low	low
Fusional languages	medium	unspecified	high
Polysynthetic languages	very high	unspecified	unspecified

Using *The World Atlas of Language Structures* (WALS)

You can explore how these morphological properties show up in the different languages of the

world on *The World Atlas of Language Structures* (WALS). In WALS, Chapter 20 describes fusion, Chapter 21 describes exponence, and Chapter 22 describes synthesis. Each article is associated with one or more maps showing how the different properties are distributed across a sample of the world's languages. You can even combine the data from two or more maps to see how the different properties interact.

WALS is put together from a descriptive functionalist framework. As such, the terminology in some chapters might differ from what we learn in class, but you can often figure it out if you read the explanations in the chapters.

However, as you explore WALS, you should keep in mind that it has a couple of weaknesses. First, the authors of each chapter on WALS are experts in the topic of their chapters, but not on all of the languages in the database (which would be impossible!). Often, then, the authors have classified languages based on a surface understanding of the language from referencing grammars and other published works. Second, mostly only spoken languages are represented in the database, with the exception of Chapters 139 and 140, which are dedicated to signed languages.

Key takeaways

- Language differ in how many meanings are put in a single morpheme and how many morphemes are put in a single word.
- Isolating languages have a near one-to-one morpheme-to-word ratio, with each morpheme containing one meaning.
- Agglutinative languages have many morphemes in each word, but each morpheme still only contains one meaning.
- Fusional languages put multiple meanings into some morphemes and multiple morphemes into each word.
- Polysynthetic languages have many morphemes in a single word, often the equivalent of a sentence in other languages. There may be multiple roots in a single word, as well as inflectional affixes from more than one part of speech.
- These four morphological types conflate three properties. Synthesis is the measure of how many morphemes may combine in a single word. Fusion is the property of how

phonologically separable morphemes are from their hosts. Exponence is the property of how many meanings can be encoded into a single morpheme.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1623#h5p-34

References and further resources

Comedy and satirical linguistics

🎉 Phlogiston, Phineas Q. 2007. Cartoon theories of linguistics part 3: Morphological typology. *Speculative Grammarian* CLII (3). https://specgram.com/CLII.3/09.phlogiston.cartoon.3.html.

Sources for examples

Compton, Richard, and Christine Pittman. 2010. Word-formation by phase in Inuit. *Lingua* 120: 2167–2192.

Lieber, Rochelle. 2016. *Introducing Morphology*. 2nd edition. Cambridge: Cambridge University Press.

Academic sources

🔍 Bickel, Balthasar, and Johanna Nichols. 2013a. Fusion of Selected Inflectional Formatives. In WALS Online (v2020.3) [Data set], ed. Matthew S. Dryer and Martin Haspelmath. Zenodo. https://wals.info/chapter/20.

🔍 Bickel, Balthasar, and Johanna Nichols. 2013b. Exponence of Selected Inflectional Formatives. In WALS Online (v2020.3) [Data set], ed. Matthew S. Dryer and Martin Haspelmath. Zenodo. https://wals.info/chapter/21.

🔍 Bickel, Balthasar, and Johanna Nichols. 2013c. Inflectional Synthesis of the Verb. In WALS Online (v2020.3) [Data set], ed. Matthew S. Dryer and Martin Haspelmath. Zenodo. https://wals.info/chapter/22.

Mattissen, Johanna. 2004. A structural typology of polysynthesis. *Word* 55 (2).

9.4. WORDHOOD: AN UNSOLVED PROBLEM

In Section 6.2, we learned about how the word *word* can be used in many different ways. In addition, we defined many technical terms we can use to distinguish these different uses. But we still haven't defined what a word is. There is a reason for that! The definition of *word* is an unsolved problem in linguistics. There is no definition of *word* that works for all known languages.

Some attempts at defining *word*

In this section, we will consider many attempts at definitions of *word* and discuss how they fall short. Before you read this section, though, take a moment and attempt to write a definition of *word*. As you read through these possible definitions, compare them to the one you wrote. Do any of the problems discussed here also apply to your definition?

Based on spelling

We could attempt to define words based on how they appear in writing. One possible definition, then could be as in (1).

(1) Definition attempt 1: A string of letters written with a space on each end.

The first problem with defining *word* based on **orthography** is that not all languages are written down. Definition (1), then, would mean that languages without a writing system would have no words at all, which does not match speaker intuition. Even languages with a writing system don't all indicate word boundaries. Classical Chinese, Ancient Greek, and Classical Latin were all written without spaces between words.

In languages that have a writing system that uses spaces, the use of spaces doesn't always correspond exactly to word units. For example, English compound words can be written as one word, with a hyphen, or with a space, as illustrated in (2).

(2) a. Compound words with spaces: *hot dog, high school, common sense, dining room, first aid, peanut butter, post office, prime minister, search engine, remote control, washing machine, role model*

b. Hyphenated compound words: *sister-in-law, check-in, far-fetched, free-for-all, know-how, merry-go-round, one-sided, well-being, up-to-date, self-esteem, freeze-dried, cage-free*

c. Compound words without spaces: *greenhouse, airport, bathroom, basketball, daylight, desktop, firefly, grandmother, grapefruit, grasshopper, lipstick, mailbox*

Whether a compound word has a space, a hyphen, or is written as one word is just convention, and it can

change over time. For example, in older texts, you may see *ice-cream* written with a hyphen, whereas in modern texts it tends to be written with a space instead, as *ice cream*. However, there is evidence, as we will discuss in Chapter 10 on compounding, that compound words behave as a single word, regardless of how they are written.

Based on uninterruptibility

We could attempt to define words based on their interruptibility. Perhaps a word is a unit that cannot have anything inserted into its middle.

(3) Definition attempt 2: A unit that cannot have anything added to it, except at the edges.

One of the big problems with this definition is that it doesn't account for **infixes**. For example, the Lakhota first singular subject marker *-wa-* is an infix.[1] This is shown in (4), where the roots *máni*, *aphé*, and *hoxpé* have *-wa-* inserted in between the first and second syllable to form the first-person singular.

(4)	a.	máni	'he walks'	ma-wá-ni	'I walk'
	b.	aphé	'he hits'	a-wá-phe	'I hit'
	c.	hoxpé	'he coughs'	ho-wá-xpe	'I cough'

(Lakhota; Albright 2000: 2)

If we maintain the the definition in (3), that a word is a unit that can only be added to at its edge, data like (4) would force us to conclude that the roots *máni*, *aphé*, and *hoxpé* each consist of two words. This is problematic, though, because the roots only contain one morpheme—that is, one piece of meaning! How could one morpheme be spread across two words?

Another challenge to the idea that a word is not interruptible are phrasal verbs. Phrasal verbs, such as *turn on*, *blow up, break down,* and *call off* are a subcategory of compound words that usually consist of a verb and a preposition. Phrasal verbs can have past tense markers inserted in between their two components, as in *called off*. Phrasal verbs are also well-known for allowing the object to occur in between the verb and the preposition, as shown in (5) to (8).

1. The Lakhota first person singular subject marker alternates between being a prefix and an infix, depending on the stem word.

(5)	a.	**turn on** the lights
	b.	**turn** the lights **on**
(6)	a.	**blow up** the building
	b.	**blow** the building **up**
(7)	a.	**break down** the box
	b.	**break** the box **down**
(8)	a.	**call off** your dog
	b.	**call** your dog **off**

English phrasal verbs, on their own, are perhaps not the strongest piece of evidence, since we could perhaps argue that they are two words, not one.

German has a similar pattern that is even more puzzling.[2] In German, verbs with separable prefixes are written with one word when there is an auxiliary, for example *aufstehen* in (9b) and *zurückgeben* in (10b). But in (9a) and (10a), when there is an auxiliary, the word is split into two pieces which aren't even next to each other! The verb portions of the particle verbs, *stehst* in (9a) and *gebe* in (10a), are the second word in each sentence, while the preposition portions of the particle verbs, *auf* in (9a) and *zurück* in (10a), are at the end of each sentence.

(9) *German*

a.	Wann	**stehst**	du	morgen	**auf**?
	when	stand.pres.2sg	you	tomorrow	up
	'When will you get up tomorrow?'				

b.	Wann	willst	du	morgen	**aufstehen**?
	when	want.pres.2sg	you	tomorrow	stand.up
	'When do you want to get up tomorrow?'				

2. German data provided by Katharina Pabst.

(10) *German*

a.	Ich	**gebe**	meiner	Tante	ihr	Buch	**zurück**.
	I	give.pres.1sg	1sg.dat.f.poss	aunt	3sg.f.dat	book	give.back
	'I give my aunt her book back.'						

b.	Ich	muss	meiner	Tante	ihr	Buch	zurückgeben.
	I	must.pres.1sg	1sg.dat.f.poss	aunt	3sg.f.dat	book	**give.back**
	'I have to give my aunt her book back.'						

Based on independence

Perhaps we should define words, not based on whether we can interrupt them, but based on whether they can stand on their own.

(11) Definition attempt 3: A unit that can be pronounced in isolation.

The problem with this definition is, first of all, that some subcomponents of words can stand on their own. For example, clippings like *math*, *flu*, and *fridge* can stand on their own. Does that mean that *mathematics*, *influenza*, and *refrigerator* are not words, but phrases? Our intuitions (or mine at least) says that can't be right, because that would mean that the parts of *mathematics*, *influenza*, and *refrigerator* that don't show up in *math*, *flu*, and *fridge* are also words.

Secondly, some words, especially functional words, can't really stand on their own, such as *the* and *of*.

Based on the interface between semantics and phonology

Perhaps, then, we could define words based on their function as a form-meaning pair. In that case, a word would be a string of sounds or signs that, together, are associated with a meaning.

(12) Definition attempt 4: A form-meaning pair.

However, this definition does not quite work either. There are, of course, form-meaning pairs that are bigger than a word as well as form-meaning pairs that are smaller than a word.

It is easy to find form-meaning pairs that are smaller than a word. Any morpheme is a form-meaning pair, thus any multimorphemic word consists of form-meaning pairs that are smaller than a word. For example, *teapot* consists of two form-meaning pairs, *tea* and *pot*.

There are also form-meaning pairs that are bigger than a word, which are called **idioms**. Idioms are phrases with non-compositional meaning. In other words, it is the entire phrase that is paired with a meaning. Some examples of English idioms are listed in (13).

(13)	a.	*break a leg*	to wish someone luck before a performance
	b.	*beat around the bush*	explain or request something indirectly
	c.	*hit the sack*	go to bed
	d.	*on the ball*	prepared, ready
	e.	*raining cats and dogs*	raining very hard
	f.	*spill the beans*	tell a secret
	g.	*under the weather*	sick

Based on phonological domains

Perhaps we can define words based on the limits of phonological processes, such as stress assignment.

(14) Definition attempt 5: The domain in which phonological processes such as stress occur.

The first problem with this definition is that functional words such as *the* and *of* often do not receive stress at all, in which case they would not count as words.

The second problem is that the domain of phonological processes vary, both within the same language and across different languages. For example, the two roots in a compound word sometimes behave like they're in the same phonological domain, and sometimes they do not.

Let's consider vowel harmony in Finnish, which causes all vowels in the same word to match in backness. In (13), the adessive marker appears as *-llä* if the vowels in the stem are front vowels, as in (13a), and as *-lla* if the vowels in the stem are back vowels, as in (13b).

(15)	a.	pöydä-llä
		table-ADESS
		'on the table'
	b.	kadu-lla
		street-ADESS
		'in the street'

(Finnish; Julien 2002: 24)

The same pattern does not hold in compound words, as shown in (16). The first root, *pää* 'head', has front vowels, while the second root, *kaupunki* 'city', has back vowels. Vowel harmony does not cross the boundary between the two compound words.

(16)	pää-kaupunki	
	head-city	
	'capital'	

(Finnish; Julien 2002: 24)

In contrast, compound words in Greek do behave as a single domain for stress assignment, as shown in (17). In this compound word, there is only one stress, marked by the accent on the final vowel.

(17)	ksilo-θimonyá	
	wood-stack	
	'wood stack'	

(Greek; Julien 2002: 17)

If this compound word behaved like two separate words, we would expect there to be a stress on each root. Instead, there is only stress on the head of the compound word.

Based on speaker intuition

Let's attempt one last definition of *word*. Perhaps we can consider words to be a unit according to the intuition of the language users.

(18) Definition attempt 6: A unit that stands on its own according to the intuition of language users.

The difficulty with this definition is that literate language users don't normally have intuitions about wordhood that are separate from their writing system. Illiterate language users are becoming more difficult to find, due to global advances in education. The rules from the writing system might reflect underlying grammatical principles, but they also might not.

What do we do now?

We do not have a consistent way of defining words across all contexts and languages. This could mean lots of different things. It could mean that we just haven't identified the right definition yet. It could mean that there is no universal definition of *word*, but that it is defined language-by-language. It could also mean that the word *word* refers to more than one thing, and using the same word in all of these different contexts has confused us! Finally, it could mean that there really is no such thing as words in the grammar at all. Let's look at these last two ideas more closely.

Different kinds of *word*

One possible solution is to claim that there are different domains that are sometimes called *words*: phonological words, grammatical words, and orthographic words. Sometimes these different domains coincide, and refer to the same string, but sometimes they don't. The different wordhood domains could be defined as follows:

- Phonological words are a unit in the prosodic structure of a sentence. They are the domains in which word-level phonological processes occur, including stress assignment. Phonological words are the smallest unit that can stand on their own.
- Grammatical words are a unit in morphosyntactic structure. They correspond to the terminal nodes in syntax trees. In other words, they are the units that syntax can manipulate. This definition is probably the least satisfying of these three, since we sometimes put morphemes that cannot stand on their own in terminal nodes of syntactic trees.
- Orthographic words are a unit we use in writing, based on spelling convention.

In other cases, linguists divide the notion of *word* into lexeme, word token, or word form, as discussed in Section 8.2.

Maybe words aren't real

Another possible solution to this problem is to assume that words aren't actually a real component in our model of human language. Julien (2002) is one linguist who hypothesizes this:

> ...my working hypothesis in the following will be that 'word' in the nonphonological sense is a distributional concept. That is, if a given string of morphemes is regarded as a word, it simply means that the morphemes in question regularly appear adjacent to each other and in a certain order. The reason the morphemes show such behavior is to be found in their syntax. But notably, the structural relation between the morphemes is not directly relevant for the word status of the string; it only matters insofar as some structural arrangements of morphemes may result in independent distribution and internal cohesion, whereas others may not.
>
> Crucially, if wordhood cannot be associated with any particular structural morpheme configuration, it follows that grammar cannot have at its disposal any specific word-forming devices. If a word is just the accidental outcome of the manipulation of morphemes that takes

> place in syntax, it must be the case that words come into being in our perception; that is, words are *perceived* rather than *formed*.
>
> (Julien 2002: 36)

Basically, Julien is arguing that words are just strings of morphemes that appear together frequently enough that we treat them in a special way, rather than from some special configuration in syntactic structure.

Key takeaways

- There is no way to define *word* that works for all languages because orthographic, phonological, morphological, syntactic, and semantic boundaries don't all align in the same way across words and across languages.
- One possible explanation for this is that phonological words, grammatical words, and orthographic words are three different kinds of units that only sometimes coincide.
- Another possible explanation is that words are strings of morphemes that appear together frequently but that don't correspond to any particular structure.

Check yourself!

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=1625#h5p-35

References and further resources

Academic sources

Albright, Adam. 2000. The productivity of infixation in Lakhota. *UCLA Working Papers in Linguistics* 0: 1–19.

Julien, Marit. 2002. *Syntactic heads and word formation*. Oxford: Oxford University Press.

CHAPTER 9 PRACTICE EXERCISES

Data analysis

Exercise 1. Dictionary search

Beginner [9.2]

Look at the dictionary entry for the word meaning 'canoe' in this Ojibwe dictionary.

a. What is the citation form for this word?

b. Which word forms does it list?

c. Identify all of the word tokens of the word for canoe on the website. List two examples here.

d. How many lexemes do you think are represented by this one dictionary entry? Explain your reasoning.

Exercise 2. Linguistic analysis

Intermediate [9.3]

Consider the following data from four different languages. Based on the data provided here, classify each language either as isolating, agglutinative, fusional, or polysynthetic. Explain the reasoning behind your classification choice.

Abbreviations used in these examples include 1, 2, 3 = 1st, 2nd, 3rd person, ACC = accusative (object pronoun), CPL = completive, DAT = dative (indirect object pronoun), DET = determiner, FUT = future, NOM = nominative (subject pronoun), PL = plural, PROG = progressive (action in progress), SG = singular

(1)	q'ai-hwagait-tsagam-yu'k-det	
	still-CPL-toward_shore-move-3PL	
	'they moved still far away toward the shore'	

(Tsimshian; Boas 1911: 325, as cited in Mattissen 2004: 191)

(2)	Le	are	k-u-k'yek	le	t'oyt'ik.
	DET	3SG	PROG-3SG-kick	DET	ball
	'S/he is/was kicking the ball.'				

(K'iche'; data from Elsa Sacalshot)

(3)	a.	bɨ	vuch-a-th	tsɨ
		1SG.NOM	see-1SG-2SG.ACC	2SG.ACC
		'I will see you.'		

(Kashmiri; Wali & Koul 1997: 228, as cited in Manetta 2014: 246)

	b.	su	vuch-iy	(tse)
		3SG.NOM	see-2SG.DAT	2SG.DAT
		'He will see you.'		

	c.	su	vuch-i	təmis
		3SG.NOM	see-FUT	3SG.DAT
		'He will see him.'		

(Kashmiri; Manetta 2014: 246)

(4)	Kua	kitia	e	au	a	Moka.
	PERF	see	ERG	1SG	ABS	Moka
	'I saw Moka.'					

(Niuean; Massam 2020: 4)

Communication and study skills

Exercise 3. Wordhood

Advanced [9.4]

Polysynthetic languages cause particular challenges for the concept of wordhood. Choose one of the attempted definitions of *word* from Section 9.4 and discuss whether that definition could handle polysynthetic languages.

Exercise 4. Dictionary entry production

Intermediate [9.2]

Think of a slang word or a word that has only recently entered English or another language you speak. Construct a dictionary entry for the word. Include the citation form, inflectional forms (if any), part of speech, pronunication, a definition, and examples of the word in use.

Hint: Use a published (paper or online) dictionary as a model for how to format the dictionary entry.

Exercise 5. Word games

Advanced [9.2, 9.4]

Look up the rules of a word game and answer the following questions. You can use the rules from the game *Codenames* copied below or choose another word game.

a. Think about the rules of the game. Which uses of *word* are relevant to the game mechanics (for example, *lexeme*, *word form*, *word token*, etc...)? Does the game need to limit productive formation of new words, or would creative formation of new words make the game more fun?

b. If the game defines what is or is not a word for the purposes of the game, does the game's definition of a word have any problems? Can you think of anything that you would consider a word that is excluded by the game's definition, or anything that you would not consider a word that is included? If you were involved in the game design, would you recommend any changes to how *word* is defined in the instructions?

c. If the game does not define what is a word, how would you define a word for the purposes of this game? Remember to use non-technical terminology, so that players will be able to understand the instructions. Why did you choose the definition you chose?

In the game *Codenames*, one player gives clues to the rest of their team to help them pick out their team's secret words out of a set of 25 words. The catch is that the clue can only be one word long! Because of this, the game instructions spend some time defining what counts as a single word. Here is an excerpt.

Sometimes you have to make judgment calls about what is valid and what is not. Different groups may prefer to play the game differently.

Compound words

English has three ways to write a compound word. *Greenhouse* is one word. *Pack rat* is two words. *Mother-in-law* is hyphenated. Technically, only *greenhouse* can be a one-word clue.

You can decide to allow any compound words. However, in no case should a player be allowed to invent compound words. *Lunar squid* is not a valid clue for MOON and OCTOPUS.

Proper names

Proper names are always valid clues if they follow the other rules. *George* is a valid clue, but you might want to specify whether you mean *George Washington* or *George W. Bush*. Your group can agree to count proper names as one word. This would also allow titles such as *The Three Musketeers*. Even if you don't allow multi-word proper names, you might want to make an exception for place names like *New York*.

Spymasters should not be allowed to make up names, not even names that turn out to be real. *Sue Mee* is not a valid clue for CHINA and LAWYER.

Acronyms and abbreviations

Technically, CIA is not one word. But it is a great clue. You can decide to allow common abbreviations like *UK*, *lol*, and *PhD*. And words like *laser*, *radar*, and *sonar* are always allowed, even thought they originated as acronyms.

Homonyms

Some people prefer to allow a more liberal use of homonyms. You can allow *knight* to be a clue for *night*-related things if that makes the game more fun for you.

(Chvátil 2015: 7)

Research and application

Exercise 6. Etymology

Beginner-Intermediate [9.1]

Look up the following words in an etymological dictionary. What is their word formation process? How can you tell? Be sure to use the technical terms for different formation processes discussed in Section 9.1.

Note: Sometimes the dictionary entry will tell you directly. Other times, you will have to read the etymology and decide for yourselves based on the description of the word's history.

a. Scuba (n)
b. Brainwash (n)
c. Camcorder (n)
d. Traceable (adj)
e. Bandersnatch (n)
f. Party (v)
g. DIY (adj/n)
h. Gym (n)
i. Boomerang (n)
j. Scavenge (v)
k. Exam (n)
l. Okay (adj/adv)
m. Wabe (n)
n. Keyboard (n)
o. Labradoodle (n)
p. Yuppie (n)

Exercise 7. WALS

Advanced [9.3]

Choose one language you don't know much about and look it up in WALS chapters 20, 21, and 22.

a. Provide a brief introductory statement about your language. (e.g., What is its name, what language family is it from, and where is it spoken?)
b. Describe your language in terms of fusion, exponence, and synthesis.

c. Based on what you found, does your language fall neatly into one of the traditional four categories? Why or why not?

Language journal

At the end of Chapter 1, we introduced the Language Journal. You should have picked a language to build a profile for as you work through this textbook. You can add to your profile with these questions.

Exercise 8

Beginner [9.2]

Find a dictionary for your language. Pick a dictionary entry, and identify the lexeme, citation form, word forms, and word tokens in that entry.

Exercise 9

Intermediate [9.1]

Is there an etymological dictionary available for your language? If so, find and describe the etymology of three different words in your language with three different word formation processes.

Exercise 10

Advanced [9.3]

Describe the synthesis, fusion, and exponence of your language. Does your language fit well into one of the categories of isolating, agglutinative, fusional, or polysynthetic? Explain your reasoning.

References and further resources

Sources for examples

Chvátil, Vlaada. 2015. *Codenames*. Czech Games Edition.

Manetta, Emily. 2014. Markedness and syncretism in Kashmiri differential argument encoding. In Pritha Chandra and Richa Srishti (eds.), *The Lexicon-syntax interface: Perspectives from South Asian languages*. Amsterdam: John Benjamins. 245–270.

Massam, Diane. 2020. *Niuean*. Oxford: Oxford University Press.

Mattissen, Johanna. 2004. A structural typology of polysynthesis. *Word* 55 (2).

8. SEGMENTING WORDS INTO MORPHEMES

[Placeholder.]

10. COMPOUNDING

[Placeholder.]

9. INFLECTION VS. DERIVATION

[Placeholder.]

11. HIERARCHICAL STRUCTURE IN WORDS

[Placeholder.]

11.1 HIERARCHICAL STRUCTURE IN COMPOUND WORDS

As we saw in the previous chapter, some languages, including English, have a productive process through which new words can be formed that involves putting two words together. The resulting words are called **compounds**.

Table 1. Compounding in English

rice	pot	→	rice-pot
swan	boat	→	swan-boat
phone	case	→	phone-case

This process is, theoretically, infinitely **recursive**, meaning that we can continue to make new words from existing words.

Table 2. More compounding in English

rice-pot	rack	→	rice-pot-rack
swan-boat	jacket	→	swan-boat-jacket
phone-case	store	→	phone-case-store

Rice-pot has two morphemes, *rice* and *pot*. When we put them together, we make something whose meaning is the combination of both. This idea is represented in terms of **bracketing structure**.

(1) a. *rice-pot* : [[rice] [pot]]
b. *swan-boat* : [[swan] [boat]]

Inside each pair of brackets is one meaning "unit." So *rice-pot* involves three pairs of brackets: one around *rice*, one around *pot* and one around the compound *rice-pot*. Same with *swan-boat*.

This way of representation allows us to iteratively represent more complex compounds, as in (2).

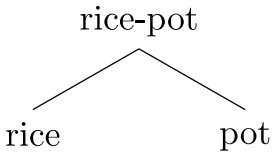
(2) a. *rice-pot-rack* : [[[rice] [pot]] [rack]]
b. *swan-boat-jacket* : [[[swan] [boat]] [jacket]]

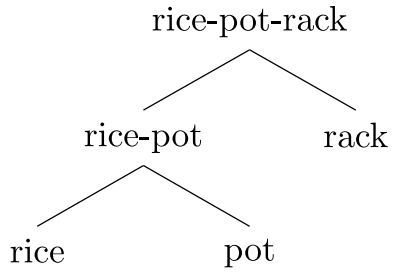
When we build larger compounds, we are creating a meaningful unit, and then adding to it. So first I create

the word *rice-pot*, and it has a meaning like, "a pot for rice." Then I create the word *rice-pot-rack*, which has the meaning "a rack for pots for rice." In other words, the meaning of the entire word depends on the meaning of its parts.

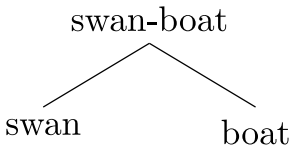
Of course, the more complex our compounds become, the more difficult it is to read the bracketing structure. So another way to express the exact same information is by using a tree, as in (3) and (4).

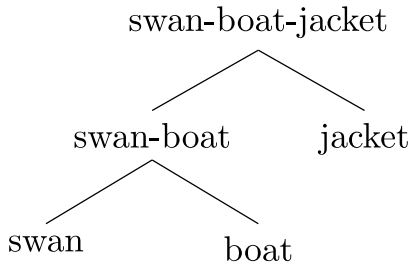
(3)





(4)

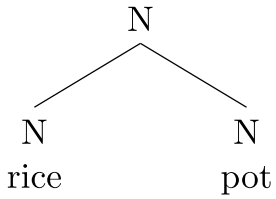


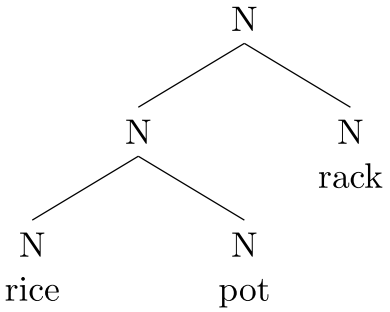


Each **node** in the tree corresponds to one pair of brackets. Thus, trees and brackets provide the *exact* same amount of information, it's just that trees do it in a visually more appealing way. But it's always possible to state a bracketing structure as a tree, and vice versa.

I can actually make my representations slightly more informative by adding in *category* information. The following trees and bracketing structures give us all the preceding information, and they additionally tell us what the category of each morpheme is.

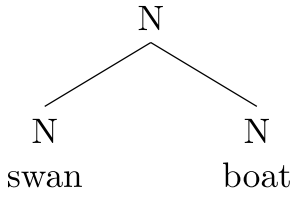
(5) a.

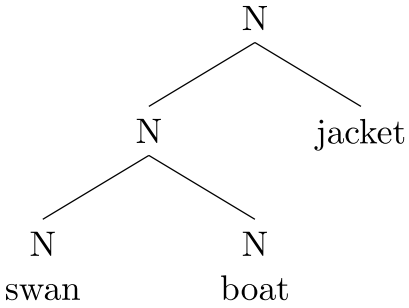




b. [$_N$ [$_N$ rice] [$_N$ pot]] [$_N$ [$_N$ [$_N$ rice] [$_N$ pot]] [$_N$ rack]]

(6) a.

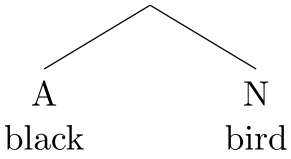




b. [$_N$ [$_N$ swan] [$_N$ boat]] [$_N$ [$_N$ [$_N$ swan] [$_N$ boat]] [$_N$ jacket]]

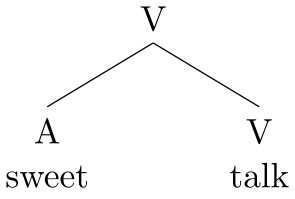
Note that it's not necessary to explicitly state the compound words that are formed at each junction, because I can just look lower in the tree to figure out what that word is. The reason it's good to list the category, though, is that you can combine more than one category. We therefore want to know which category "projects," that is, provides the part of speech to the resulting word.

(7) *black-bird* :



[$_N$ [$_A$ black] [$_N$ bird]]

(8) *sweet-talk* :



[$_V$ [$_A$ sweet] [$_V$ talk]]

Hierarchical structure

These tree and bracketed diagrams illustrate one property of language: words and sentences have **hierarchical structure**. A hierarchical structure is a way of organizing elements. You might encounter hierachical structure in a workplace, where, for example, a worker might report to their team leader, the team leader might report to a department head, and the department head might report to the CEO.

In morphology trees, though, instead of reporting to people higher up the chain of command, hierarchical

structure represents containment relationships. The nodes higher up in the structure contain the nodes that branch off below them.

Structural ambiguity

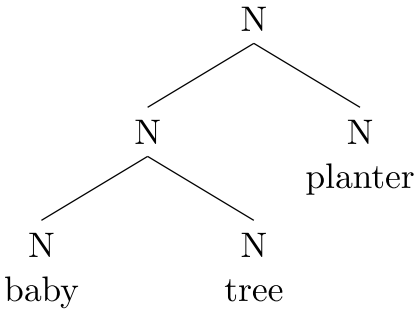
What does the following compound mean?

(9) baby-tree-planter

Many of you will say that it means, "someone who plants baby trees." And many of you will say that it means "a tree-planter who is a baby." Both of these are right; it's possible to get both readings of the compound.

If we wanted to represent each meaning, we would choose different structures. The "baby-trees" meaning would have the following representations.

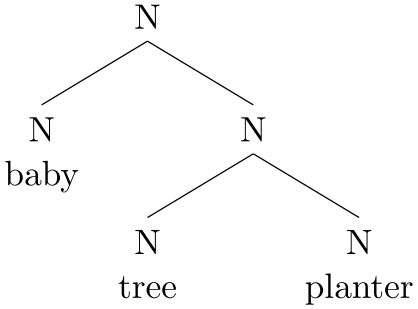
(10) "A planter of baby trees" :



The reason is that we want "baby tree" to be a unit of meaning, i.e., a **constituent**, because that meaning describes the kind of planter it is.

On the other hand, the "tree-planting-baby" meaning would have the following representation

(11) "A tree-planter who is a baby" :



The reason is that we want "tree-planter" to be a unit of meaning first, because ultimately we're describing a kind of tree-planter.

This kind of ambiguity is called **structural ambiguity**, because we can represent the ambiguity structurally.

There is another kind of ambiguity called **lexical ambiguity**. The word *bank* is lexically ambiguous, because it means two different things (a financial institution and the side of a river). But it's not structurally ambiguous because the two meanings do not correspond to different structures.

The take-away point here is that **different meanings correspond to different structures**. More abstractly, the different meanings are the result of packaging the information in different ways.

Check yourself!



An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.openedmb.ca/wordandsentencestructures/?p=2106#h5p-36

References and further resources

Attribution

This section is adapted from the following CC BY NC SA source:

Gluckman, John. n.d. Chapter 3: Brackets and trees. *The science of syntax*. https://pressbooks.pub/syntax/chapter/brackets-and-trees

12. MORPHOPHONOLOGY

[Placeholder.]

13. PRODUCTIVITY AND WORD FORMATION

[Placeholder.]

14. THE LEXICON

[Placeholder.]

15. CLITICS

[Placeholder.]

PART III. SYNTAX

Chapters 16-24 focus on syntax.

16. CONSTITUENCY AND HIERARCHY

[Placeholder.]

17. PHRASE STRUCTURE RULE GRAMMARS

[Placeholder.]

18. SYNTAX TREES

[Placeholder.]

19. BINDING THEORY

[Placeholder.]

20. PREDICATES AND ARGUMENTS

[Placeholder.]

21. TYPES OF CLAUSES

[Placeholder.]

22. INTRO TO X-BAR

[Placeholder.]

23. INTRO TO MOVEMENT

[Placeholder.]

24. DIFFERENT WAYS OF DOING SYNTAX

[Placeholder.]

PART IV. MORPHOLOGY-SYNTAX INTERFACE

[Placeholder.]

25. Y MODEL OF GRAMMAR

[Placeholder.]

26. MORPHOSYNTACTIC FEATURES, SELECTION, AND AGREEMENT

[Placeholder.]

27. LANGUAGE UNIVERSALS AND DIVERSITY

[Placeholder.]

28. DISTRIBUTED MORPHOLOGY

[Placeholder.]

GLOSSARY

ablaut

synonym: internal change
A morphological process that changes the vowel in the stem.

acquisition

Developing a skill, habit, or ability subconsciously, by instinct, instead of consciously.

acronym

When the first letter or two of each word in a phrase is combined to form a new word, which is pronounced as a whole.

active voice

A clause where there is an agent in subject position. The opposite of the passive voice.

adjunct

An optional modifying phrase. The opposite of an argument.

adposition

A preposition or a postposition. A word describing the relationship between a noun and another part of the phrase.

affix

A bound morpheme that attaches to a stem.

affixes

bound morphemes that are attached to a stem, usually linearly.

agglutinative language

A language with one meaning per morpheme and multiple morphemes per word.

algorithm

A procedure, process, or system of rules used to solve a problem.

allomorph

A variant form of a morpheme.

allomorphy

When the form of a morpheme varies depending on context.

alphabetism

synonym: initialism
When the first letter or two of each word in a phrase is combined to form a new word, which is pronounced as a string of letters.

analogy

A language change process whereby one word form adopts a pattern used by another word form.

analytic

A language with a low degree of synthesis. In other words, a language with few morphemes per word.

anthropological linguistics

The study of the relationship between language and culture.

argument

An argument is something that is selected for and required by the head of a phrase. Arguments include grammatical roles like "subject, object, indirect object, etc."

aspect

Verbal inflectional that indicates the internal time structure of an event, such as indicating that an action is complete, ongoing, repeated, etc.

attested

A word or construction that exists in a particular corpus, or, more generally, that has been observed in use in a natural setting.

backformation

A process whereby a language user reanalyzes a monomorphemic word to be polymorphemic and removes the additional morphemes.

backronym

When a word that already exists is claimed to be an acronym.

bare form

A word without any inflectional morphemes applied to it, or the simplest form in a paradigm.

base

The portion of a word that hosts an affix.

blend

A word formed by combining portions of two or more roots. At least one of the roots must be truncated in order to be a blend.

borrowing

When a word from one language is adopted by another language.

bound morpheme

A morpheme that must attach to a stem.

bracketing structure

A representation using brackets to indicate constituents and categories in morphology or syntax.

calque

A borrowed phrase which is a direct translation of a phrase in another language.

case

A morphology property of nouns in many languages. Case can either be structural, and indicate the grammatical role (subject/object) or semantic, and indicate the thematic role of the noun.

categorical

A pattern or rule with no exceptions.

Chomskyan

Any theories or models that are based on the work of Noam Chomsky.

circumfix

An affix that is inserted on both sides of the stem.

citation

The attribution of quotes and information to their original sources.

citation form

The word form used to refer to a lexeme.

clipping

The formation of a new word by shortening an existing one.

closed parts of speech

Parts of speech that cannot easily have new words added to them.

cognitive science

The study of the mind.

common noun

A noun that does not refer to a specific entity by name, but rather refers to a type of entity.

comparative

An inflectional form of adjectives meaning "to a greater degree of X."

competence

Having acquired the rules of grammar in a particular language, regardless of the ability to produce grammatical language in a particular instance.

complementary distribution

When the distributions of two elements are opposites. Each element always appears in contexts the other element can never appear in.

complex word

A word containing two or more morphemes.

compositionality

The observation that the meaning of a sentence is derived by the meaning of its parts and the way that they are combined.

compound word

A word with two or more roots.

computational linguistics

The study of how we can model and process language with computers.

concatenative morphology

Morphological processes that result in a linear string of morphemes, such as prefixation or suffixation.

confirmation bias

The tendency for humans to look for evidence that is consistent with their belief system.

constituent

A group of words or morphemes that behave as a unit.

constructed language

A language designed for a particular purpose, as opposed to a natural language.

conversion

Changing the part of speech of a word without changing its form.

copula

A word used to introduce non-verbal predicates, including adjectival *(I am happy*), nominal (*I am a teacher),* and prepositional (*I am down the street*) predicates. It has roughly the meaning of an equal sign. In English, the verb BE is used to form copular constructions.

corpus

plural: corpora
A body of language data collected from real-world use.

crip linguistics

A critical approach to the study of linguistics from the approach of disability studies, studying how people with differing abilities communicate, without taking a deficit approach.

cumulative exponence

When multiple meanings are encoded in a single morpheme.

derivation

A morphological process that creates a new lexeme.

descriptive grammar

A set of rules that describe how language is used based on observation.

descriptively adequate

A model that accounts for all observed data, plus also speaker judgments.

diachronic

Describing how something (such as language) has changed over time.

direct object

An argument of the verb, the entity which undergoes the action.

discourse analysis

The study of how multiple sentences are combined and interact, including turn-taking in conversation.

discourse markers

Words or morphemes used to manage the structure and flow of discourse.

Distributed Morphology

A model of grammar developed in the early 1990s with three main hypotheses: (a) Late Insertion, the idea that syntax is constructed using abstract items, and vocabulary items are inserted late in the process; (b) Underspecification, the idea that vocabulary items are inserted based on being the best match to the features, and do not have to be a perfect match; and (c) Syntactic Hierarchical Structure All the Way Down, the idea that morphological rules aren't separate from syntax, but that instead, syntactic rules are used to construct both words and sentences.

ditransitive

Having three arguments, usually a subject, direct object, and indirect object.

double object construction

A ditransitive construction where both objects are NPs and the indirect object comes first, as in *I gave you a book*.

E-language

External language.
The forms of language that are produced by individuals or in a community.

embedded clause

A clause that is included inside of a complex sentence, as a modifier or dependent on the main clause.

etymologists

People who study the historical development of words.

etymology

The study of the historical origins of words and morphemes, including change over time in their form and/or meaning.

explanatorily adequate

A model that accounts for all observed data and speaker judgments, as well as provides an explanation for how language is acquired.

exponence

A scale indexing how many meanings can be combined into a single morpheme.

falsifiable hypothesis

A hypothesis which is possible to be proven wrong by some test.

folk etymology

A commonly believed etymology of a word that is not historically accurate.

free morpheme

A morpheme that can stand on its own.

functional parts of speech

Parts of speech that encode grammatical functions, such as determiners, complementizers, and conjunctions.

fusion

A scale indexing the degree to which morphemes are phonologically separable from their stems.

fusional language

A language which tends to combine multiple meanings into a single morpheme.

generative grammar

Any model of grammar that uses rules to generate or "build" a language structure. This model should be able to produce all of the grammatical sentences in the language and no others.

genericization

When a brand name becomes the common name for all similar items.

genitive

Case marking possession.

gloss

Provide a morpheme-by-morpheme translation of linguistic data

head

A head is the main word in the phrase which determines the properties of the phrase as a whole. In syntax, it is typically a terminal node, and is realized as a morpheme or word.

hierarchical structure

The organization of elements into ranks or levels, where each level contains or is in charge of the lower levels.

historical linguistics

The study of how language changes over time.

homonym

When two or more words have the same form but different meaning.

hypercorrection

The over-application of a rule to additional contexts.

hypothesis

A proposed analysis or explanation made based on the observation of data, to be tested over the course of scientific investigation.

I-language

Internal language.
The system of grammatical rules that an individual language user has in their mind.

idiom

A phrase with non-compositional meaning.

indirect object

An argument of the verb, the goal or endpoint of an action.

infix

An affix that is inserted internally to the stem.

inflection

A morphological process that adds grammatical information, such as number, tense, gender, or case.

initialism

synonym: alphabetism
When the first letter or two of each word in a phrase is combined to form a new word, which is pronounced as a string of letters.

internal change

synonym: ablaut
A morphological process that changes the vowel in the stem.

International Phonetic Alphabet

An internationally recognized set of symbols developed in order to transcribe speech sounds with a one-to-one correspondence between sound and symbol.

intransitive

Not having an object.

Most commonly used to describe verbs that have only a single argument, the subject, although other parts of speech are also occasionally described as being transitive or intransitive.

isolating language

A language with more or less one morpheme per word.

language acquisition

The study of how children and adults acquire and learn language.

lexeme

a word in the abstract sense, including all of its different inflected forms

lexical ambiguity

When a word or sentence is ambiguous because one of the morphemes has a homonym.

lexical category

synonyms: part of speech, syntactic category
We classify words into part of speech based on their syntactic function. Some examples of part of speech include noun, verb, adjective, and preposition.

lexical conditioning

When allomorphy is triggered by the lexical properties of the stem, such as its semantics. This also includes allomorphy which is not predictable from a pattern.

lexical parts of speech

Parts of speech that encode the meaning of the sentence, such as nouns, verbs, adjectives, and adverbs.

lexicographer

A person who compiles a dictionary.

lexicon

The part of our memory where we store the vocabulary we have learned.

linguistic elicitation

Working with a user of a language to collect linguistic data from that language, through asking them to translate sentences to and from the language or by asking them to describe different situations or contexts in the language.

linguistic typology

The study of the classification of languages according to their structural properties.

linguistics

The scientific study of language.

loan translation

A borrowed phrase which is a direct translation of a phrase in another language.

loanword

A word that has been borrowed from another language.

metalanguage

the language being used to discuss the object language.

metathesis

IPA: /məˈtæθəˌsɪs/
The reordering of phonological segments.

Middle English

The variety of English spoken after the Norman invasion of England in 1066 until approximately the year 1500.

Minimalism

A research program/framework in generative syntax, beginning in the 1990s and continuing through the present day, with the goal of modeling a maximally efficient computational system to account for the attested syntactic patterns of the languages of the world.

modality

The type of signal through which language is transmitted and perceived, such as spoken or signed.

model

A scientific model is a representation of a system, an organism, a process, or a phenomenon.

Modern English

The varieties of English spoken from approximately the year 1500 until the present day.

morpheme

The smallest meaningful unit in language.

morphological conditioning

When allomorphy is triggered by morpological context, usually inflectional features of the stem.

morphology

The study of the structure of words; the study of the systematic co-variation of form and meaning.

morphosyntactic feature

The grammatical properties of a word, morpheme, or syntactic head which are relevant to its morphology or syntax.

morphosyntax

The study of the structure of words, phrases, and sentences. The intersection of the fields of morphology and syntax.

mutual intelligibility

When speakers from two different language varieties can understand each other.

native speaker

Someone who has been speaking the language in question from early childhood.

nativization

The process of a borrowed word adopting the phonology and possibly the morphology of the language it is borrowed into.

natural language

Language that arose out of use in a community, rather than a language that was designed or constructed.

necessary evidence

Evidence that must be the case if your hypothesis is true.

negative concord

Agreement between multiple negative words in a sentence.

negative evidence

Evidence that something is not possible, does not occur, or is absent.

neologism

The coining of a new word.

neurolinguistics

The study of how language is related to the anatomy of the brain.

node

Any point where multiple branches meet in a syntax tree, as well as the endpoint of any branch.

nominative

Case marking that appears on the subjects of finite clauses.

non-compositional meaning

A word or phrase with meaning that is not predictable from the combination of its components.

non-concatenative

Morphological processes that do not result in a linear string of morphemes, such as suprafixes, transfixes, or ablaut.

non-finite

synonym: infinitive
Not marked for tense.

nonce word

A made-up word for one-time use.

object language

the language under discussion or being analyzed.

observationally adequate

A model that accounts for all observed data.

Old English

The variety of English spoken before the Norman invasion of England in 1066.

open parts of speech

Parts of speech that can easily have new words added to them.

original research

Research that involves the creation, discovery, or documentation of new knowledge and ideas.

orthography

The writing system of a language.

overregularization

The process of applying a rule in contexts where it should not apply, common in child speech.

paradigm

The collection of all the word forms in a lexeme is called the paradigm for that lexeme, or the set of all the inflected forms of a word.

parasynthesis

A phenomenon in which one meaning or group of meanings is encoded simultaneously by more than one morphological element or process.

part of speech

synonyms: lexical category, syntactic category
We classify words into part of speech based on their syntactic function. Some examples of part of speech include noun, verb, adjective, and preposition.

partitive

Case used to mark a variety of meanings including partialness and incompleteness.

passive voice

A grammatical structure in which the subject of a clause is deleted or included in an optional 'by'-phrase, and the object of the clause is promoted to subject position.

past participle

A form of the English verb used in the perfect (e.g., *have sung* or *have eaten*) or in the passive (e.g., *was sung* or *was eaten*).
The past participle is often marked with *-en* or shares the same form as the past tense, but there are also many verbs with an irregular past participle.

peer review

A process where a piece of scholarship is reviewed by peers of the author. In the context of academic sources, the author and the author's peers would be expected to be experts on the topic.

perfect aspect

The perfect aspect is used to indicate an event which is completed by the indicated tense. Thus, in the "present perfect," the event of the verb has been completed by the present moment. In the "past perfect," the event of the verb has been completed at some past moment. In English, perfect is expressed using a form of the auxiliary verb HAVE and the past participle form of the following verb, e.g., "John has eaten sushi."

performance

The production of language that conforms to the rules of grammar of a particular language, regardless of whether the rules in question have been acquired by the language user.

phonetics

The study of the physical properties of language, including how it is produced and perceived.

phonological allomorph

Allomorphy that is predictable based on phonological context.

phonological conditioning

When allomorphy is triggered by phonological context.

phonology

The study of the systematic rules and constraints that characterize human speech sounds and signs in and across languages.

phrase

Any grammatically well-formed string of words.

polymorphemic

Containing multiple morphemes.

polysynthetic language

A language which combines many morphemes into a single, complex word, which often corresponds to an entire sentence.

portmanteau

A word formed by combining portions of two or more roots. At least one of the roots must be truncated in order to be a portmanteau.

possessive

A grammatical marking indicating possession.

postposition

A word describing the relationship between a noun and another part of the phrase and that occurs after the noun. See also *adposition.*

poverty of the stimulus argument

An argument used to support Universal Grammar based on the observation that language is infinite and the premise that infinite systems are unlearnable.

pragmatics

The study of the meaning of language in context.

predicate

The part of a clause that describes the event that applies to the subject. Typically the verb phrase, but non-verbal predicates also exist.

prefix

A morpheme that attaches to the beginning of a stem.

preposition

A word describing the relationship between a noun and another part of the phrase and that occurs before the noun. See also *adposition.*

prepositional ditransitive

A ditransitive construction where the direct object comes first as an NP and the indirect object comes second as a PP, as in *I gave a book to you*.

prepositional phrase

A phrase with a preposition as its head, frequently modified by a noun phrase.

prescriptive grammar

A set of rules that describe someone's opinion of how language should be used.

present participle

An inflectional form of verbs used in the progressive aspect, formed by suffixing *-ing* to a verb stem.

productivity

The property of language that allows language users to create novel utterances and have those utterances be understood by others.

progressive aspect

The progressive is an aspectual category that indicates an incomplete or ongoing action with respect to a particular event at a time. In English, progressive aspect is expressed using a form of the verb BE and the present participle form of the verb, e.g., 'John is eating sushi.'

proper noun

A noun that refers to a specific entity by its name.

proposition

A statement that can be true or false.

prosody

The patterns of stress and intonation in a language.

psycholinguistics

The study of how language is processed and planned.

quantifer

A kind of determiner that does not refer to specific entities but indicates some proportion of a group, for example *each*, *every*, *all*, *many*, and *some*.

recursion

Iteratively applying a rule to its own output.

Reduplication

A morphological process involving the full or partial copying of the stem.

regularization

A historical process of language change whereby irregular forms adopt the majority pattern.

replicable

Research that can be repeated in different contexts with similar results.

root

The core morpheme of a word.

root creation

Coining a word without basis on any previously existing words.

root-and-pattern morphology

A way of forming words common in Semitic languages, in which words are formed by adding transfixes to triconsonantal roots. The transfixes determine the arrangement of the consonants and the vowels that are inserted in between the three consonants of the root, and also sometimes additional consonants.

Scientific Method

The method of inquiry used in science whereby researchers observe data, and then formulate, test, and revise hypotheses based on their observations.

scope

The part of the sentence that is modified by a semantic operator.

segmented

(of a word) Broken up into morphemes.

selection

The requirements of a head for what it must combine with. The head selects the type, number, and semantic role of what it combines with. There is semantic selection (s-selection) and category selection (c-selection).

semantics

The study of the conventional meaning of language, including the meaning of individual words (lexical semantics) and how meaning is derived from their combination (compositional semantics).

sentence

A grammatically well-formed string of words that communicates a complete thought.

serial verb

A construction where the predicate consists of a string of multiple verbs.

sibilant

High-pitched consonant pronounced with the tongue tip, such as /s z ʃ ʒ t͡ʃ d͡ʒ/.

simplex word

A word containing only a single morpheme.

social science

The study of societies.

sociolinguistics

The study of how language is used in its social context, including variation based on broad social categories like age, gender, and region, and the intersection of language and identity.

stem

The portion of a word that hosts an affix.

strong suppletion

Allomorphy which is not predictable based on phonological context and has no phonological similarity to other forms in the paradigm.

structural ambiguity

When a word or sentence can be associated with more than one hierarchical structure, each resulting in different possible meanings.

subject

What a sentence is about. Usually a noun phrase. In English, a subject will appears before the verb and triggers agreement on the verb.

subject-auxiliary inversion

When an auxiliary verb appears in front of the subject, usually in order to form a question. Formally, this is analyzed as T-to-C movement.

sufficient evidence

Evidence that demonstrates your hypothesis is true.

suffix

A morpheme that attaches to the end of a stem.

superlative

An inflectional form of adjectives meaning "to the highest degree of X."

suppletion

Allomorphy which is not predictable from phonological context.

suprafix

An affix that is encoded through the prosodic structure of the stem, such as through stress, length, or pitch.

synchronic

Describing something (such as language) at a particular point in time.

synonym

A word that shares the same or similar meaning with another word in the same language.

syntactic category

synonyms: lexical category, part of speech
We classify words into part of speech based on their syntactic function. Some examples of part of speech include noun, verb, adjective, and preposition.

syntax

The study of the structure of sentences and phrases; the study of how we combine words and encode the relationships between them.

synthesis

A scale indexing how many morphemes are contained in a word.

taboo avoidance

The coining of a word in order to avoid use of a word that is socially unacceptable.

tense

Verbal inflection that indicates when the event took place, such as past, present, or future marking.

transfix

An affix that is interwoven among the segments of the root, such as in root-and-pattern morphology in Semitic languages.

triconsonantal root

A root consisting only of three consonants, as is common in root-and-pattern morphology in Semitic languages.

truncation

The formation of a new word by shortening an existing one.

unattested

A pattern or piece of data that doesn't show up in our data source.

underlying form

the form of a word or morpheme before phonological processes have been applied to it

Universal Grammar (UG)

Chomsky's theory that there is an innate, language-specific genetic component underlying the human capacity for language.

voiced

Pronounced with vocal fold vibration.

voiceless

Pronounced without vocal fold vibration.

Weak suppletion

Allomorphy that is not predictable from phonological context but nonetheless has some similarity with other forms in the paradigm.

word family

A set of words related by derivation.

word form

A particular form of a lexeme, one of the possible inflected forms of a lexeme.

word token

a particular occurrence of a word

wug test

An experimental study design whereby participants manipulate novel word forms, testing whether a word formation rule is productive.

LIST OF ABBREVIATIONS

Parts of speech abbreviations

Abbreviation	Meaning
Adj	adjective
Adv	adverb
C	complementizer
Conj	conjunction
D	determiner
N	noun
P	preposition
T	tense
V	verb

Glossing abbreviations

Abbreviation	Meaning
1	first person
2	second person
3	third person
ABS	absolutive **case**
ACC	accusative **case**
CN	**common noun** marker
DUR	durative **aspect**
ERG	ergative **case**
F	feminine noun class
FUT	future **tense**
HAB	habitual **aspect**
M	masculine noun class
NOM	nominative **case**
OBJ	object
PFV	perfective **aspect**
PL	plural number
PN	**proper noun** marker
PolQP	polarity question particle
PRS	present **tense**
PST	past **tense**
REAL	realis mood
SG	singular number
SUBJ	subjunctive mood
VTI	transitive verb with inanimate goal

Other abbreviations

Abbreviation	Meaning
AI	artificial intelligence
CJL	Canadian Journal of Linguistics
IPA	International Phonetic Alphabet
OED	Oxford English Dictionary
UG	Universal Grammar

LISTS OF TABLES AND FIGURES

Tables

Chapter 1. What are morphology and syntax?

Section 1.1. What is morphology?

Table 1. Words with *un–*

Chapter 2. The scientific method in morphosyntax

Section 2.3. Prescriptivism and descriptivism

Table 1. Some contractions of standardized English
Table 2. Examples of prescriptive rules in English
Table 3. Change in the comparative and superlative forms of fun

Section 2.4. The lexicon and dictionaries

Table 1. The differences between mental lexicons and dictionaries

Section 2.5. Becoming a linguist: Identifying academic sources

Table 1. Kinds of publications

Chapter 4. Linguistic data in morphosyntax

Section 4.2. Methods of data collection

Table 1. Classification of data collection methods

Section 4.7. Becoming a linguist: Glossing spoken language data

Table 1. Acceptability markings

Section 4.8. Becoming a linguist: Glossing signed language data

Table 1. Common symbols in signed language glosses (adapted from Baker et al. 2016: 338-339)
Table 2. Common abbreviations for non-manual markers in ASL (adapted from Baker et al. 2016: 339-340)

Chapter 4 practice exercises

Table 1. Mini Yoruba dictionary
Table 2. Yoruba pronoun paradigm

Chapter 6. Kinds of morphemes and morphological processes

Section 6.10. Guided reading: Is n't a clitic or an affix?

Table 1: Non-standard transciption used in Zwicky and Pullum (1983)

Chapter 7. Predicates, arguments, and clauses

Section 7.2. Arguments

Table 1. Kinds of transitivity
Table 2. Ditransitive constructions

Section 7.4. Statements, questions, and commands

Table 1. Combinations of main and embedded clause functions

Chapter 8. Parts of speech

Section 8.1. Determining part of speech

Table 1: Traditional semantic definitions of parts of speech

Section 8.2. Nouns

Table 1. English derivational morphemes
Table 2. English personal pronouns
Table 3. The meanings of some common cases.
Table 4. Niuean personal pronouns (Seiter 1980: 49)
Table 5. Inflectional paradigms of different gender singular nouns in Russian
Table 7. Deverbal nouns (V → N) (adapted from Haspelmath and Sims 2010: 87)
Table 8. Deadjectival nouns (A → N) (adapted from Haspelmath and Sims 2010: 87)
Table 9. Denominal nouns (N → N) (adapted from Haspelmath and Sims 2010: 87)

Section 8.4. Adjectives and adverbs

Table 1. The inflectional paradigms of some English adjectives
Table 2. Deverbal adjectives (V → A) (adapted from Haspelmath and Sims 2010: 89)
Table 3. Denominal adjectives (N → A) (adapted from Haspelmath and Sims 2010: 89)
Table 4. Deadjectival adjectives (A → A) (adapted from Haspelmath and Sims 2010: 89)

Section 8.5. Functional parts of speech

Table 1. Subclasses of the determiner category
Table 2. Examples of some gender-neutral neopronouns in English
Table 3. Some possible parts of speech of question words in English
Table 4. Subclasses of the tense category

Chapter 9.Words

Section 9.1. Word formation processes

Table 1. Examples of derivation in English
Table 2. The productivity of -ish and -ness

Section 9.3. Packaging words and morphemes

Figures

Chapter 1. What are morphology and syntax?

Section 1.1. What is morphology?

Figure 1. An untidy workspace. Photo by SN.CHE. Used under Pexels license.
Figure 2. A selection of pumpkins. Photo by Allie Reefer. Used under Pexels license.
Figure 3. Language play. Photo by Thirdman. Used under Pexels license.

Section 1.3. Morphosyntax as a subfield of linguistics

Figure 1: A toddler kicking a ball in a field. Photo by Paul Ney. Used under CC BY-NC-SA license.

Chapter 2. The scientific method in morphosyntax

Section 2.2. Evidence and hypotheses

Figure 1. A decorative ghost. Photo by Juan Vargas. Used under Pexels license.
Figure 2. Four macarons. Photo by Arminas Raudys. Used under Pexels license.

Section 2.4. The lexicon and dictionaries

Figure 1: The Dictionary of Proto-Algonquian by David Pentland. This dictionary includes a lot of etymological information about numerous Algonquian languages, which would not be included in the mental lexicons of speakers. Photo by Julie Doner. Reproduced with permission.

Section 2.6. Becoming a linguist: Scientific writing in morphosyntax

Figure 1: A model of a carbon atom. Photo by SrKellyOP. Used under CC0 1.0 license.

Chapter 3. Theories of grammar and language acquisition

Section 3.1. Becoming a linguist: Scientific writing in morphosyntax

Figure 1. A speed limit sign. Photo by Shahnoor Habib Munmun. Used under CC BY 3.0 license.
Figure 2. A bookshelf at a university library in the Indigenous languages section. This section contains reference and pedagogical grammars, as well as dictionaries, texts in Indigenous languages, and academic books on specific topics in Indigenous linguistics. What kinds of books can you spot in this picture? Photo by Julie Doner.
Figure 3: A page from a pedagogical grammar of Yoruba (Mosadomi 2011). Yoruba is a member of the Niger-Congo language family and is spoken primarily in Nigeria with over 43 million native speakers (Eberhard et al. 2023).
Figure 4: A picture of page 35 of Bloomfield (1957), a classic reference grammar of Nishnaabemwin (also known as Eastern Ojibwa). Nishnaabemwin belongs to the Algonquian language family and is spoken by about 220 people in Canada, according to the 2021 census. It is spoken in and around Lake Huron, especially Manitoulin Island (Eberhard et al. 2023). Photo by Julie Doner.

Section 3.2. Generative grammar

Figure 1: Noam Chomsky in 2004. Photo by Duncan Rawlinson. Used under CC BY-NC 2.0 license.
Figure 2: A bicycle. Photo by LUM3N. Used under Pixabay Content License.

Chapter 4. Linguistic data in morphosyntax

Section 4.8. Becoming a linguist: Glossing signed language data

Figure 1. Differences between ASL and STS in handshape meanings (Anderson et al. 2022 Section 3.8).
Figure 2. Different orientations of the right hand as seen from the signer's point of view (Anderson et al. 2022 Section 3.8).

Chapter 6. Kinds of morphemes and morphological

processes

Section 6.2. Affixes

Figure 1. A stack of five books. Photo by Julie Doner.
Figure 2. A stack of books with one book inside of another book. Photo by Julie Doner.

Section 6.8. Becoming a linguist: Anatomy of an academic article

Figure 1. The first page of an academic journal article, Bjorkman (2022), published by the Canadian Journal of Linguistics. Used under CC BY 4.0 license.

Chapter 8. Parts of speech

Section 8.2. Nouns

Figure 1: A schematic representation of nominative-accusative case alignment.
Figure 2: A schematic representation of ergative-absolutive case alignment.

GLOSS TEMPLATE HTML AND CSS

These HTML templates and the accompanying CSS stylesheet facilitate the presentation of glosses in online linguistic documentation. They offer a clean, organized structure using basic HTML tables with standard row and data cell tags, a structure which ensures accessibility for screen readers and keyboard navigation without needing the header cells or captions required for accessibility in standard use of HTML tables. The templates are easily customizable to fit various linguistic projects, and the stylesheet provides a cohesive and consistent gloss design across the templates.

Gloss Template

(1)	a.	John	goes	to	the	store$_1$
	b.	subject$_2$	verb	preposition Small caps	article	noun
	c.	“John goes to the store.”				

Gloss Template HTML

```
<code>
<div class=”glosscontainer”>
<table class=”interlinear”>
<tbody>
<tr>
<td class=”list1">(1)</td>
<td class=”list2">a.</td>
```

```
<td class=”spacer”>John</td>
<td class=”spacer”>goes</td>
<td class=”spacer”>to</td>
<td class=”spacer”>the</td>
<td class=”spacer”>store<sub>1</sub></td>
</tr>
<tr>
<td><span style=”color: #fff;”>No Data</span></td>
<td class=”list2">b.</td>
<td class=”spacer”>subject<sub>2</sub></td>
<td class=”spacer”>verb</td>
<td class=”spacer”>preposition
<p class=”small-caps”>Small caps</p>
</td>
<td class=”spacer”>article</td>
<td class=”spacer”>noun</td>
</tr>
<tr>
<td><span style=”color: #fff;”>No Data</span></td>
<td class=”list2">c.</td>
<td colspan=”5">”John goes to the store.”</td>
</tr>
</tbody>
</table>
</div>
</code>
```

Sign Language Gloss Template

			re
(37)	a.	$INDEX_2$ BITE-BAR	LONG!
		‘Bite the bar for a long time!’	
			hn
			re
	b.	PLEASE $INDEX_2$ BITE!	
		‘Please, bite it!’	

Sign Language Gloss Template HTML

```
<code>
<div class=”glosscontainer”>
<table class=”interlinear-SL”>
<tr>
<td></td>
<td></td>
<td class=”non-manual” colspan=”4">re</td>
</tr>
<tr>
<td class=”list1">(37)</td>
<td class=”list2">a.</td>
<td class=”spacer-left cell-border”>INDEX<sub>2</sub></td>
<td class=”spacer-center cell-border” colspan=”2">BITE-BAR</td>
<td class=”spacer-right cell-border”>LONG!</td>
</tr>
<tr>
<td></td>
```

```
<td></td>
<td colspan="4">'Bite the bar for a long time!'</td>
</tr>
<tr>
<td></td>
<td></td>
<td class="spacer"></td>
<td class="non-manual" colspan="2">hn</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td class="spacer cell-border"></td>
<td class="spacer cell-border"></td>
<td></td>
</tr>
<tr>
<tr>
<td></td>
<td></td>
<td class="non-manual" colspan="3">re</td>
<td></td>
</tr>
<tr>
<td></td>
<td class="list2">b.</td>
<td class="spacer-left cell-border">PLEASE</td>
<td class="spacer-center cell-border">INDEX<sub>2</sub></td>
<td class="spacer-right cell-border">BITE!</td>
<td class="spacer"></td>
</tr>
<tr>
<td></td>
<td></td>
```

```
<td colspan=”4">’Please, bite it!'</td>
</tr>
</table>
</div>
</code>
```

CSS Master Stylesheet for both templates

```
/* CSS Document */

/* Container for all elements */

.glosscontainer {
width: 50%; /* Adjust this width up to 100%, whatever your preference */
margin-left: auto; /* Delete if you don’t want the tables centered */
margin-right: auto; /* Delete if you don’t want the tables centered */
}

/* Table styling */

.interlinear {
border-collapse: collapse;
width: 100%;
}

.interlinear td {
border: 1px solid #ededed; /* Delete ‘1px solid #’ and replace with ‘none’ when glosses are all complete. */
padding: 8px;
}

/* Table top specific styling */

.list1 {
```

```
width: 3%;
text-align: center;
vertical-align: middle;
}
.list2 {
width: 2.5%;
}
.spacer {
width: 12%;
text-align: right; /* Change to 'center' in .spacer if needed */
vertical-align: middle;
}
/* Inline container styling */
.inline-container {
display: flex;
flex-wrap: nowrap; /* Prevents the items from wrapping */
align-items: center; /* Vertically centers the content in the container */
justify-content: center; /* Horizontally centers the content in the container */
}
.inline-container p {
margin-right: 10px; /* Optional: Adds some space between the paragraphs */
}
/* Sets italics for the bottom row */
.interlinear tr:last-child {
font-style: italic;
}
/* Small caps styling */
.small-caps {
font-variant: small-caps;
padding-top: 0;
margin: 0;
}
```

```
/* For sign language glosses ONLY*/

.interlinear-SL {
border-collapse: collapse;
width: 50%; /* Keeps table at a optimal width – can be adjusted */
}

td.cell-border {
border-top: solid 4px; /* Creates the border above cells */
}

td.spacer-left {
text-align: left; /* Controls text alignment in td cell */
vertical-align: middle;

}

td.spacer-center {
text-align: center; /* Controls text alignment in td cell */
vertical-align: middle;
}

td.spacer-right {
text-align: right; /* Controls text alignment in td cell */
vertical-align: middle;
}

.interlinear-SL td {
padding-top: 5px; /* Controls spacing between cells to prevent collapse */
padding-bottom: 5px;
}

.non-manual {
text-align: right; /* Alignment for non-manual ‘hn’ and ‘re’ elements */
}
```