

Lecture 1: Introduction and Encodings

LING-351 Language Technology and LLMs

Instructor: Hakyung Sung

August 26, 2025

*Acknowledgment: These course slides are based on materials by Lelia Glass @ Georgia Tech (Course: Language & Computers)

Table of contents

1. Introduction
2. Lesson plan
3. What is language?
4. Language vs. Writing
5. Encoding
6. Digital encoding of writing
7. Wrap-up

Introduction

- Instructor: Dr. Hakyung Sung

- Instructor: Dr. Hakyung Sung
 - PhD in Linguistics, MS in Computer Science @ University of Oregon

Course logistics

- Instructor: Dr. Hakyung Sung
 - PhD in Linguistics, MS in Computer Science @ University of Oregon
- Grader: Bea (Bey-uh) Pulido

Course logistics

- Instructor: Dr. Hakyung Sung
 - PhD in Linguistics, MS in Computer Science @ University of Oregon
- Grader: Bea (Bey-uh) Pulido
 - a second year Experimental Psychology graduate student

Course logistics

- Instructor: Dr. Hakyung Sung
 - PhD in Linguistics, MS in Computer Science @ University of Oregon
- Grader: Bea (Bey-uh) Pulido
 - a second year Experimental Psychology graduate student
- Time: Tu/Th 2:00PM-3:15PM

Course logistics

- Instructor: Dr. Hakyung Sung
 - PhD in Linguistics, MS in Computer Science @ University of Oregon
- Grader: Bea (Bey-uh) Pulido
 - a second year Experimental Psychology graduate student
- Time: Tu/Th 2:00PM-3:15PM
- Office: EAS 3173

Course logistics

- Instructor: Dr. Hakyung Sung
 - PhD in Linguistics, MS in Computer Science @ University of Oregon
- Grader: Bea (Bey-uh) Pulido
 - a second year Experimental Psychology graduate student
- Time: Tu/Th 2:00PM-3:15PM
- Office: EAS 3173
- Office hour: TuTh 3:30-4:30 in-person, or Zoom by appointment

Course logistics

- Instructor: Dr. Hakyung Sung
 - PhD in Linguistics, MS in Computer Science @ University of Oregon
- Grader: Bea (Bey-uh) Pulido
 - a second year Experimental Psychology graduate student
- Time: Tu/Th 2:00PM-3:15PM
- Office: EAS 3173
- Office hour: TuTh 3:30-4:30 in-person, or Zoom by appointment
- Course website:
https://hksung.github.io/Fall25_LING351/

Course logistics

- Instructor: Dr. Hakyung Sung
 - PhD in Linguistics, MS in Computer Science @ University of Oregon
- Grader: Bea (Bey-uh) Pulido
 - a second year Experimental Psychology graduate student
- Time: Tu/Th 2:00PM-3:15PM
- Office: EAS 3173
- Office hour: TuTh 3:30-4:30 in-person, or Zoom by appointment
- Course website:
https://hksung.github.io/Fall25_LING351/
- Email: hks gla@rit.edu

Learning goals

- We'll explore the interaction between language and technology.
Topics including:

Learning goals

- We'll explore the interaction between language and technology.
Topics including:
 - Writing assistance tools

Learning goals

- We'll explore the interaction between language and technology.
Topics including:
 - Writing assistance tools
 - Computer-assisted language learning

Learning goals

- We'll explore the interaction between language and technology.
Topics including:
 - Writing assistance tools
 - Computer-assisted language learning
 - Chatbots

Learning goals

- We'll explore the interaction between language and technology.
Topics including:
 - Writing assistance tools
 - Computer-assisted language learning
 - Chatbots
 - Machine translation

Learning goals

- To understand these systems, basic coding skills are helpful

Learning goals

- To understand these systems, basic coding skills are helpful
- We will do hands-on exercises during class (Python tutorials)

Learning goals

- To understand these systems, basic coding skills are helpful
- We will do hands-on exercises during class ([Python tutorials](#))
- [No prior coding experience is required—tutorials will start from the very beginning](#)

- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit [NLTK]*

- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit* [NLTK]
- Glass, I., Dickinson, M., Brew, C., & Meurers, D. (2024). *Language and Computers* (2nd edition) [LC].

Updated reading materials

- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit* [NLTK]
- Glass, I., Dickinson, M., Brew, C., & Meurers, D. (2024). *Language and Computers* (2nd edition) [LC].
- All books are available as pdf. (publicly available).

Updated reading materials

- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit* [NLTK]
- Glass, I., Dickinson, M., Brew, C., & Meurers, D. (2024). *Language and Computers* (2nd edition) [LC].
- All books are available as pdf. (publicly available).
- Please check course website.

Final grading components

[a × b] a = number; b = points

- Exercises [4 × 10]: 40%
- Assignments [2 × 10] 20%
- Paper presentations [2 × 5] 10%
- Online exams 30%
 - Midterm [1 × 15] 15%
 - Final [1 × 15]: 15%

Final grading components

- Exercises [4 × 10]: 40%
- These are individual assignments, usually based on the work you complete during in-class lab sessions.

Final grading components

- Exercises [4 × 10]: 40%

Week	Date	Topic	Readings	Due (Friday, 11:59 pm)
1	8/26	Introduction, Encoding	[LC] Ch.1	
	8/28	Writer's aids: Spelling errors	[LC] Ch.2.1-2.3	
2	9/2	Writer's aids: Grammar errors	[LC] Ch.2.5-2.8	
	9/4	Computer-assisted language learning	[LC] Ch. 3	
3	9/9	Text as data	[LC] Ch. 4.1-4.3	
	9/11	Python tutorial 1		Exercise 1
4	9/16	Python tutorial 2		
	9/18	Python tutorial 3		Exercise 2
5	9/23	Python tutorial 4		
	9/25	Python tutorial 5		Exercise 3
9	10/21	Building a chatbot	[LC] Ch. 8.3	
	10/23	Prompt engineering		Exercise 4

Final grading components

- Exercises [4 × 10]: 40%

Week	Date	Topic	Readings	Due (Friday, 11:59 pm)
1	8/26	Introduction, Encoding	[LC] Ch.1	
	8/28	Writer's aids: Spelling errors	[LC] Ch.2.1-2.3	
2	9/2	Writer's aids: Grammar errors	[LC] Ch.2.5-2.8	
	9/4	Computer-assisted language learning	[LC] Ch. 3	
3	9/9	Text as data	[LC] Ch. 4.1-4.3	
	9/11	Python tutorial 1		Exercise 1
4	9/16	Python tutorial 2		
	9/18	Python tutorial 3		Exercise 2
5	9/23	Python tutorial 4		
	9/25	Python tutorial 5		Exercise 3
9	10/21	Building a chatbot	[LC] Ch. 8.3	
	10/23	Prompt engineering		Exercise 4

- Please bring your laptop on these days!

Final grading components

- Assignments [2 × 10] 20%
- Paper presentations [2 × 5] 10%
- https://youtube.com/shorts/Yg7WrDt5I1E?si=12YMKYi_OJRj9c6r

- To apply NLP technologies to a given domain, we need a basic understanding of that domain.

Final grading components

- To apply NLP technologies to a given domain, we need a basic understanding of that domain.
- We'll research articles on the use of NLP in the (1) humanities, (2) social sciences, (3) language studies, and (4) the impact of LLMs, with an emphasis on conceptual understanding.

Final grading components

- To apply NLP technologies to a given domain, we need a basic understanding of that domain.
- We'll research articles on the use of NLP in the (1) humanities, (2) social sciences, (3) language studies, and (4) the impact of LLMs, with an emphasis on conceptual understanding.
- All paper links are on the course website!

Final grading components

- To apply NLP technologies to a given domain, we need a basic understanding of that domain.
- We'll research articles on the use of NLP in the (1) humanities, (2) social sciences, (3) language studies, and (4) the impact of LLMs, with an emphasis on conceptual understanding.
- All paper links are on the course website!
- 2 people will be grouped to present papers in each area.

Final grading components

- Weeks 10-13

10	10/28	Prompt engineering		
	10/30	Paper presentation (Papers 1, 2)		
11	11/4	Paper presentation (3, 4)		
	11/6	Paper presentation (5, 6)		
12	11/11	Paper presentation (7, 8)		
	11/13	Paper presentation (9, 10)		Assignment 1
13	11/18	Paper presentation (11, 12)		
	11/20	Paper presentation (13, 14)		
14	11/25	Paper presentation (15, 16)		
	11/27	Thanksgiving break (No class)		
15	12/2	Paper presentation (17, 18)		
	12/4	Final wrap-up		Assignment 2

- Week 6

6	9/30	Word vectors	[LC] Ch. 4.4	
	10/2	Text classification	[LC] Ch. 5	Student presentation topics submission

Final grading components

- Assignments [2 × 10]: 20%

10	10/28	Prompt engineering		
	10/30	Paper presentation (Papers 1, 2)		
11	11/4	Paper presentation (3, 4)		
	11/6	Paper presentation (5, 6)		
12	11/11	Paper presentation (7, 8)		
	11/13	Paper presentation (9, 10)		Assignment 1
13	11/18	Paper presentation (11, 12)		
	11/20	Paper presentation (13, 14)		
14	11/25	Paper presentation (15, 16)		
	11/27	Thanksgiving break (No class)		
15	12/2	Paper presentation (17, 18)		
	12/4	Final wrap-up		Assignment 2

- Each group presents twice (Rounds 1–9; Rounds 10–18).

Final grading components

- Assignments [2 × 10]: 20%

10	10/28	Prompt engineering		
	10/30	Paper presentation (Papers 1, 2)		
11	11/4	Paper presentation (3, 4)		
	11/6	Paper presentation (5, 6)		
12	11/11	Paper presentation (7, 8)		
	11/13	Paper presentation (9, 10)		Assignment 1
13	11/18	Paper presentation (11, 12)		
	11/20	Paper presentation (13, 14)		
14	11/25	Paper presentation (15, 16)		
	11/27	Thanksgiving break (No class)		
15	12/2	Paper presentation (17, 18)		
	12/4	Final wrap-up		Assignment 2

- Each group presents twice (Rounds 1–9; Rounds 10–18).
- For each round, students will also submit a short assignment summarizing what they learned from (1) the presented studies and (2) other presentations.

Final grading components

- Assignments [2 × 10]: 20%

10	10/28	Prompt engineering		
	10/30	Paper presentation (Papers 1, 2)		
11	11/4	Paper presentation (3, 4)		
	11/6	Paper presentation (5, 6)		
12	11/11	Paper presentation (7, 8)		
	11/13	Paper presentation (9, 10)		Assignment 1
13	11/18	Paper presentation (11, 12)		
	11/20	Paper presentation (13, 14)		
14	11/25	Paper presentation (15, 16)		
	11/27	Thanksgiving break (No class)		
15	12/2	Paper presentation (17, 18)		
	12/4	Final wrap-up		Assignment 2

- Each group presents twice (Rounds 1–9; Rounds 10–18).
- For each round, students will also submit a short assignment summarizing what they learned from (1) the presented studies and (2) other presentations.
- Assignments are released at the start of each round and due at the end of the presentation day.

Final grading components

- Online exam: 30%
 - Midterm [1 × 15]: 15%
 - Final [1 × 15]: 15%

Grading policy

- **2-hr grading window:** Any assignment submitted online will automatically have a 2-hour grading window. This will be applied by the system, and no action is required from students.

Grading policy

- **2-hr grading window:** Any assignment submitted online will automatically have a 2-hour grading window. This will be applied by the system, and no action is required from students.
- **Late penalty:** Late assignments will incur a 10% deduction per day, for up to 5 days (e.g., 1 day late = 10% off). After 5 days, the assignment will receive a grade of zero.

Grading policy

- **2-hr grading window:** Any assignment submitted online will automatically have a 2-hour grading window. This will be applied by the system, and no action is required from students.
- **Late penalty:** Late assignments will incur a 10% deduction per day, for up to 5 days (e.g., 1 day late = 10% off). After 5 days, the assignment will receive a grade of zero.
- **Extenuating circumstances:** Whenever possible, please request an official document that can prove the circumstances—this allows me to accommodate you fairly while respecting your privacy.

Grading policy

- **2-hr grading window:** Any assignment submitted online will automatically have a 2-hour grading window. This will be applied by the system, and no action is required from students.
- **Late penalty:** Late assignments will incur a 10% deduction per day, for up to 5 days (e.g., 1 day late = 10% off). After 5 days, the assignment will receive a grade of zero.
- **Extenuating circumstances:** Whenever possible, please request an official document that can prove the circumstances—this allows me to accommodate you fairly while respecting your privacy.
 - If that is not possible, contact me as soon as you can. Extensions are generally not granted retroactively.

Grading policy

- **2-hr grading window:** Any assignment submitted online will automatically have a 2-hour grading window. This will be applied by the system, and no action is required from students.
- **Late penalty:** Late assignments will incur a 10% deduction per day, for up to 5 days (e.g., 1 day late = 10% off). After 5 days, the assignment will receive a grade of zero.
- **Extenuating circumstances:** Whenever possible, please request an official document that can prove the circumstances—this allows me to accommodate you fairly while respecting your privacy.
 - If that is not possible, contact me as soon as you can. Extensions are generally not granted retroactively.
- **No extensions will be granted for the online exam.**

- For the group works, all members are expected to contribute their time and effort equally.

- For the group works, all members are expected to contribute their time and effort equally.
- Each submission will include a section outlining both individual and group contributions, which will be evaluated separately.

- For the group works, all members are expected to contribute their time and effort equally.
- Each submission will include a section outlining both individual and group contributions, which will be evaluated separately.
- Collaboration with AI tools is permitted, but you are responsible for the quality and integrity of the work produced.

- For the group works, all members are expected to contribute their time and effort equally.
- Each submission will include a section outlining both individual and group contributions, which will be evaluated separately.
- Collaboration with AI tools is permitted, but you are responsible for the quality and integrity of the work produced.
- You must acknowledge and document how AI tools were used in your work (including individual exercises).

Any questions?

Lesson plan

- Spelling challenges in writing

- Spelling challenges in writing
- Types of spelling errors

- Spelling challenges in writing
- Types of spelling errors
- Building a simple spell checker

- Spelling challenges in writing
- Types of spelling errors
- Building a simple spell checker
- Extending to more complex systems

- Spelling challenges in writing
- Types of spelling errors
- Building a simple spell checker
- Extending to more complex systems
- Spell checkers across different languages

- Spelling challenges in writing
- Types of spelling errors
- Building a simple spell checker
- Extending to more complex systems
- Spell checkers across different languages
- Wrap-up

- Spelling challenges in writing
- Types of spelling errors
- Building a simple spell checker
- Extending to more complex systems
- Spell checkers across different languages
- Wrap-up

Key idea: Spelling errors are annoying

- Spelling challenges in writing
- Types of spelling errors
- Building a simple spell checker
- Extending to more complex systems
- Spell checkers across different languages
- Wrap-up

Key idea: ~~Spelling errors are annoying~~ There is no one-size-fits-all solution for spelling errors (error types; different language systems)

What is language?

Charles Hockett's Design Features of Language (1960)

- **Modality:**
Spoken language is produced with the vocal tract and perceived by the auditory system; *Signed* language is produced with the body and perceived visually.
- **Intentionality:**
Language is produced deliberately for communication.
- **Transitoriness:**
Language is ephemeral unless recorded.
- **Interchangeability:**
Anything you can hear, you can also say.
- **Total feedback:**
Speakers can hear themselves and monitor their speech.
- **Primacy of communication:**
Language is used primarily for communication—not as a secondary function.
- **Semanticity:**
Specific words or signs are linked to specific meanings.
- **Arbitrariness:**
The connection between a sign and its meaning is largely conventional.
- **Discreteness:**
Continuous variation is categorized into discrete mental units.
- **Displacement:**
Language allows reference to things not present—past, future, imaginary.
- **Prevarication:**
Language can be used to lie or deceive.
- More:
https://en.wikipedia.org/wiki/Hockett%27s_design_features

Which are Languages?

Let's test Hockett's design features!

Are the following systems *languages*?

Why or why not?

Is *Music* a Language?

- Can music express specific meanings?

Is *Music* a Language?

- Can music express specific meanings?
- Can it refer to imaginary or absent things?

Is *Music* a Language?

- Can music express specific meanings?
- Can it refer to imaginary or absent things?
- Can music be used to lie?

Is *Music* a Language?

- Can music express specific meanings?
- Can it refer to imaginary or absent things?
- Can music be used to lie?
- Is the relationship between musical symbols and meanings arbitrary?

Is *Music* a Language?

- Can music express specific meanings?
- Can it refer to imaginary or absent things?
- Can music be used to lie?
- Is the relationship between musical symbols and meanings arbitrary?

More discussions + a small break

- How many of Hockett's features does *music* meet?
- Is *Python* a Language?
- Is *Mathematics* a Language?

Language vs. Writing

- Tell stories, ask questions, learn, plan, imagine alternate realities

- Tell stories, ask questions, learn, plan, imagine alternate realities
- Coordinate with others and build culture

Language is *technology*

- Tell stories, ask questions, learn, plan, imagine alternate realities
- Coordinate with others and build culture
- All human societies use it

Language is *technology*

- Tell stories, ask questions, learn, plan, imagine alternate realities
- Coordinate with others and build culture
- All human societies use it
- Estimated age: 100,000–200,000 years

Language is *technology*

- Tell stories, ask questions, learn, plan, imagine alternate realities
- Coordinate with others and build culture
- All human societies use it
- Estimated age: 100,000–200,000 years
- Evidence? Archaeological findings (e.g., symbolic beads, tools, burial sites)



Figure 1: Clay tablet inscribed with the earliest known writing system, cuneiform—recording the receipt of barley and malt (around 3000 BCE, left)—and a close-up of cuneiform text on a mudbrick (around 1200 BCE).

Sourced from: <https://en.wikipedia.org/wiki/Cuneiform>

What is *Writing*?

Writing is another amazing technology!

- Records language, which is otherwise ephemeral

What is *Writing*?

Writing is another amazing technology!

- Records language, which is otherwise ephemeral
- Makes language usable across time and space

What is *Writing*?

Writing is another amazing technology!

- Records language, which is otherwise ephemeral
- Makes language usable across time and space
- Key to history, law, science, culture

Writing is another amazing technology!

- Records language, which is otherwise ephemeral
- Makes language usable across time and space
- Key to history, law, science, culture
- Not all people or societies use writing

What is *Writing*?

Writing is another amazing technology!

- Records language, which is otherwise ephemeral
- Makes language usable across time and space
- Key to history, law, science, culture
- Not all people or societies use writing
- Estimated age: 5,000–6,000 years

Case 1. Same writing system, different languages:

Case 1. Same writing system, different languages:

- Latin alphabet used in: English, French, German, Vietnamese

Case 1. Same writing system, different languages:

- Latin alphabet used in: English, French, German, Vietnamese

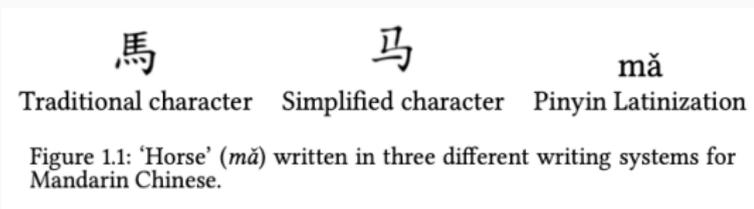
Case 2. Same language, different writing systems:

Case 1. Same writing system, different languages:

- Latin alphabet used in: English, French, German, Vietnamese

Case 2. Same language, different writing systems:

- Chinese: traditional vs. simplified vs. pinyin (Latinized)



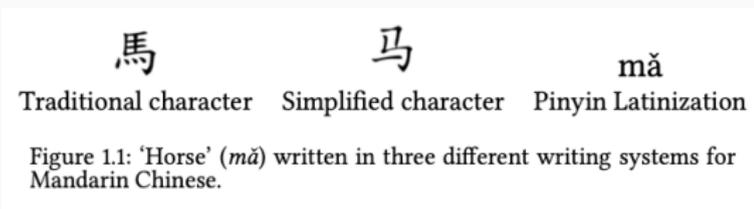
- Turkish: Arabic script (pre-1928) vs. Latin script (modern)

Case 1. Same writing system, different languages:

- Latin alphabet used in: English, French, German, Vietnamese

Case 2. Same language, different writing systems:

- Chinese: traditional vs. simplified vs. pinyin (Latinized)



- Turkish: Arabic script (pre-1928) vs. Latin script (modern)
- Japanese: 1 language, 3 scripts—hiragana, katakana, kanji

Spot the misconception

“French is written in the English alphabet.”

Spot the misconception

“French is written in the English alphabet.”

Hmm... **Something's off!**

Is there such a thing as an “*English*” *alphabet*?

Spot the misconception

“French is written in the English alphabet.”

Hmm... **Something's off!**

Is there such a thing as an “*English*” *alphabet*?

- Both English and French use the **Latin alphabet**, a writing system shared by many languages.

Spot the misconception

“French is written in the English alphabet.”

Hmm... **Something's off!**

Is there such a thing as an “*English*” alphabet?

- Both English and French use the **Latin alphabet**, a writing system shared by many languages.
- But! Each language uses it differently:
 - French includes letters with diacritics: é, è, ê, ç
 - English doesn't use those in native words.

Spot the misconception

“French is written in the English alphabet.”

Hmm... **Something's off!**

Is there such a thing as an “*English*” alphabet?

- Both English and French use the **Latin alphabet**, a writing system shared by many languages.
- But! Each language uses it differently:
 - French includes letters with diacritics: é, è, ê, ç
 - English doesn't use those in native words.
- So, it's not that French borrows “English's” alphabet— they both adapt a shared system for their own phonology and grammar.

How language and writing work in language technology?

What is NLP?

What is NLP?

Natural Language Processing (NLP) is the field that enables computers to understand, analyze, and generate human language.

What is NLP?

Natural Language Processing (NLP) is the field that enables computers to understand, analyze, and generate human language.

- To process language with computers, NLP requires a way to **encode language** → that's where **writing systems** come in.

What is NLP?

Natural Language Processing (NLP) is the field that enables computers to understand, analyze, and generate human language.

- To process language with computers, NLP requires a way to **encode language** → that's where **writing systems** come in.
- Evolution of writing technologies: **clay** → **papyrus** → **printing press** → **digital text**

What is NLP?

Natural Language Processing (NLP) is the field that enables computers to understand, analyze, and generate human language.

- To process language with computers, NLP requires a way to **encode language** → that's where **writing systems** come in.
- Evolution of writing technologies: **clay** → **papyrus** → **printing press** → **digital text**
- **Digital writing allows for new forms of communication and makes language machine-readable.**

Any questions?

Encoding

What is encoded in writing?

- Language = (mostly arbitrary) sound-meaning pairs

Three major systems:

What is encoded in writing?

- Language = (mostly arbitrary) sound-meaning pairs
- Writing encodes **sound, meaning, or syllables**, but usually not all three.

Three major systems:

What is encoded in writing?

- Language = (mostly arbitrary) sound-meaning pairs
- Writing encodes **sound**, **meaning**, or **syllables**, but usually not all three.

Three major systems:

- **Alphabetic:** symbol → sound

What is encoded in writing?

- Language = (mostly arbitrary) sound-meaning pairs
- Writing encodes **sound**, **meaning**, or **syllables**, but usually not all three.

Three major systems:

- **Alphabetic**: symbol → sound
- **Syllabic**: symbol → syllable

What is encoded in writing?

- Language = (mostly arbitrary) sound-meaning pairs
- Writing encodes **sound**, **meaning**, or **syllables**, but usually not all three.

Three major systems:

- **Alphabetic**: symbol → sound
- **Syllabic**: symbol → syllable
- **Logographic**: symbol → meaning

1. Alphabetic systems

- Each character = one sound or articulatory gesture

1. Alphabetic systems

- Each character = one sound or articulatory gesture
- English has some exceptions:

1. Alphabetic systems

- Each character = one sound or articulatory gesture
- English has some exceptions:
 - Silent letters: *knee, debt*

1. Alphabetic systems

- Each character = one sound or articulatory gesture
- English has some exceptions:
 - Silent letters: *knee, debt*
 - One sound, multiple letters: *running, phone*

1. Alphabetic systems

- Each character = one sound or articulatory gesture
- English has some exceptions:
 - Silent letters: *knee, debt*
 - One sound, multiple letters: *running, phone*
 - One letter, multiple sounds: *tax*

1. Alphabetic systems

- Each character = one sound or articulatory gesture
- English has some exceptions:
 - Silent letters: *knee, debt*
 - One sound, multiple letters: *running, phone*
 - One letter, multiple sounds: *tax*
 - Homophones: *bank (river/finance)*

1. Alphabetic systems

- Each character = one sound or articulatory gesture
- English has some exceptions:
 - Silent letters: *knee, debt*
 - One sound, multiple letters: *running, phone*
 - One letter, multiple sounds: *tax*
 - Homophones: *bank (river/finance)*
- Examples: Latin, Greek, Cyrillic alphabets

1. Alphabetic systems (Example)

Table 1.1: The Cyrillic alphabet used for Russian.

а	б	в	г	д	е	ё	ж	з	и	й
[a]	[b]	[v]	[g]	[d]	[je]	[jo]	[ʒ]	[z]	[i]	[j]
<hr/>										
к	л	м	н	о	п	р	с	т	у	ф
[k]	[l]	[m]	[n]	[o]	[p]	[r]	[s]	[t]	[u]	[f]
<hr/>										
х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
[x]	[ts]	[tʃ]	[ʂ]	[ʃʃ]	[-]	[i]	[j]	[e]	[ju]	[ja]

- The Cyrillic alphabet is used for Russian and other nearby languages.
- Some letters resemble Latin characters, but others are unique.

- Each character = exactly one sound

International Phonetic Alphabet (IPA)

- Each character = exactly one sound
- Useful for linguists: consistent sound representation across languages

International Phonetic Alphabet (IPA)

- Each character = exactly one sound
- Useful for linguists: consistent sound representation across languages
- Different charts for (1) vowels and (2) consonants

International Phonetic Alphabet (IPA)-Vowels

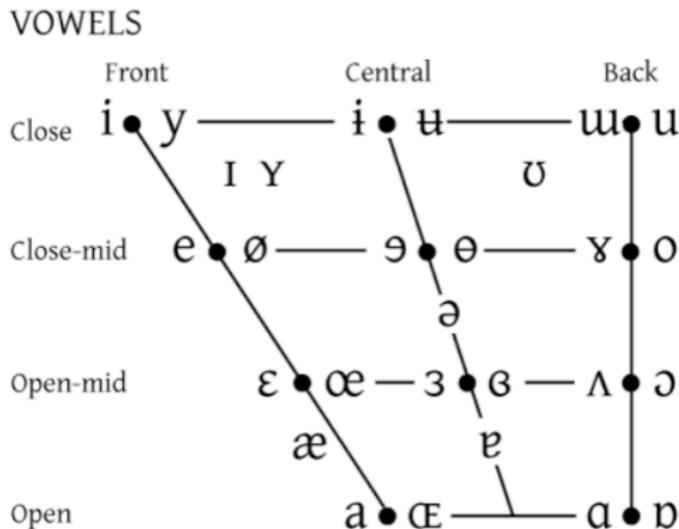


Figure 1.2: International Phonetic Alphabet of vowels (<https://commons.wikimedia.org/wiki/File:Ipa-chart-vowels.png>)

International Phonetic Alphabet (IPA)-Consonants

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill				r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

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

Figure 1.3: International Phonetic Alphabet of consonants (https://commons.wikimedia.org/wiki/Category:IPA_consonant_charts)

Figure 2: Textbook, p. 8

2. Syllabic systems

- Syllabic systems map symbols to whole syllables; larger sound units than alphabetic systems.

2. Syllabic systems

- Syllabic systems map symbols to whole syllables; larger sound units than alphabetic systems.
- Syllable: a unit of pronunciation having one vowel sound, with or without surrounding consonants.

2. Syllabic systems

- Syllabic systems map symbols to whole syllables; larger sound units than alphabetic systems.
- Syllable: a unit of pronunciation having one vowel sound, with or without surrounding consonants.
- All human languages have syllables, but syllable structure varies by languages.

2. Syllabic systems

- Syllabic systems map symbols to whole syllables; larger sound units than alphabetic systems.
- Syllable: a unit of pronunciation having one vowel sound, with or without surrounding consonants.
- All human languages have syllables, but syllable structure varies by languages.
- **Syllabary: A set of written characters representing syllables**

2. Syllabic systems

- Syllabic systems map symbols to whole syllables; larger sound units than alphabetic systems.
- Syllable: a unit of pronunciation having one vowel sound, with or without surrounding consonants.
- All human languages have syllables, but syllable structure varies by languages.
- Syllabary: A set of written characters representing syllables
- **Japanese:** simple syllables (e.g., *sashimi*, *omasake*) → few combinations → syllabaries work well.

2. Syllabic systems

- Syllabic systems map symbols to whole syllables; larger sound units than alphabetic systems.
- Syllable: a unit of pronunciation having one vowel sound, with or without surrounding consonants.
- All human languages have syllables, but syllable structure varies by languages.
- Syllabary: A set of written characters representing syllables
- **Japanese:** simple syllables (e.g., *sashimi*, *omasake*) → few combinations → syllabaries work well.
- **English:** allows complex clusters (e.g., *spark*) → many possible syllables → syllabaries become impractical.

3. Logographic systems

- Symbol is a meaning (not sound)



Figure 1.8: U.S. National Park Service symbols (pictographs).
(http://commons.wikimedia.org/wiki/File:National_Park_Service_sample_pictographs.svg)

Figure 3: p. 14

3. Logographic systems

- Symbol is a meaning (not sound)
- No pure logographic systems for human language
- Examples: icons, signage (e.g., national park symbols)



Figure 1.8: U.S. National Park Service symbols (pictographs).
(http://commons.wikimedia.org/wiki/File:National_Park_Service_sample_pictographs.svg)

Chinese Characters

- Represent syllables
- Combine logographic and phonetic elements:
“semantic-phonetic compounds”
- Over time: symbols become more abstract



Figure 4: p. 14

Logographic Writing: Semantic-Phonetic Compounds

- Chinese characters often combine:



Logographic Writing: Semantic-Phonetic Compounds

- Chinese characters often combine:
 - a **semantic element** (gives a clue to meaning)



Logographic Writing: Semantic-Phonetic Compounds

- Chinese characters often combine:
 - a **semantic element** (gives a clue to meaning)
 - a **phonetic element** (gives a clue to pronunciation)



Logographic Writing: Semantic-Phonetic Compounds

- Chinese characters often combine:
 - a **semantic element** (gives a clue to meaning)
 - a **phonetic element** (gives a clue to pronunciation)



- Example: **mā** “mother” is written with:

Logographic Writing: Semantic-Phonetic Compounds

- Chinese characters often combine:
 - a **semantic element** (gives a clue to meaning)
 - a **phonetic element** (gives a clue to pronunciation)



- Example: **mā** “mother” is written with:
 - Left side: woman – semantic component

Logographic Writing: Semantic-Phonetic Compounds

- Chinese characters often combine:
 - a **semantic element** (gives a clue to meaning)
 - a **phonetic element** (gives a clue to pronunciation)



- Example: **mā** “mother” is written with:
 - Left side: woman – semantic component
 - Right side: **mǎ** “horse” – phonetic component

Logographic Writing: Semantic-Phonetic Compounds

- Chinese characters often combine:
 - a **semantic element** (gives a clue to meaning)
 - a **phonetic element** (gives a clue to pronunciation)



- Example: **mā** “mother” is written with:
 - Left side: woman – semantic component
 - Right side: mǎ “horse” – phonetic component
- **Tone is important:**

Logographic Writing: Semantic-Phonetic Compounds

- Chinese characters often combine:
 - a **semantic element** (gives a clue to meaning)
 - a **phonetic element** (gives a clue to pronunciation)



- Example: **mā** “mother” is written with:
 - Left side: woman – semantic component
 - Right side: mǎ “horse” – phonetic component
- Tone is important:
 - **mǎ (horse) = down-up tone**

Logographic Writing: Semantic-Phonetic Compounds

- Chinese characters often combine:
 - a **semantic element** (gives a clue to meaning)
 - a **phonetic element** (gives a clue to pronunciation)



- Example: **mā** “mother” is written with:
 - Left side: woman – semantic component
 - Right side: mǎ “horse” – phonetic component
- Tone is important:
 - mǎ (horse) = down-up tone
 - **mā (mother) = high flat tone**

- Chinese: semantic + phonetic compounds (as we just discussed in the previous slide)

Hybrid systems

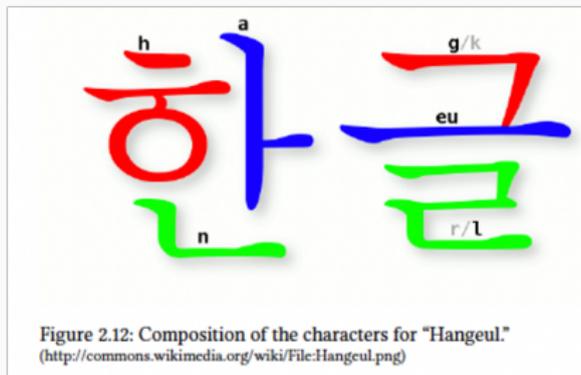
- Chinese: semantic + phonetic compounds (as we just discussed in the previous slide)
- Korean: syllable blocks built from alphabetic elements

Hybrid systems

- Chinese: semantic + phonetic compounds (as we just discussed in the previous slide)
- Korean: syllable blocks built from alphabetic elements

Hybrid systems

- Chinese: semantic + phonetic compounds (as we just discussed in the previous slide)
- Korean: syllable blocks built from alphabetic elements



- Diacritics? (e.g., accents, tone marks; $i \rightarrow \acute{i}, \hat{i}, \bar{i}, \check{i} / j \rightarrow \hat{j}$)

- Diacritics? (e.g., accents, tone marks; $i \rightarrow \acute{i}, \hat{i}, \bar{i}, \check{i} / j \rightarrow \hat{j}$)
- Word/sentence/paragraph boundaries (spaces, punctuation)

- Diacritics? (e.g., accents, tone marks; $i \rightarrow \acute{i}, \hat{i}, \bar{i}, \check{i} / j \rightarrow \hat{j}$)
- Word/sentence/paragraph boundaries (spaces, punctuation)
- Capitalization? Italics? Quotation marks?

Writing system design choices

- Diacritics? (e.g., accents, tone marks; $i \rightarrow \acute{i}, \hat{i}, \bar{i}, \check{i} / j \rightarrow \hat{j}$)
- Word/sentence/paragraph boundaries (spaces, punctuation)
- Capitalization? Italics? Quotation marks?
- Direction: Left-to-right, right-to-left, top-to-bottom

Writing system design choices

- Diacritics? (e.g., accents, tone marks; $i \rightarrow \acute{i}, \hat{i}, \bar{i}, \check{i} / j \rightarrow \hat{j}$)
- Word/sentence/paragraph boundaries (spaces, punctuation)
- Capitalization? Italics? Quotation marks?
- Direction: Left-to-right, right-to-left, top-to-bottom
- **Boustrophedon: alternating direction per line**

Digital encoding of writing

How is *writing* encoded on a computer?

- Bits = 0 or 1

How is *writing* encoded on a computer?

- Bits = 0 or 1
- Bytes = 8 bits (can represent $2^8 = 256$ values)

How is *writing* encoded on a computer?

- Bits = 0 or 1
- Bytes = 8 bits (can represent $2^8 = 256$ values)
- ASCII: 7 bits = 128 characters

How is *writing* encoded on a computer?

- Bits = 0 or 1
- Bytes = 8 bits (can represent $2^8 = 256$ values)
- ASCII: 7 bits = 128 characters
- UTF-8: variable-length encoding using 8-bit blocks

How is speech encoded on a computer?

Waveform

Representation of sound amplitude over time.

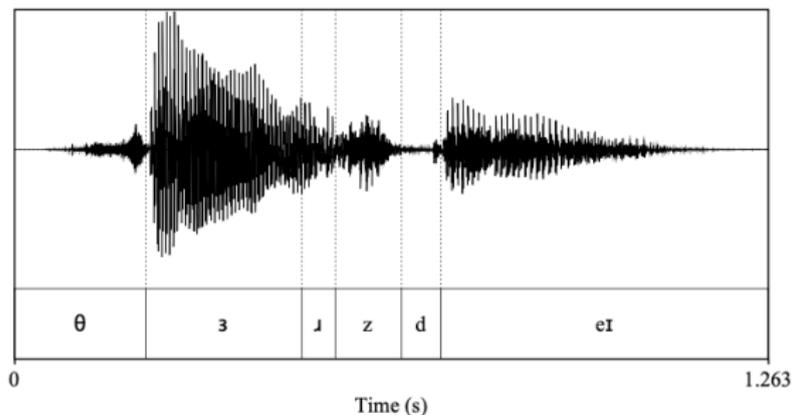


Figure 1.15: A waveform for *Thursday*.

How is speech encoded on a computer?

Spectrogram

Representation of frequency components over time (color/brightness = intensity).

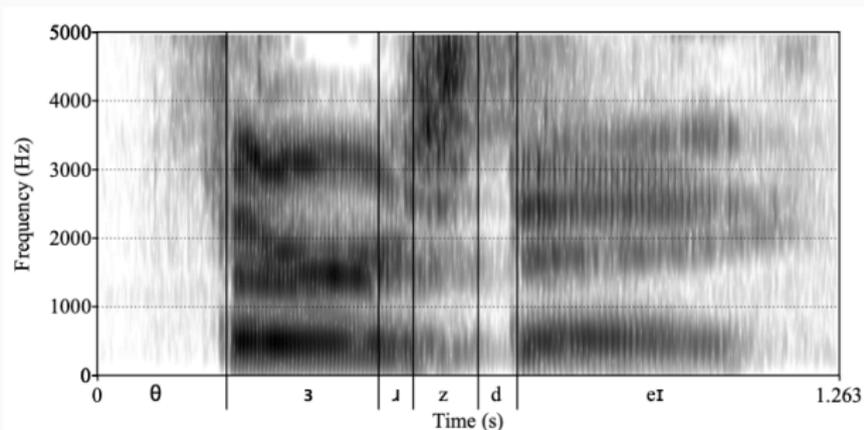


Figure 1.16: A spectrogram for *Thursday*.

- Language can be transmitted across time/space at scale

Consequences of digital writing

- Language can be transmitted across time/space at scale
- Humans understand language qualitatively

Consequences of digital writing

- Language can be transmitted across time/space at scale
- Humans understand language qualitatively
- Computers process it quantitatively (bits, bytes)

Consequences of digital writing

- Language can be transmitted across time/space at scale
- Humans understand language qualitatively
- Computers process it quantitatively (bits, bytes)
- **Question:** If computers only see numbers, how might they ever approximate meaning?

Consequences of digital writing

- Language can be transmitted across time/space at scale
- Humans understand language qualitatively
- Computers process it quantitatively (bits, bytes)
- **Question:** If computers only see numbers, how might they ever approximate meaning?
- We'll come back to this topic when learning about the **word vectors**.

Wrap-up

- Can different languages share the same writing system?

- Can different languages share the same writing system?
- Yes — e.g., English and Spanish both use the Latin alphabet.

- Can different languages share the same writing system?
- Yes — e.g., English and Spanish both use the Latin alphabet.
- Writing systems can be categorized as:

- Can different languages share the same writing system?
- Yes — e.g., English and Spanish both use the Latin alphabet.
- Writing systems can be categorized as:
 - *Alphabetic systems*

- Can different languages share the same writing system?
- Yes — e.g., English and Spanish both use the Latin alphabet.
- Writing systems can be categorized as:
 - Alphabetic systems
 - Syllabic systems

- Can different languages share the same writing system?
- Yes — e.g., English and Spanish both use the Latin alphabet.
- Writing systems can be categorized as:
 - Alphabetic systems
 - Syllabic systems
 - Logographic systems