

Morphology

Dereje T. Abzaw

Recap...

- Basic ground for NLP, introduction, definitions
- Challenges
- Scope and Coverage
- Levels of Languages
- Methods and Resources

Outline

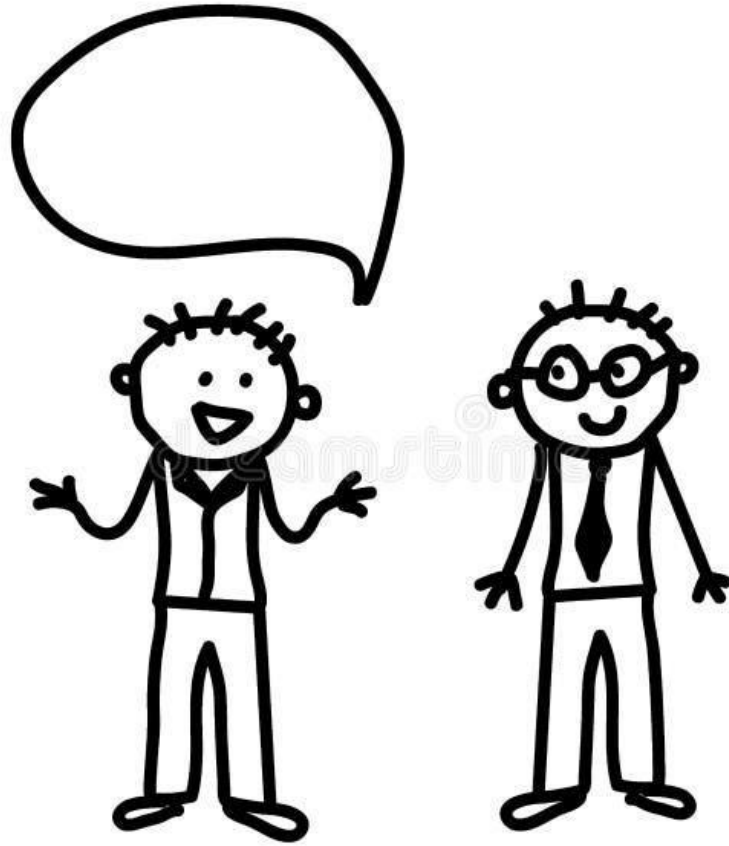
- Language
- Communication
- Natural Language
- Text Processing
- Morphology and Morphemes

Language

- What is a **language**?
 - A language is a system of communication used by a particular community or country, consisting of written **symbols**, **sounds**, **gestures**, or a combination of these, organized by **rules** for the purposes of **communication**.
- 7000+ languages
- The number, however, is constantly changing due to factors such as **language death**.
- Languages are not evenly distributed across the globe
 - Around 2000 is from Africa
 - 80+ Languages

Communication?

- Communication is the process of sharing information, thoughts, feelings, and messages between individuals or groups through various channels and mediums.
- It can occur
 - Verbally,
 - Non-verbally - Textual
 - Gesture



	hoy		läwe		hāwt		may		śāwt		re's		sat		š		qaf		qaf
ሀ	hā	ለ	lä	ሐ	hā	መ	mā	ሠ	śā	ረ	rā	ሰ	sā	ሸ	šā	ቀ	qā	ቂ	quā
ሁ	hu	ሉ	lu	ሐ	hu	ሙ	mu	ሡ	śu	ሩ	ru	ሱ	su	ሹ	śu	ቁ	qu		
ሂ	hi	ሊ	li	ሐ	hi	ሚ	mi	ሢ	śi	ሪ	ri	ሲ	si	ሺ	ši	ቂ	qi	ቃ	qui
ሃ	ha	ላ	la	ሐ	ha	ማ	ma	ሣ	śa	ራ	ra	ሳ	sa	ሻ	ša	ቃ	qa	ቄ	qua
ሄ	he	ሌ	le	ሐ	he	ሜ	me	ሤ	śe	ራ	re	ሴ	se	ሼ	še	ቄ	qe	ቅ	que
ሀ	ha/ø	ል	la/ø	ሐ	ha/ø	ም	ma/ø	ሥ	śa/ø	ሮ	ra/ø	ሰ	sa/ø	ሸ	ša/ø	ቅ	qa/ø	ቃ	qua
ሀ	ho	ሎ	lo	ሐ	ho	ሞ	mo	ሦ	śo	ሮ	ro	ሶ	so	ሾ	šo	ቆ	qo		

	bet		täwe		č		härm		nāhas		ñ		'älf		kaf				
በ	bä	ተ	tä	ቸ	čä	ኀ	hā	ኡ	huā	ነ	nā	ኘ	ñä	አ	'ä	ከ	kā	ኰ	kuā
ቡ	bu	ቱ	tu	ቸ	ču	ኀ	hu			ኑ	nu	ኙ	ñu	ኡ	'u	ኩ	ku		
ቢ	bi	ቲ	ti	ቸ	či	ኀ	hi	ኡ	hui	ኒ	ni	ኘ	ñi	ኢ	'i	ኪ	ki	ኰ	kui
ባ	ba	ታ	ta	ቸ	ča	ኃ	ha	ኣ	hua	ና	na	ኘ	ña	ኣ	'a	ካ	ka	ካ	kua
ቤ	be	ቲ	te	ቸ	če	ኄ	he	ኣ	hue	ኔ	ne	ኘ	ñe	ኤ	'e	ኬ	ke	ኮ	kue
ብ	ba/ø	ት	ta/ø	ቸ	ča/ø	ኀ	ha/ø	ኡ	hua	ን	na/ø	ኘ	ña/ø	እ	'a/ø	ከ	ka/ø	ኰ	kua
ቦ	bo	ቶ	to	ቸ	čo	ኀ	ho			ኖ	no	ኘ	ño	አ	'o	ኮ	ko		



Making Sense of it all - Natural Language + Text Processing

We have seen the basics so far

- Language
- Communication
- What is Natural Language then?
 - Natural language refers to any language that has developed naturally in humans through use and social interaction, without conscious planning or premeditation.
 - It **contrasts** with **artificial** or **constructed** languages , which are deliberately invented and structured by individuals or groups for specific purposes.

Text Processing

- Text processing refers to the automated method of manipulating or analyzing a text data to obtain useful information, perform specific tasks or solve problems.
- Involves a wide range of techniques and technologies
 - Searching Words
 - Replacing Words
 - Natural language processing (NLP)
 - Text mining
 - Sentiment Analysis
- The goal of text processing is often to transform **unstructured text** into a **structured form** that can be further analyzed or processed by computers.

Contd ...

- **Text Editing:** Basic text processing involves editing and formatting text, which can be done using text editors or word processors.
 - adding,
 - deleting,
 - finding, and
 - replacing text.
- **Data Cleaning and Preparation:** Involves removing irrelevant, redundant, or erroneous data from text, as well as standardizing and formatting text data in preparation for further analysis.
- **Information Extraction:** The process of automatically extracting structured information from unstructured text, such as entities (names, places, dates), relationships between entities, and specific facts or events.
- **Text Mining:** Involves extracting high-quality information from text.
 - Pattern recognition,
 - Sentiment analysis,
- **Natural Language Processing (NLP):** A more advanced form of text processing that focuses on enabling computers to understand, interpret, and generate human language in a way that is both valuable and meaningful.
- **Sentiment Analysis:** Identifies the sentiment expressed in a piece of text, determining whether it is positive, negative, or neutral, and to what degree.

Morphology

- The term morphology is Greek and is a makeup of
 - **“Morph”** - meaning 'shape, form , structure',
 - **“Ology”** - which means ‘the study of something’.
- **Morphology** is the branch of linguistics concerned with the **study of the structure and form of words** in a language.
- It deals with how words are formed, their internal structures, and how they relate to other words in the same language.
- Morphology examines the smallest units of meaning or grammatical function in a language, known as **morphemes**, and how these units combine to form words.

Word Counts?

- How many words are in the English Language?
 - a. Oxford English Dictionary (OED): The OED lists over **600,000** words, including words currently in use, as well as those that are obsolete but have historical significance.
 - b. Merriam-Webster's Third New International Dictionary: Contains about **470,000** entries.
 - c. Collins English Dictionary: Features over **722,000** words, meanings, and phrases.
- It's important to note that these figures include not just "**root**" words but also **derived forms**, **technical** and **scientific**, **slang**, and **regional** variations.
- Additionally, the English language is constantly evolving, with new words being added to dictionaries every year due to technological advancements, cultural changes, and other influences.

Contd ...

- Which of the following words are you familiar with? (Merriam, Dictionary.com)
 - “Rizz”
 - “Barbiecore”
 - “Sustainable Fashion”
 - “Generative AI”

Morphemes

1. **Free morphemes:** These are morphemes that can stand alone as words with their own meaning.
 - **For example,** "book" and "run" are free morphemes because they have meaning on their own and can function as independent words.
2. **Bound morphemes:** These are morphemes that cannot stand alone and must be attached to other morphemes to convey meaning.
 - Bound morphemes include **prefixes, suffixes, infixes(?)**, and **circumfixes**.
 - **For example,** in the word "**unhappiness**", "un-" and "-ness" are bound morphemes attached to the free morpheme "happy".

★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★

**IT'S GOING TO BE
LEGEND**

★ ★ **WAIT FOR IT...** ★ ★

DARY

★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ 



Contd ...

- true, truer, truest, truly, untrue, truth, truthful, truthfully, untruthfully, untruthfulness.
- Untruthfulness → un+ -true+ -th+ -ful+ -ness (Productive morphemes)
- Unhappiness → un+ -happi+ -ness → un+ -happy+ -ness
 - “un” is a negative marker
 - “ness” expresses a state or quality
- Horses → horse+ -s → (“s” shows plurality)
- Walking → walk+ -ing → (“ing” represent a sense of duration for walk)

Switching things to local

- አልጠየኩልህምና
- ስለህንፃዎቹ
- በወዳጆቻቸው