University of Science

Computational Linguistics Center

Introduction to Natural Language Processing

Section 1:
Introduction to Natural Languages

Lecturer: Assoc.Prof. Dr. Dinh Dien

LANGUAGES IN THE WORLD

- Differ: Natural Languages (e.g. Vietnamese, English, French, etc.) vs. Artificial Languages (e.g. C, Pascal,…; Morse; Braille; etc.
- From now: language = natural language.
- How many different (natural) languages are there in Vietnam ?
- ~ 54 (Vietnamese and 53 ethnic languages)
- How many different languages are there in the world ?
- ~ 7015 !
- Distribution of users: very unequally (100M vs. <100)

LANGUAGE POPULATION

Rank	Language Name	Primary Country	Population
1	CHINESE, MANDARIN	China	885,000,000
2	SPANISH	Spain	332,000,000
3	ENGLISH	United Kingdom	322,000,000
4	BENGALI	Bangladesh	189,000,000
5	HINDI	India	182,000,000
6	PORTUGUESE	Portugal	170,000,000
7	RUSSIAN	Russia	170,000,000
8	JAPANESE	Japan	125,000,000
9	GERMAN, STANDARD	Germany	98,000,000
10	CHINESE, WU (Ngô)	China	77,175,000
11	JAVANESE	Indonesia, Java, Bali	75,500,800
12	KOREAN	Korea, South	75,000,000
13	FRENCH	France	72,000,000
14	*VIETNAMESE*	*Vietnam*	*67,662,000*
15	TELUGU	India	66,350,000
16	CHINESE, YUE (Việt)	China	66,000,000

Endangered Languages

- Vanishing Languages:
- One language dies every 14 days.
- By the next century nearly half of the roughly 7,000 languages spoken on Earth will likely disappear,
- as communities abandon native tongues in favor of English, Mandarin, or Spanish.
- What is lost when a language goes silent?
- Cultural treasures, historical lessons, mankind knowledge, etc.

THE ORIGIN OF LANGUAGES

- Who invented English?
- Who invented Vietnamese?
- Differ: voice (natural) vs. writings (manmade)
- Only popular languages have writings.
- Vietnamese writings: who invented ?
- before 10th century: has no (using Chinese writings);
- from 10th-19th century: using Nôm writings
- Most famous works (in ~ 1000 years) written in Nôm: literature (*The Tale of Kiều*) / history/culture/ Agriculture/ Geography/ Traditional Medicine/…
- Nơm writings borrow Chinese characters: one for sound, one for meaning:

e.g.: 爸(*ba/father*), 巴(*ba,3*), 罘(*4*), 靡(*year*), 靸(*5*), …

The Tale of Kiều in Nôm

𡨸 𣍐 沖 埃 得 些

Trăm năm trong cõi người ta

字 才 字 命 窖 羅 恄 饒

Chữ tài chữ mệnh khéo là ghét nhau

.....

浪 𣍐 嘉 靖 朝 明

Rằng: Năm Gia Tịnh triều Minh

罕 方 滂 朗 𠊚 京 凭 傍

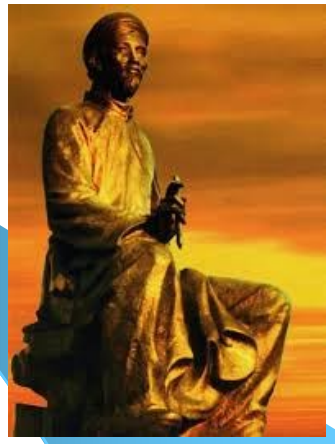
Bốn phương phẳng lặng hai kinh vững vàng

固 茹 員 外 户 王

Có nhà viên ngoại họ Vương

家 資 擬 拱 常 常 堛 中

Gia tư nghỉ cũng thường thường bậc trung



Nguyen Du
(1766-1820)

- *nghĩ (think)*
- *nghỉ (he)*

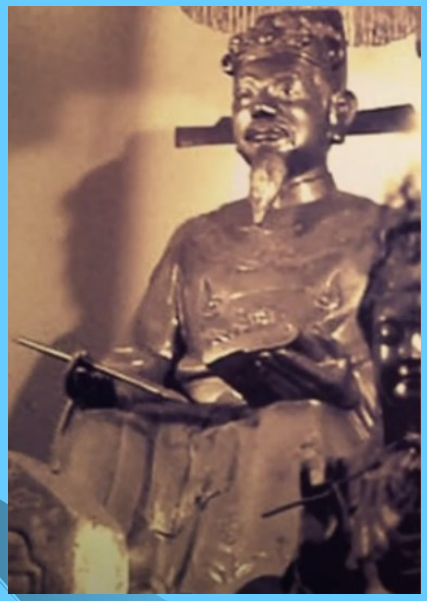
WRITINGS = artificial

- Vietnamese alphabet:
- since 19th century – till now:
- from Latin-letters + diacritics: *ba, bốn, năm,*
- Latin alphabet: from Greek (Y);
- Greek from Phoenician …

Alexandre de Rhodes (1591-1660)

Α Alpha	Β Beta	Γ Gamma	Δ Delta	Ε Epsilon	Ζ Zeta
Η Eta	Θ Theta	Ι Iota	Κ Kappa	Λ Lambda	Μ Mu
Ν Nu	Ξ Xi	Ο Omicron	Π Pi	Ρ Rho	Σ Sigma
Τ Tau	Υ Upsilon	Φ Phi	Χ Chi	Ψ Psi	Ω Omega

α Alpha	β Beta	γ Gamma	δ Delta	ε Epsilon	ζ Zeta
η Eta	θ Theta	ι Iota	κ Kappa	λ Lambda	μ Mu
ν Nu	ξ Xi	ο Omicron	π Pi	ρ Rho	σ Sigma
τ Tau	υ Upsilon	φ Phi	χ Chi	ψ Psi	ω Omega



Hàn Thuyên

13th century

WRITINGS = artificial

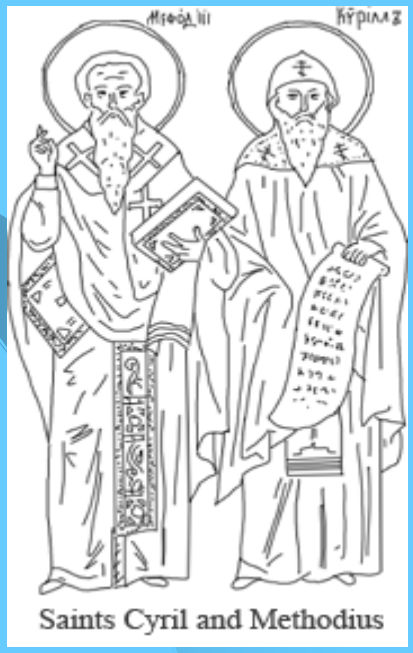
Russian alphabet: from Cyrillic; Greek

Мы учим язык *(We are learning a language)*

Москва

путин

Russian Alphabet					
А	ah	К	kah	Х	khah
Б	beh	Л	ehl	Ц	tseh
В	veh	М	ehm	Ч	chYah
Г	geh	Н	ehn	Ш	shah
Д	deh	О	o	Щ	shchyah
Е	yeh	П	peh	Ъ	tviordiy znak
Ё	yo	Р	ehr	Ы	i
Ж	zheh	С	ehs	Ь	myagkiy znak
З	zeh	Т	teh	Э	eh
И	ee	У	oo	Ю	Yoo
Й	ee kratkoye	Ф	ehf	Я	Yah



Saints Cyril and Methodius

1000 AD

WRITINGS = artificial

Korean alphabet (Hangeul):

우리는 언어를 배우고 있어요

King Sejong (1397; 1418-1450; 1443)

한국 (H-a-n k-u-k)

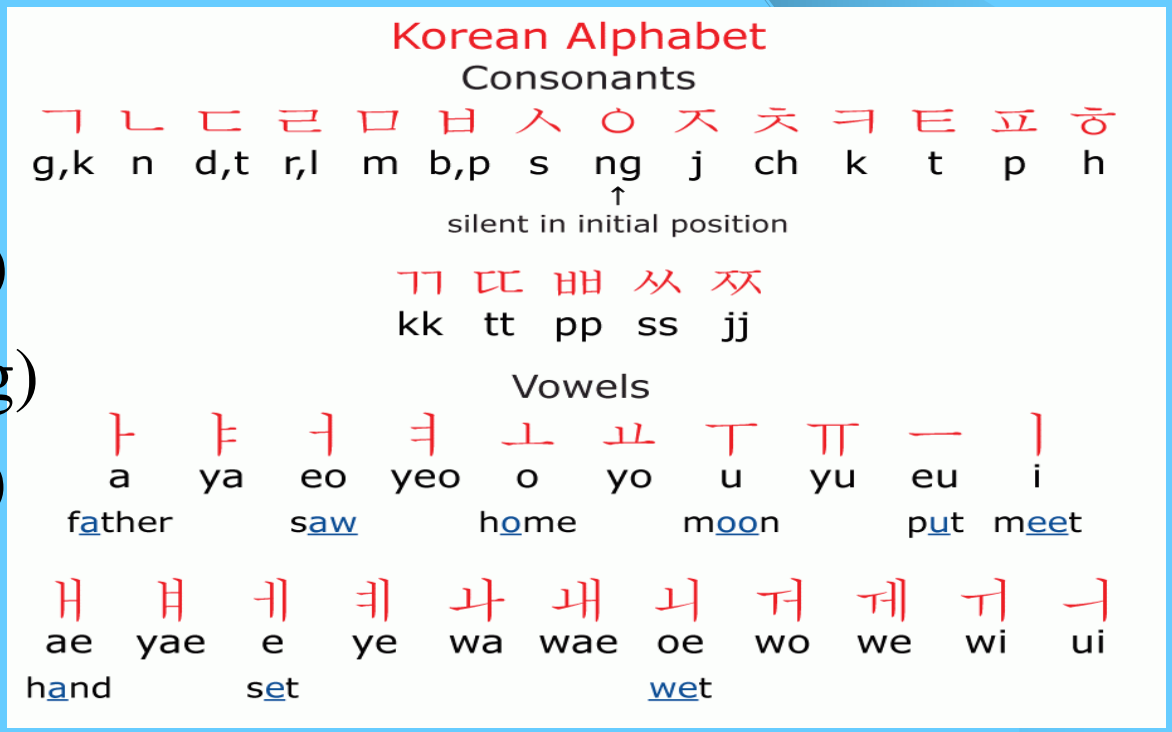
미국 (M-i k-u-k)

중국 (Tr-u-ng k-u-k)

삼성 (S-a-m S-eo-ng)

학생 (h-a-k S-ae-ng)

준비 (j-u-n b-i)

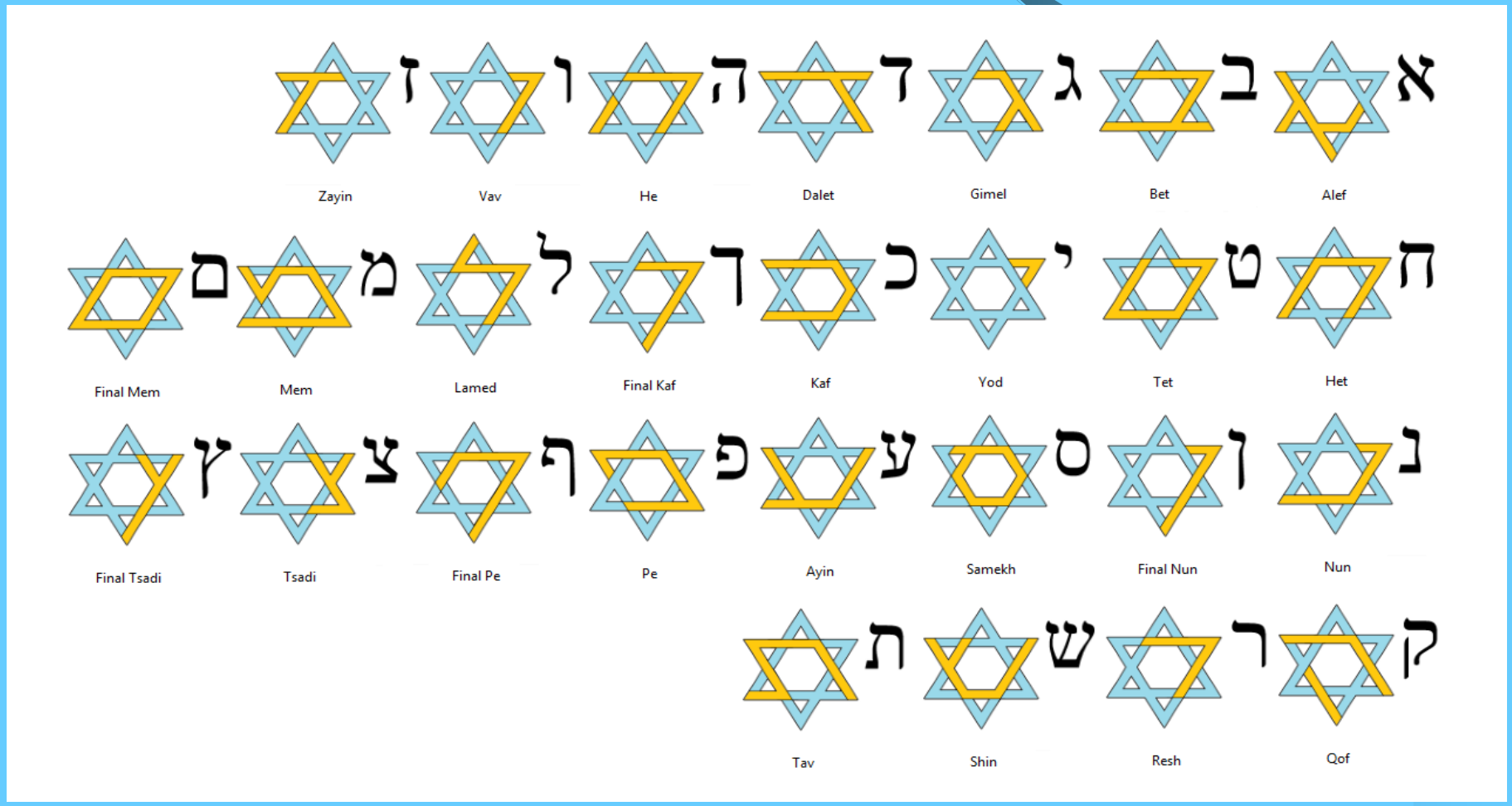


Hebrew alphabet:

(David star)

ירושלים (y-r-sh-l-y-m) =>yerushalayim

אַבָּא => aba



THE ORIGIN OF LANGUAGES

- [Genesis]: At the beginning, the whole world had one language and a common speech, settled in the same land named Shinar.
- As the population was growing, they decided to build a tall "reach to the heavens", proud symbol of how great they had made their nation.
- God did not like the pride and arrogance, God caused the people to suddenly speak different languages so they could not communicate and work together to build the tower.
- This caused the people to scatter across the land with different languages as nowadays.
- The tower was named The Tower of Babel because the word Babel means “confusion”.

The Tower of Babel

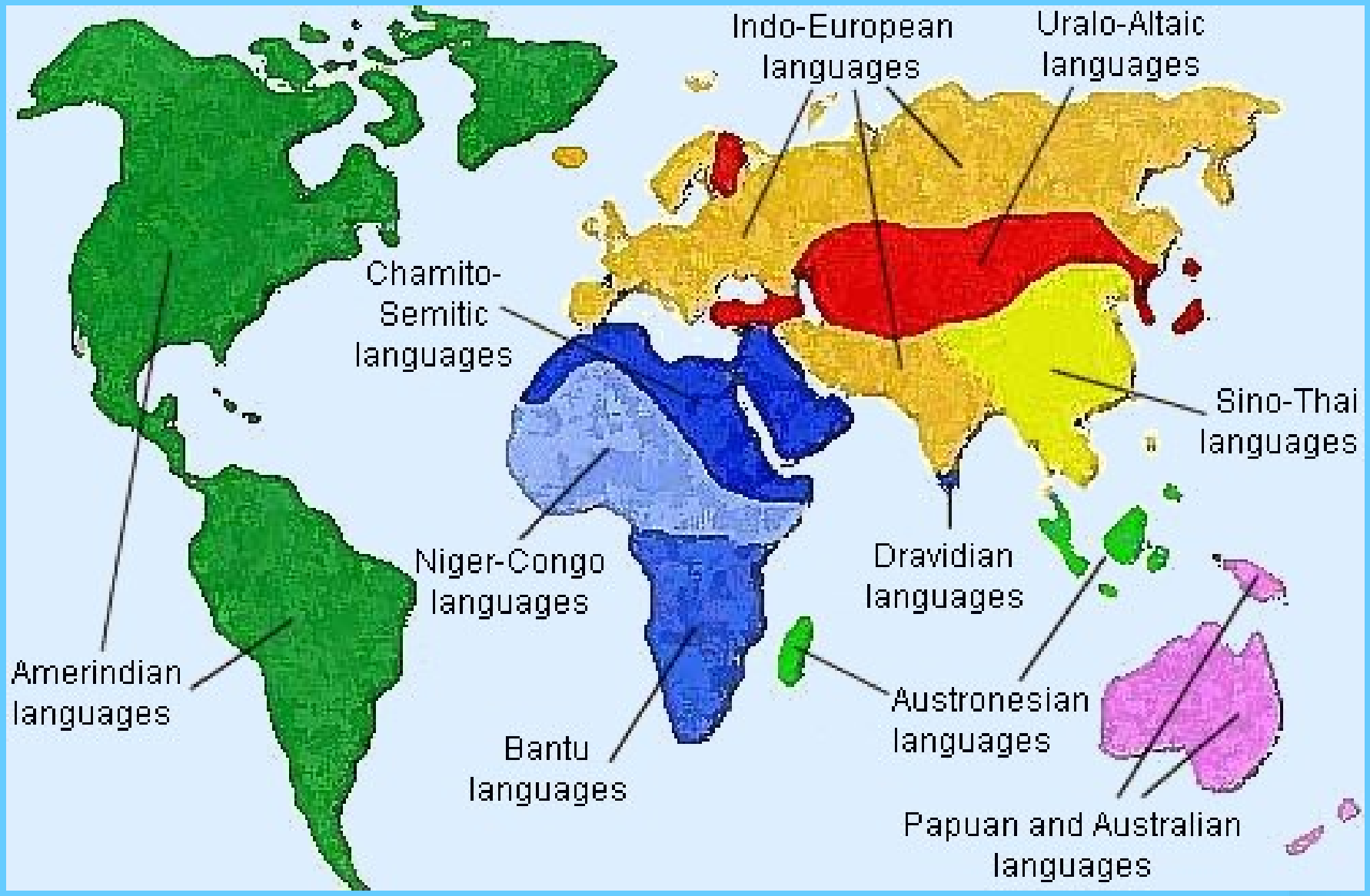
This is only a legend !

THE CLASSIFICATION OF LANGUAGES

There are 2 main kinds of **classification** of **languages**:
(1) genetic (or genealogical) and (2) typological.
The purpose of genetic **classification** is to group **languages** into families according to their degree of diachronic relatedness
There are 5 genres of languages:

1. Indo-Euro: India, Iran, Bantic, Slave, Roman, Greek, German (German, ***English,*** Dutch,...).
2. Semitic: Semit, Egypt, Kusit, Becbe,...
3. Turkish: Turkish, Tatar, Uyghur,…
4. Sino-Tibetan: Sino (Chinese), Tibeto-Burman,...
5. Austro-Asia: Nahali, Munda, Nicoba, Mon-Khmer. Mon-Khmer branch: Viet-Muong group; Viet-Muong group: Muong and ***Vietnamese***.

THE CLASSIFICATION OF LANGUAGES: GENRE



Typological Classification of Languages

Definition

- Languages are described by their *types* rather than by their origins and relationships
- The type under which languages are classified follows morphological classification (changing the form; inflected by tense, case, number, gender, mood, etc.)

Language Typology

1. Inflecting/flectional/fusional
2. Isolating
3. Agglutinating/agglutinative
4. Polysynthetic/incorporating

Flexional/Fusional/Inflecting Languages

- Grammatical devices like *affixes* or internal changes in words to show grammatical relationships (tense, case, person, number, gender, mood, form, …)
- Ex. book:n, book-s, mouse-mice
- walk:v, walk-s, walk-ing, walk-ed
- He (sub) – him (obj) , I (sub) –me (obj)
- E.g.: *I see him* vs. *He sees me.*
- I walk; He walks; I walked; I am walking; …

- Inflections of French:

When the verb ALLER is conjugated, it looks like this:

Je vais – I go, I am going
Tu vas – you go, you are going
Il va – he goes, he is going
Elle va – she goes, she is going
Nous allons – we go, we are going
Vous allez – you go, you are going
Ils vont – they go, they are going
Elles vont – they go, they are going

Le verbe être

Person	Verb	Translation
Je	**suis**	*I am*
Tu	**es**	*You are*
Il/Elle	**est**	*He/She is*
Nous	**sommes**	*We are*
Vous	**êtes**	*You are*
Ils/Elles	**sont**	*They are*

Il y a un petit livre dans la petite maison.
(There is a small book in the small house)

Flexional Languages

Infinitive	читать to read
я - I	чита-ю
ты - you	чита-ешь
он, она - he,she	чита-ет
мы - we	чита-ем
вы - you	чита-ете
они - they	чита-ют

I read a book => Je lis un livre => я читаю книгу.
This is my book => C'est mon livre => Это моя книга.
This book is interesting => Ce livre est intéressant =>
Эта книга интересная.
These books are interesting=> Ces livres sont intéressants
Эти книги интересны.

- Inflections of Russian:

	1st person	2nd person	3rd person (masc.)	3rd person (fem.)	3rd person (neut.).
English	*I, Me*	*You*	*He, Him*	*She, Her*	*It*
Nominative Case	Я	Ты	Он	Она	Оно
Accusative Case	Меня	Тебя	Его	Её	Его
Genitive Case	Меня	Тебя	Его	Её	Его
Dative Case	Мне	Тебе	Ему	Ей	Ему
Instrumental Case	Мной	Тобой	Им	Ей	Им
Prepositional Case	Мне	Тебе	Нём	Ней	Нём

- Inflections of Latin:

1st and 2nd Declension (-ā and –o stem) Adjectives				
		bonus, bona, bonum, *good*		
		STEM **bono-** (M.)	STEM **bonā-** (F.)	STEM **bono-** (N.)
SING.	NOM.	bon**us**	bon**a**	bon**um**
	GEN.	bon**ī**	bon**ae**	bon**ī**
	DAT.	bon**ō**	bon**ae**	bon**ō**
	ACC.	bon**um**	bon**am**	bon**um**
	ABL.	bon**ō**	bon**ā**	bon**ō**
	VOC.	bon**e**	bon**a**	bon**um**
PLUR.	NOM.	bon**ī**	bon**ae**	bon**a**
	GEN.	bon**ōrum**	bon**ārum**	bon**ōrum**
	DAT.	bon**īs**	bon**īs**	bon**īs**
	ACC.	bon**ōs**	bon**ās**	bon**a**
	ABL.	bon**īs**	bon**īs**	bon**īs**

Magister discipulos amat.
Discipuli magistrum amant.

Iūppiter est deus et in Olympō habitat. Terram spectat et puellam Eurōpam videt. Eurōpa pulchra est et Iūppiter puellam dēsīderat. Iūppiter sē in taurum trānsfōrmat quod Eurōpa est timida.

Eurōpa taurum spectat et taurus puellam portat. Nunc puella nōn est timida. Taurus fugitat et Eurōpam ad insulam Crētam portat. Deus et puella in insulā habitant.

Isolating languages

- It is an unalterable unit whose function in the sentence is not usually marked by some grammatical device (affix, auxiliary) but only by position.
- Since the boundaries of syllables and morphemes *coincide*, these languages are sometimes referred to as monosyllabic.

Isolating Languages

- Examples: Chinese, Vietnamese, Thai, Laos, and many languages of South East Asia
- Ex (Chinese):我看他 *wo kan ta*

  “I see him”; “I am seeing him”

  他看我朋友 *Ta kan wo peng you*

  “He sees my friend”

Agglutinating/Agglutinative Languages

- A type of flexional language with the exception that the morphemes attached have a separate existence (= free morpheme)
- Implication: the boundaries between the morphemes are always clear because their shape remains the same

Agglutinating languages

Examples: Japanese, Korean, Turkish,...

VERB CONJUGATION - る

	present	past
affirmative	たべる たべます	たべた たべました
negative	たべない たべません	たべなかった たべませんでした

*informal
*formal

MajideJapanese.com

マジで majide

Conjugating 食べる: 一段 (type II) verb. (Supplementary Comments)

	Affirmative		Negative	
	Plain/Informal	Polite/Formal	Plain/Informal	Polite/Formal
Non-past	食べる	食べます	食べない 食べぬ(x) 食べず(に)(x)	食べません
Past	食べた	食べました	食べなかった	食べませんでした
Te-form	食べて	食べまして	食べなくて 食べないで	食べませんで
Conditional	食べたら	食べましたら	食べなかったら	食べませんでしたら
Provisional	食べれば	食べますなら(ば)	食べなければ	食べませんなら(ば)
Potential(#) & Passive	食べられる	食べられます	食べられない	食べられません
Causative	食べさせる 食べさす	食べさせます 食べさします	食べさせない 食べささない	食べさせません 食べさしません
Caus-Pass	食べさせられる	食べさせられます	食べさせられない	食べさせられません
Volitional/ Hortative	食べよう 食べる[よう/こと]にしよう	食べましょう 食べる[よう/こと]にしましょう	食べまい(+) 食べない[よう/こと]にしよう	食べますまい 食べない[よう/こと]にしましょう
Conjectural	食べるだろう	食べるでしょう	食べないだろう	食べないでしょう
Alternative	食べたり	食べましたり	食べなかったり	食べませんでしたり
Imperative	食べろ	食べなさい	食べるな	食べなさるな

Example:

Turkish

ev → house (nom. sg.)

ev-ler → houses (nom. pl.)

ev-i → his/her house (sg.+poss.)

ev-ler-i → his/her houses (pl.+poss.)

ev-den → in front of the house (sg.+abl.)

ev-ler-den → in front of the houses (pl.+abl.)

Japanese

私は本を読みます: I read a book

私は本を読みました : I read(pst) a book

私は本を読みません: I do not read a book

私は本を読みませんでした : I did not read a book

Ex: tabesaserareru

- *tabe* “eat” (the base)
- *sase* “the causative element (i.e. to cause someone to do something)
- *rare* “the passive form”
- *ru* “the infinitive”

Polysynthetic/Incorporating Languages

- These languages make use of affixation and often incorporate what English would represent with nouns and adverbs.
- The word forms are often very long and morphologically complex
- Languages: Inuktitut (Baffin Island Eskimo), Oneida)

Word Order Typology: syntax

Ví dụ: I eat rice Tôi ăn cơm

S V O S V O

1. SVO: 32.4 - 41.8 %, e.g. **English**, Chinese, French, **Vietnamese**, Thai, Bulgaria, ...
2. SOV: 41 - 51.8 %, e.g. Japanee, Korean, Mongolian, Turki, Eskimo,...
3. VOS: 9 - 18 %, e.g. Cakchiquel (Guatemala), Huave (Mexico),...
4. VSO: 2 - 3 %, e.g. Tagalo, Egypt(old), Hebrew (Bible), Ireland,...
5. OVS: 1 % , e.g. Apalai (Brazil), Barasano (Columbia), Panare (Venezula),..
6. OSV: 1 %, e.g. Apurina, Xavante (Brazil),..

word order

- I read a book
- 나는 책을 읽고;
- 私は本を読みます

MultiLingual Parallel Corpus

Chúng ta học một ngôn ngữ.

We are learning a language.

Nous apprenons une langue.

我们 学习 一门 语言。

言語を 習います。

우리는 언어를 배우고 있어요.

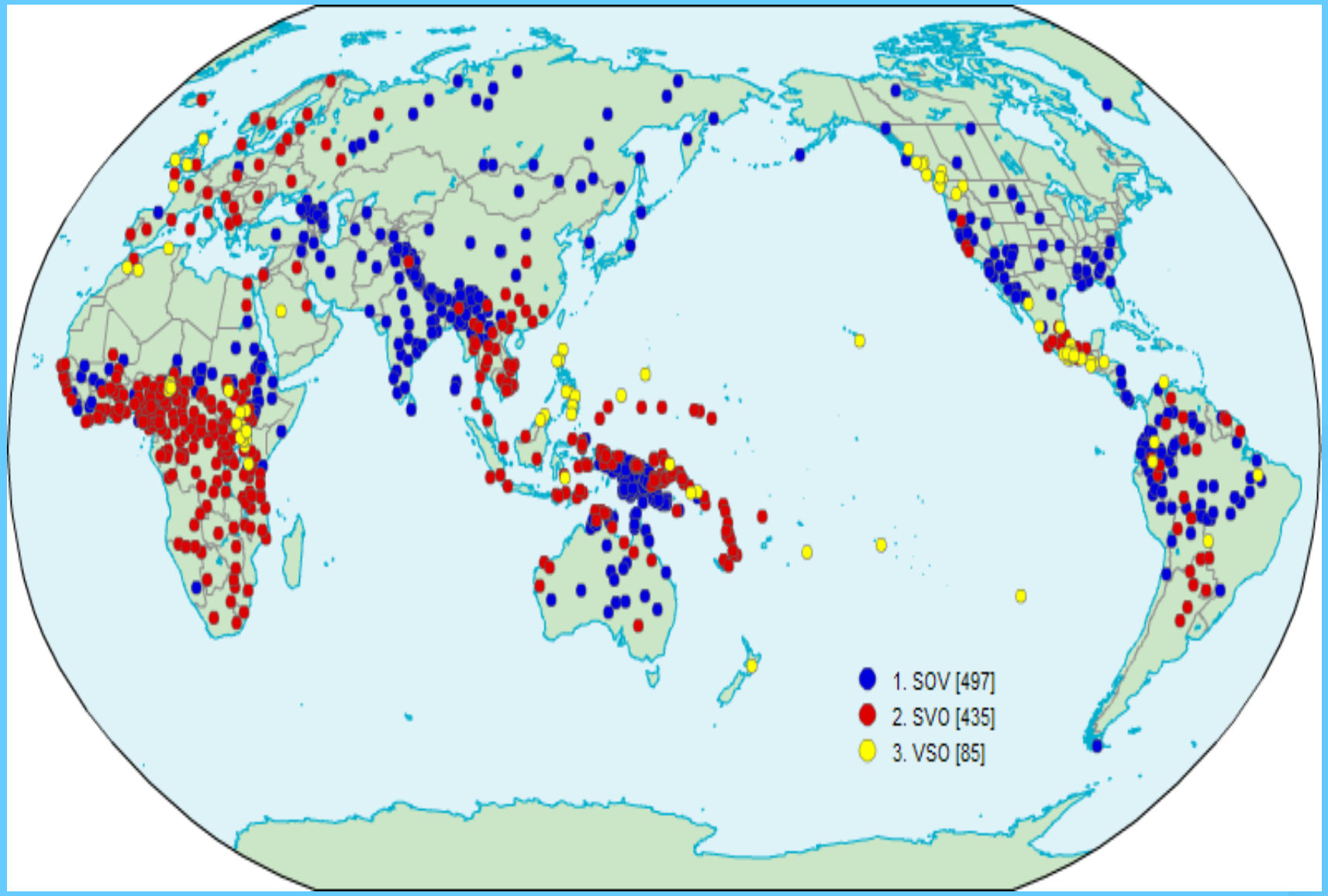
Wir lernen eine Sprache.

Мы учим язык.

Ni lernas lingvon.

(https://www.50languages.com/)

Word Order typology map



Characteristics of Natural Languages

- NL is a social phenomenon, not a natural phenomenon, or personal or biology (not hereditary).
- NL is the most important means of communication between humans.
- NL is the special semiotic system: differ: “signifier” (sound/image) vs. “signified”(concept).
- Ex: in the traffic signal system: “red light” (signifier) => stoppage (signified).
- Ferdinand de Saussure: “NL seems like a chess-board”. The value of each chessman is regulated by the system of the chess-board.
- => The meaning of a word is dependent on the context.

THE SYSTEM OF NATURAL LANGUAGES

NL consists of following linguistic units:

1. Phoneme: the smallest unit of voice.
2. Morpheme: the smallest unit carrying the meaning
3. Word: the free morpheme.
4. Phrase: many words
5. Sentence: at least 1 clause (SP-VP)
6. Text: system of sentences

THE SYSTEM OF NATURAL LANGUAGES

Other linguistic units:

1. Letter/alphabet:
2. Character: letter/alphabet, digit/number, symbol, ^char
3. Syllable:
4. Morpho-syllable: chữ/tự

THE SYSTEM OF NATURAL LANGUAGES

Has following aspects:

1. Phonetics, phonology: sound of linguistic units (LU): (voice/speech)
2. Morphology: the form of LUs
3. Grammar, syntax: the relationship of LUs
4. Semantics: the content (meaning) of LUs
5. Pragmatics: the purpose/usage of LUs

Writing systems

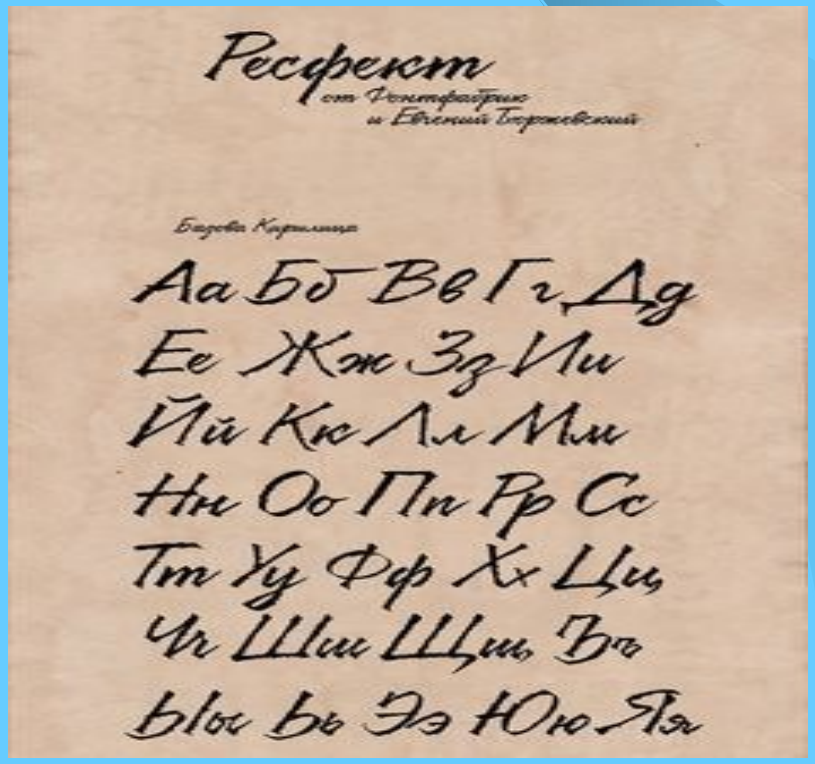
1. Alphabet (phoneme): Latin, Greek, Cyrillic
2. Abjad (alphabet without vowels): Hebrew, Arabic
3. Abugida (alphabet with vowels as features): Hangeul, Thai, Lao
4. Syllabary: Hira, Kata
5. Logography (meaning): Chinese

Α	Β	Γ	Δ	Ε	Ζ
Alpha	Beta	Gamma	Delta	Epsilon	Zeta
Η	Θ	Ι	Κ	Λ	Μ
Eta	Theta	Iota	Kappa	Lambda	Mu
Ν	Ξ	Ο	Π	Ρ	Σ
Nu	Xi	Omicron	Pi	Rho	Sigma
Τ	Υ	Φ	Χ	Ψ	Ω
Tau	Upsilon	Phi	Chi	Psi	Omega

α	β	γ	δ	ε	ζ
Alpha	Beta	Gamma	Delta	Epsilon	Zeta
η	θ	ι	κ	λ	μ
Eta	Theta	Iota	Kappa	Lambda	Mu
ν	ξ	ο	π	ρ	σ
Nu	Xi	Omicron	Pi	Rho	Sigma
τ	υ	φ	χ	ψ	ω
Tau	Upsilon	Phi	Chi	Psi	Omega

word "лишишь"

А	Б	В	Г	Д	Е
Ё	Ж	З	И	Й	К
Л	М	Н	О	П	Р
С	Т	У	Ф	Х	Ц
Ч	Ш	Щ	Ъ	Ы	Ь
Э	Ю	Я			



ALPHABET

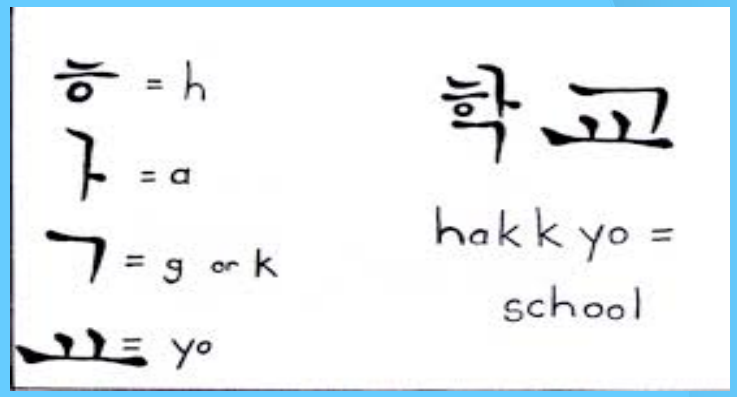
Japanese (Hiragana/Katakana):

あ a	か ka	さ sa	た ta	な na	は ha	ま ma	や ya	ら ra	わ wa		が ga	ざ za	だ da	ば ba	ぱ pa
い i	き ki	し shi	ち chi	に ni	ひ hi	み mi		り ri			ぎ gi	じ ji	ぢ ji	び bi	ぴ pi
う u	く ku	す su	つ tsu	ぬ nu	ふ fu	む mu	ゆ yu	る ru			ぐ gu	ず zu	づ zu	ぶ bu	ぷ pu
え e	け ke	せ se	て te	ね ne	へ he	め me		れ re			げ ge	ぜ ze	で de	べ be	ぺ pe
お o	こ ko	そ so	と to	の no	ほ ho	も mo	よ yo	ろ ro	を o	ん n	ご go	ぞ zo	ど do	ぼ bo	ぽ po

きゃ kya	しゃ sha	ちゃ cha	にゃ nya	ひゃ hya	みゃ mya	りゃ rya	ぎゃ gya	じゃ ja	ぢゃ ja	びゃ bya	ぴゃ pya
きゅ kyu	しゅ shu	ちゅ chu	にゅ nyu	ひゅ hyu	みゅ myu	りゅ ryu	ぎゅ gyu	じゅ ju	ぢゅ ju	びゅ byu	ぴゅ pyu
きょ kyo	しょ sho	ちょ cho	にょ nyo	ひょ hyo	みょ myo	りょ ryo	ぎょ gyo	じょ jo	ぢょ jo	びょ byo	ぴょ pyo

ALPHABET

Hangeul:



WRITING CHART

Script	Print	Name	Letter
	ם	Final Mem	ם
	נ	Nun	נ
	ן	Final Nun	ן
	ס	Samech	ס
	ע	Ayin	ע
	פ	Pay	פ
	פ	Fay	פ
	ף	Final Fay	ף
	צ	Tsadee	צ
	ץ	Final Tsadee	ץ
	ק	Koof	ק
	ר	Resh	ר
	שׁ	Shin	שׁ
	שׂ	Sin	שׂ
	ת	Tav	ת
	ת	Tav	ת

Script	Print	Name	Letter
	א	Alef	א
	בּ	Bet	ב
	ב	Vet	ב
	ג	Gimmel	ג
	ד	Dalet	ד
	ה	Hay	ה
	ו	Vav	ו
	ז	Zayin	ז
	ח	Ḥet	ח
	ט	Tet	ט
	י	Yud	י
	כּ	Kaf	כ
	כ	Chaf	כ
	ך	Final Chaf	ך
	ל	Lamed	ל
	מ	Mem	מ

JERUSALEM ירושלם

יְרוּשָׁלַיִם

Yerusha**la**yim

Examples

γραφικ,

график

그래픽

グラフィック

גרפי

Vietnamese characteristics

- Vietnamse is the isolated language typology.
- Vietnamese words have no inflections. The grammatical meaning is outside the word. Ex: *Tôi nhìn anh ấy* vs. *Anh ấy nhìn tôi (I see him* vs. *He sees me)* .
- Grammatical methods are: *word order* and *function words*. Ex*: Gạo xay vs. Xay gạo ; đang học* vs. *học rồi (learning* vs. *learned).*
- There are a special linguistic unit: morpho-syllable ("hình tiết") whose its phonetic-cover exactly coincides with its syllable (âm tiết), and morpheme (hình vị) aka "tiếng".

CHARACTERISTICS OF VIETNAMESE LANGUAGE

- The word boundary is ambiguous (not delimited by space as flexional typology languages). Ex: “học sinh học sinh học” (pupils learn biology).
- => Morphological analysis becomes difficult.
- Word Segmentation is the pre-requisite for next modules, e.g.: spelling checker, POS tagger, word frequency, …
- There is a special classifier which go accompanied with nouns, e.g. : *cái* bàn, *cuốn* sách, *bức* thư, *con* chó, *con* sông, *vì* sao, ...(same phenomena in Chinese).

CHARACTERISTICS OF VIETNAMESE LANGUAGE

- In the phonetics aspect, Vietnamese is the tone language. Each syllable carries 1 of 6 following tones: no mark (ngang); acute (sắc), breve (huyền), question mark (hỏi), tilde (ngã) and dot below (nặng).
- This is suppra-segmental phoneme (âm vị siêu đoạn tính).
- Reduplicative words: *lấp lánh, lung linh, ..*
- Spoonerism (nói lái): by exchanging the initial consonant and the nucleus and/or the tone-mark between 2 syllables within a word due to their loose links, e.g. *hiện đại -> hại điện, thầy giáo -> tháo giầy,…*

CHARACTERISTICS OF ENGLISH LANGUAGE

- English is the flexional language typology with following characteristics:
- In the running texts, the word will be inflected.
- The grammatical meaning is inside the word.
- E.g.: *I see him* vs. *He sees me.*
- Grammatical methods: suffix. Ex: *learning* vs. *learned.*
- Word formations: affix. Ex: anticomputerizational (anti-compute-er-ize-ation-al).
- The morpheme boundary is ambiguous.
- The word boundary is clear (delimited by space or punctuation marks).

ENGLISH – VIETNAMESE COMPARISON

- Due to language/cultural typology (English-VNese comparative/contrastive linguistics)=> many differences.
- E.g.: in phonetics: English (no tone), Vietnamese (tone)
- Word boundary; lexicalization: e.g.: ox – bò đực, anh – elder brother , “carry out” -> “thực hiện”;…;
- Part-Of-Speech: “thank you for your attention/N” (“cám ơn các bạn đã lắng nghe/V”)
- Word order: “head-initial” vs. “head-final”; “pre-position” vs, “post-position”. Ví dụ: “A pretty new green dress” vs. “một cái áo dài mới đẹp màu xanh”;

⇒ “Didn’t we learn this lesson *yesterday*? *Hôm qua*, mình đã không học bài này sao ?”; “They went home *quietly*.” “Họ *lặng lẽ* về nhà.” or “Họ về nhà *một cách lặng lẽ*” (Adverb positions: Manner-Place-Time).

Vietnamese NLP

Đinh Điền

Dinh Dien

딘 디엔

丁田

Динх Диэн

ディンディエン

Language	Typology	Word-Order	Writing system	
Vietnamese	Isolating	S-V-O	Alphabet	Latin++
Chinese	Isolating	S-V-O	LogoSyllabic	
English	Flexional	S-V-O	Alphabet	Latin
Russian	Flexional	S-V-O	Alphabet	Cyrillic
Japanese	Agglunative	S-O-V	Syllabic	Hira/Katagana
Korean	Agglunative	S-O-V	Abjad	Hangeul

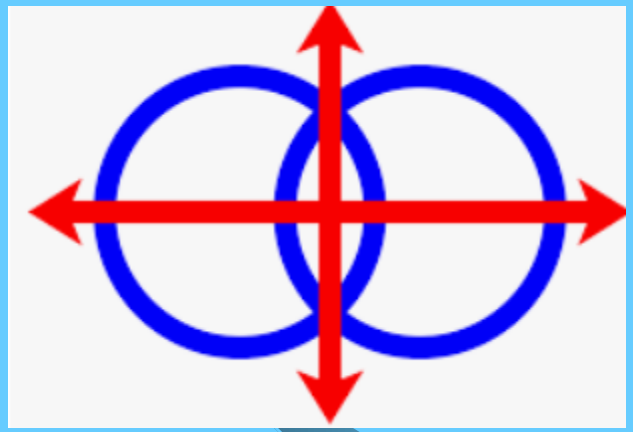
Vietnamese	English
Mạ, lúa, thóc, gạo, cơm	*horse, pony, cow, ox, Horse Power,…*
Táo quân, hát chèo,	
Con chim bay trên trời	*The bird flies in the sky*
Xưng khiêm – hô tôn	*I, you*
…	…

Differences: culture, worldview, language, history, geography, …

Appendix 1: A little bit of Esperanto

- Artificial Language invented by Zamenhof in 1887
- Unambiguous
- Easiest language => save time to learn other Indo-Euro lang.
- Several areas in Central European uses as a native language.
- Examples:
- mi kaj vi
- Mi amas vin
- Mi kaj amiko
- Mi amas amikon
- Mi amas amikinon
- Mi amas amikinon belan
- Mi ne amas amikinon malbelan
- Amikinon malbelan ne amas mi

Appendix 2: Lojban



Lojban

Constructed language

Lojban is a constructed, syntactically unambiguous human language created by the Logical Language Group. It succeeds the Loglan project. The Logical Language Group began developing Lojban in 1987. Wikipedia

Writing system: Latin and others

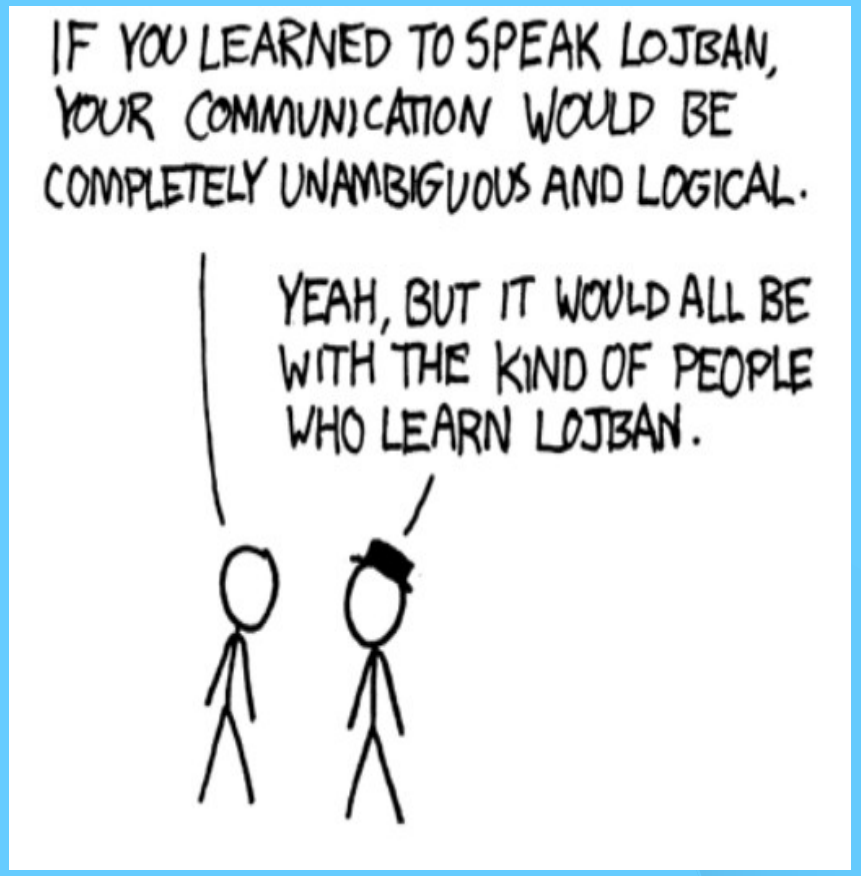
Created by: Logical Language Group

Purpose: Constructed languages > engineered languages > logical languages > Lojban

Setting and usage: a logically engineered language for various usages

Sources: Loglan


da'i ganai do crebi'o la lojban gi le se cusku be do cu mulno pavysmu je logji
.i .ie ku'i cusku fi le prenu klesi poi certu la lojban


IF YOU LEARNED TO SPEAK LOJBAN, YOUR COMMUNICATION WOULD BE COMPLETELY UNAMBIGUOUS AND LOGICAL.
YEAH, BUT IT WOULD ALL BE WITH THE KIND OF PEOPLE WHO LEARN LOJBAN.

Lojban

C code

```
char* str_fill(char string[], char filling, uint_t count) {
   if (string) {
      if (!count) while(string[count]) count++;
      string[count] = '\0';
      while (count--) string[count] = filling;
   }
   return string;
}
```

English translation

```
Function str_fill returning char pointer, taking parameters:
    'string', of type char array; 'filling', of type char; 'count', of type uint_t.
Variable list: empty.
Instruction list:
    if 'string' is not equal to 0:
    >   if 'count' is equal to 0:
    >   >   as long as the offset number 'count' of the array 'string' is not equal to 0:
    >   >   >   we add 1 to 'count'
    >   we set the offset number 'count' of the array 'string' to 0
    >   as long as 'count' is not equal to 0:
    >   >   we substract 1 from 'count'
    >   >   we copy 'filling' into the offset number 'count' of the array 'string'
    we return to the calling function giving the value 'string'
End of function.
```

Lojban translation (to be improved)

Attempt by Danr

```
to ro da zo'u la'e da du ca'e lo se judri be da toi

la .styrfil. cu pruce
  la .strin. poi judri lo lerfu ku'o ce'o
  la .filin. poi lerfu ku'o ce'o
  la .kaunt. poi mulna'u ku'o
  la .strin. poi ke'u judri lo lerfu ku'o
  lo pu'u
    va'o lo nu la'e la .strin. na du li no kei
      ba gi
        va'o lo nu la .kaunt. na du li no kei
        ze'a lo nu la'e lo sumji be la .strin. bei la .kaunt.
                    na du li no kei
        ko setca fi la .kaunt.
                  fe lo sumji be la .kaunt. bei li pa
      gi ba gi
        ko setca fi la'e lo sumji be la .strin. bei la .kaunt.
                  fe li no
      gi
        ze'a lo nu ba gi ko setca fi la .kaunt.
                                  fe lo se sumji be la .kaunt. bei li pa
                    gi la .kaunt. na du li pa kei
        ko setca fi la'e lo sumji be la .strin. bei la .kaunt.
                  fe la .filin.
```

mi milxe lo ka se mansa ge tu'a zo ze'a gi tu'a zo'oi .return.