

# Turkish in Terms of Informatics

Prof. Dr. Eşref ADALI

[adali@itu.edu.tr](mailto:adali@itu.edu.tr)

# Language

People were able to communicate and understand each other by speaking. Thanks to language, advances were made in culture, art and science. Languages take a long time to develop and their development is continuous.

In time, new words are added to the language or some words are faded out depending on the developments in the society. In fact, the way a language constructs sentences can change over time.

According to the United Nations data, it is understood that more than 4000 languages are spoken today. However, some of these languages are widely spoken, while others are used by a few people.

# Language

There is a close connection between the sophistication of languages and the development of the societies speaking that language in the field of culture, art and science. In other words, communities with insufficient language cannot be expected to be successful in the fields of culture, art and science.

Confucius (551-474 BC)

*“... If I were to take charge of a country, the first thing I would do would undoubtedly be to review its language. Because if the language is imperfect, words cannot express thought well. Duties and services cannot be performed properly if the thought cannot be expressed well. In places where duty and service cannot be performed properly, customs, rules and culture are broken. If custom, rule and culture are corrupted, justice will go astray. If justice goes astray, the people who are bewildered do not know what to do or where the business will lead. That's why nothing is as important as language!...”*

# Language Classification

## Isolating Languages

(Single syllable)

*Chinese, Vietnamese*

## Agglutinative Languages

*Turkish, Finnish*

## Fusional Languages

*India-European language :  
German, English*

## Polysynthetic Languages

*Native American*

In the isolating language family, words do not take suffixes. Words gain meaning depending on their order and emphasis in the sentence. Chinese, Tibetan, Vietnamese, and Himalayan are counted in the set of isolating languages.

In agglutinative languages, the basis of the word is the root word. New words are formed by adding derivational and inflectional suffixes to the root word, but the root never changes. There is no limit to the number of suffixes that can be added to the root word. Therefore, many words can be derived from a root word. The agglutinative language family includes **Turkish, Hungarian, Manchu, Tungusic, Finnish, Mongolian, Samoyed, Korean, and Japanese.**

Indo-European and Hami-Semitic languages are evaluated within the fusional language structure. In Indo-European languages, the body word acquires a new meaning by taking prefixes and suffixes. The number of prefixes and suffixes added to a root usually does not exceed one.

In polysynthetic languages, which are not very common, the verb merges with the other elements of the sentence. Therefore, the action can be the entire sentence. Native American languages are included in this cluster.

# Language Families



# Technical Features of Turkish

The technical features of Turkish will be made in terms of classical grammar and the technical aspects that we will introduce in this paper. As it is known, classical grammar topics are **phonology**, **prosody**, **morphology**, **syntax**, **semantics**, **discourse** and **pragmatics**. Our review will be on the features listed below.

Clarity

Ease of pronunciation and good-sounding

Vocabulary

Productivity

Prominence

Being regular

# Clarity - I

Since language is a tool used to transfer information between individuals, its **intelligibility is an important** feature. Intelligibility is the measure of how much of what the speaker is saying is correctly understood by the listeners. **Not every word spoken by the speaker will be understood by the listeners.**

$$AD = 1 + \frac{YS_1 + YS_2}{M_1 + M_2}$$

AD: Degree of intelligibility

YS: Incorrect number

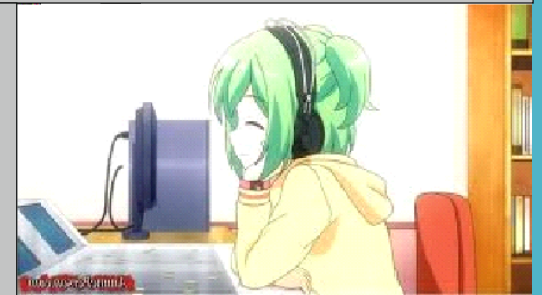
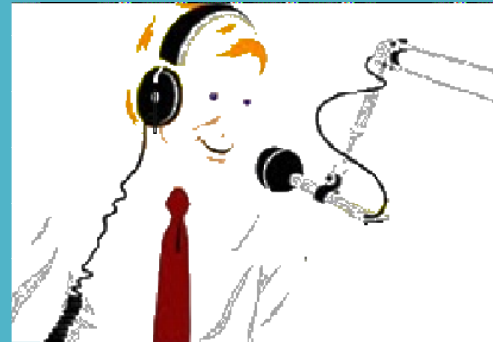
YS<sub>1</sub>: Incorrect number of first listener

YS<sub>2</sub>: Incorrect number of second listener

M : Number of words in the text read

M<sub>1</sub>: Number of words of the text read by the first reader

M<sub>2</sub>: Number of words of the text read by the second reader



# Clarity - II

## Turkish is an agglutinative language

In agglutinative languages, the basis of the word is the root. The root enlivens the basic meaning in the human mind. A Turkish speaker can predict the suffixes that can be added to a root.

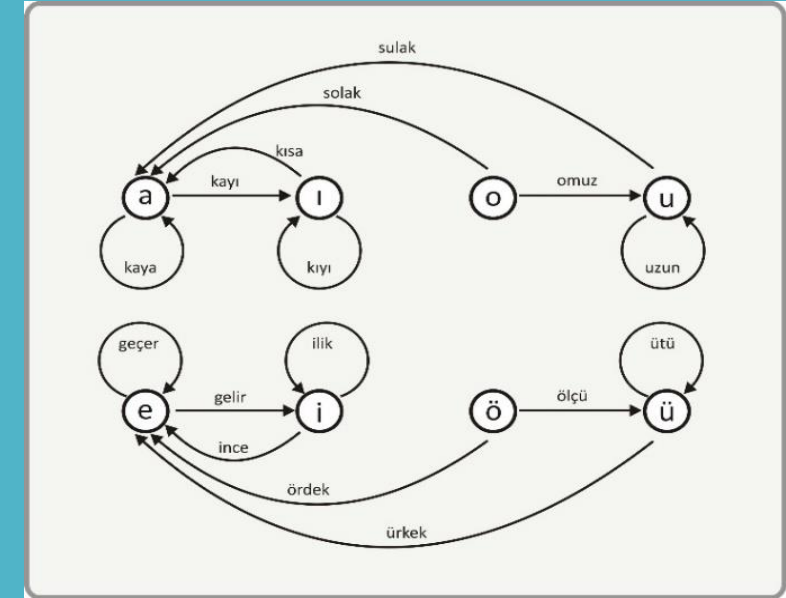
## Vowel and consonant harmony in Turkish

There are harmony rules for vowels and consonants in Turkish. These harmony rules are explained in the form of Finite State Machine representation. Vowel and consonant harmony enable the listener to predict the next vowels and consonants. Therefore, it makes it easier to understand.

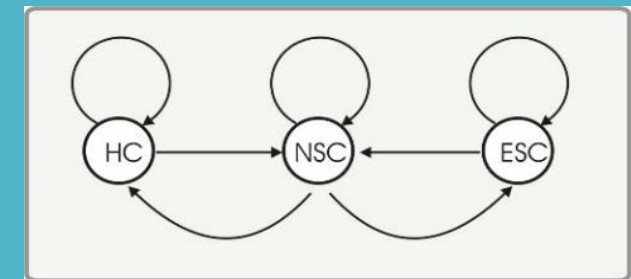
## Syllable structure of Turkish

The syllable structures in Turkish are clear and are V, VC, VCC, CV, CVC, CVCC

The syllabic structure is formed in the mind of an individual whose mother tongue is Turkish and he expects it. Therefore, he interprets the word in this sense.



Vowel harmony



Consonant harmony

HC: hard con. NSC: soft con. but not equ., ESC: soft cons wit equ.



# Ease of Pronunciation - I

Turkish as the Italian of the East.

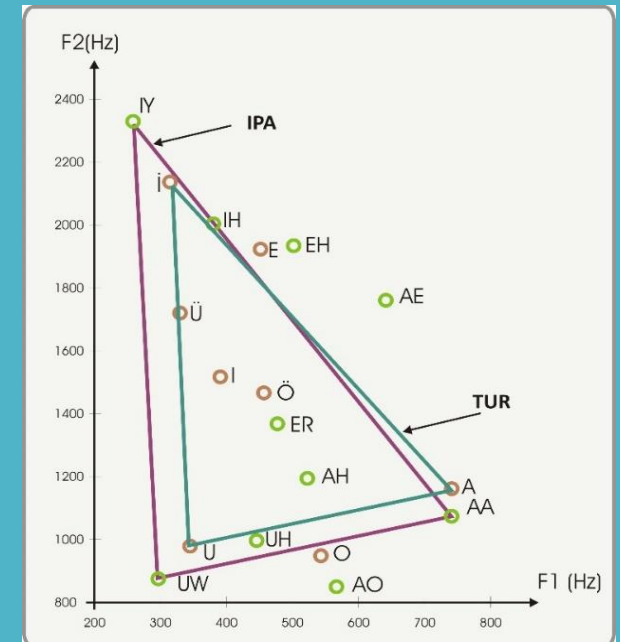
The number of vowels and consonants in every language is not the same. In Table, vowel and consonant numbers in some languages are given.

The formant frequency of vowel sounds determined by IPA (International Phonetic Alphabet) are shown in Figure together with Turkish vowels. As can be seen from this figure, the vowels of Turkish are very similar to the vowels determined by the IPA.

Pronunciation of consonants can differ from language to language. For example, there is a difference between the "R" sound in Turkish and the "R" sound in English. While the English R sound is in the form of amplitude modulation, it is in the form of frequency modulation in Turkish. Those with short sublingual membranes cannot produce the Turkish R sound because their tongue cannot tremble.

Vowels and Consonants in Languages

| Language | Number of Vowels | Number of Consonants |
|----------|------------------|----------------------|
| Turkish  | 8+5              | 21                   |
| Arabic   | 3                | 28                   |
| German   | 8                | 20                   |
| English  | 5                | 21                   |
| French   | 6                | 20                   |
| Russian  | 10               | 20                   |



Formants of IPA and Turkish

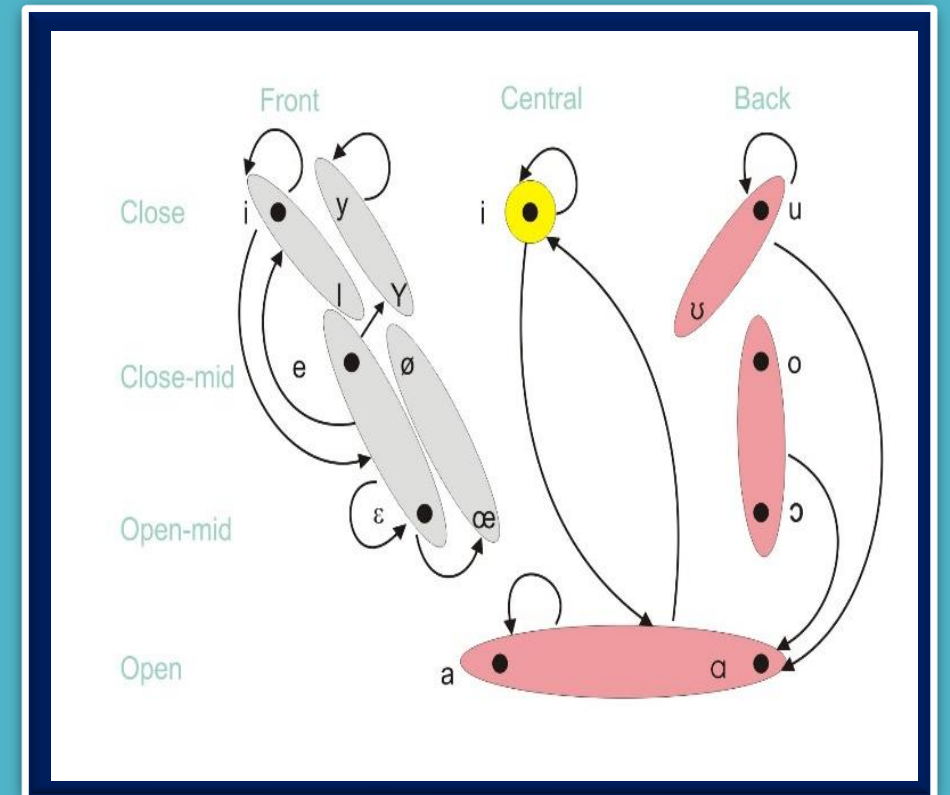
# Ease of Pronunciation - II

The tongue is an important element of the vocal organ and the sounds consist of the tongue, very roughly the vocal cords, the nasal larynx, the tongue and the lip as the tongue. It is especially important in which part of the tongue the vowels are produced.

## Clustering of Vowels in Turkish According to Their Origin

|                | Front |       | Central |       | Back |       |
|----------------|-------|-------|---------|-------|------|-------|
|                | Flat  | Round | Flat    | Round | Flat | Round |
| Close (narrow) | i     | ü     | ɨ       |       |      | u     |
| Open (wide)    | e     | ö     |         |       | a    | o     |

We can think of the human tongue as a flexible mass that is rooted and can move longitudinally and transversely. When the situation of such a mass moving from one position to another is examined, it is clear that the transition to the neighboring state/position will be easier than the transition to the far state/position. Therefore, the transition from **e** to **i** and **ö** to **ü** is easy. Similarly, the transition from **a** to **ɨ** and **o** to **u** is easy. This explains the origin of the vowel harmony. In other words, **vowel harmony is a result of the nature of the human vocal organ.**



# Energy Saver

## Turkish speaking gets tired less

- The vowel harmony feature of Turkish requires less tongue, lip and jaw movements while speaking.
- Therefore, the jaw of a Turkish-speaking person is less tired than other people.
- When Turkish and foreign speakers presenting news on television are watched, it will be seen how accurate our interpretation is.



# Ease of Pronunciation - III

## Turkish is Easy to Say

The language of a Turkish-speaking person is not difficult due to vowel and consonant harmony rules, in other words, Turkish is an easy language to pronounce.

Let's try to read these two artificial words to see the effect of vowel harmony on pronunciation.

*tentered, tintorido*

It's easier to say the first word and less strain on our jaws. Although this word is not a Turkish word, it is easy to say because the vowels are in harmony with the Turkish phonetic rules.

# Good Sounding

## Turkish sounds good

### In General

- Vowels are melody sounds of a language.
- The aesthetics of a language is directly proportional to the multiplicity of its vowels.
- Extending the duration of vowel sounds brings fluency to the language

### Turkish

- Considering that there are 3 vowels in Arabic and 5 in English, Turkish is considered rich in vowels. Therefore, it can be called a melodic language.
- Pronunciation of vowels can be extended up to twice. Ğ not only extends the duration of the vowels before it, but also allows them to soften.
- Vowel and consonant harmony add rhythm to speech.
- The place of emphases in Turkish is determined by the rules. This feature makes Turkish sound nice.

# Vocabulary

The effectiveness of a language can also be measured by the richness of its vocabulary. The size of the vocabulary of a language is directly proportional to the thinking ability of the person speaking that language. When investigating the richness of a language's vocabulary, only looking at the number of words in the dictionary does not give accurate results.

When we look at English or French dictionaries, we can see the richness of words. On the other hand, the vocabulary in Turkish dictionaries can be seen as few. In fusional languages, it is necessary to give a new name to each concrete and abstract object.

## Vocabulary Generating Ability of Turkish

| Turkish     | English       | Turkish     | English                |
|-------------|---------------|-------------|------------------------|
| Göz         | Eye           | Gözlemci    | Observer               |
| Gözlük      | Eyeglasses    | Gözlemcilik | Observation            |
| Gözlükçü    | Optician      | Gözde       | Favourite              |
| Gözlükçülük | Opticians     | Gözü        | Mirror                 |
| Gözcü       | Watchman      | Gözümlük    | Mirror stand           |
| Gözcülük    | Ophthalmology | Gözücü      | Mirror maker           |
| Gözlem      | Observation   | Gözücülük   | Mirror makers business |
| Gözleme     | Observing     |             |                        |

## Verb Structures of Turkish

| Verb type              | Turkish       | English  |
|------------------------|---------------|--|
| Active                 | Görmek        | To see   |
| Passive                | Görünmek      | To be seen   |
| Passive                | Görülmek      | To be seen   |
| Reflexive              | Giyinmek      | To dress   |
| Reciprocal             | Görüşmek      | To see each other or to discuss                    |
| Causative              | Görüştürmek   | To bring someone to see or to discuss each other   |
| Second level causative | Görüştürtmek  | To have somebodies to see or to discuss each other |
| Passive causative      | Görüştürülmek | To be brought to see or to discuss somebody        |

# Productivity

An indicator of the development of a language is the meanings attributed to words

- Additions to words in agglutinative languages give words new meanings. These features not only enrich the vocabulary of the language, but also add functionality to the language.
- Active and passive forms of verbs are seen in all languages. However, it is not as easy and regular as in Turkish to form **reflexive, reciprocal, causative, secondary causative, causative-passive** forms of verbs.

- *Soğuk hayvanları kırdı*: Cold broke (kill) animals
- *Fiyatları kırmak*: Breaking prices (discounting)
- *Kalbimi kırdı*: Broke my heart (offended)
- *Pulunu kırmak*: Breaking the checker (in backgammon game)
- *Buğdayı kırdırmak*: Crushing the wheat (coarse grinding)
- *Direksiyonu kırmak*: Breaking the steering wheel (Turning, changing direction)
- *Soğuğun belini kırmak*: Breaking the back of the cold (Seriously reducing the effect of the cold)
- *Senet kırmak*: Breaking a promissory note (turning a bill into money)
- *Dersi kırmak*: Breaking the lesson (avoiding the lesson)

# Prominence - I

It is expected that there should be a certain order in the order of the words in the sentence. In each language, certain rules are followed in ordering the words in the sentence. In some languages, if this order is changed, the meaning of the sentence is distorted or lost. In Turkish sentences, the words are arranged in a certain order. However, there is no loss of meaning when the order is changed. However, there is a difference in emphasis. For example, the meaning of these six sentences is the same, the stressed word is different. The translation of these sentences are the same in English (Wash your hands with soap).

*Elini sabunla yıka* (emphasis with soap)  
*Elini yıka sabunla* (emphasis your hand)  
*Sabunla elini yıka* (emphasis your hand)  
*Sabunla yıka elini* (emphasis with soap)  
*Yıka sabunla elini* (emphasis wash)  
*Yıka elini sabunla* (emphasis wash)

*el*: hand  
*elini*: your hand  
*sabun*: soap  
*sabunla*: with soap  
*yıka*: wash)

The sentence structure in Indo-European languages is rigid. For this reason, when the words in the sentence are changed, the meaning of the sentence changes or is distorted. The flexibility of the sentence structure increases the semantic power of Turkish and provides ability and convenience to the speakers. However, this feature causes additional difficulties in natural language processing studies.



# Prominence - II

We can also examine languages in terms of case suffixes. The meaning of case suffixes in Turkish is very precise. This precision is not at the same level in Indo-European languages.

*Müziği dinliyorum*: I am listening << (accusative) the music. ( I am listening to music)

*Seni düşünüyorum*. I am thinking << (accusative) you. (I think of you)

*Partililere konuşuyorum*: I speak >> (dative) to party members. (I am speaking to members of parties)

As it is known, there are six cases of nouns in Turkish: - (nominative), -i (accusative), -e (dative), -de (locative), -den (ablative), -in (genitive). Since these suffixes indicate direction, they can be considered as vectors. It cannot be wrong in the meanings of the case suffixes

# Being Regular

- It is thought that languages have developed with the development of humanity, so it should not be expected to have definite rules. However, this idea is not valid for Turkish. **It is as if a language convention was held 5000 years ago and the rules of Turkish were determined. These rules have not changed until today.**
- This feature of Turkish amazes linguists. The unchanging rules of Turkish have never been broken in phonetics and morphology. An irregular example can be given in the rules of morphology. As it is known, when adding possessive suffixes to a word ending with a vowel, the letter "n" is added. According to this rule, we should say *su+n+un* for the word *su*, but we say *suyun*. This situation, which we can consider as a violation of the rule, is actually due to the word *suy* being said in the past tense.

## The roots of Turkish verbs never change.

- For example, if we write the verb to go for all tenses (**git**+tim: I went; **git**+miştim: I went; **gid**+iyorum: I am going; **gid**+eceğim: I will go; **gid**+erim; I go), we will clearly see this property. In the example we gave, the letter "t" can turn into "d" due to the harmony rule, and this is also regular.
- In the Indo-European language family, there is not even a letter similarity between the present tense and the past tense of the verb to go.

| Language | Tens            |                  |
|----------|-----------------|------------------|
|          | Simple present  | Past             |
| Turkish  | <b>Gid</b> erim | <b>Git</b> tim   |
| English  | I go            | I went           |
| French   | Je vais         | Je suis allé     |
| Persian  | Harem reft      | Reftem           |
| Spanish  | Yo voy          | Yo fuit          |
| German   | ich gehe        | ich bin gegangen |
| Italian  | andrò           | sono andato      |

Eşref ADALI

Türkçe  
Doğal Dil İşleme

Eşref ADALI

Türkçe  
Doğal Dil İşleme

AKÇAĞ