Lecture 19: Languages of the World

Yong-cheol Lee
University of Pennsylvania

Ling 001, Fall 2011
November 16, 2011

ROADMAP

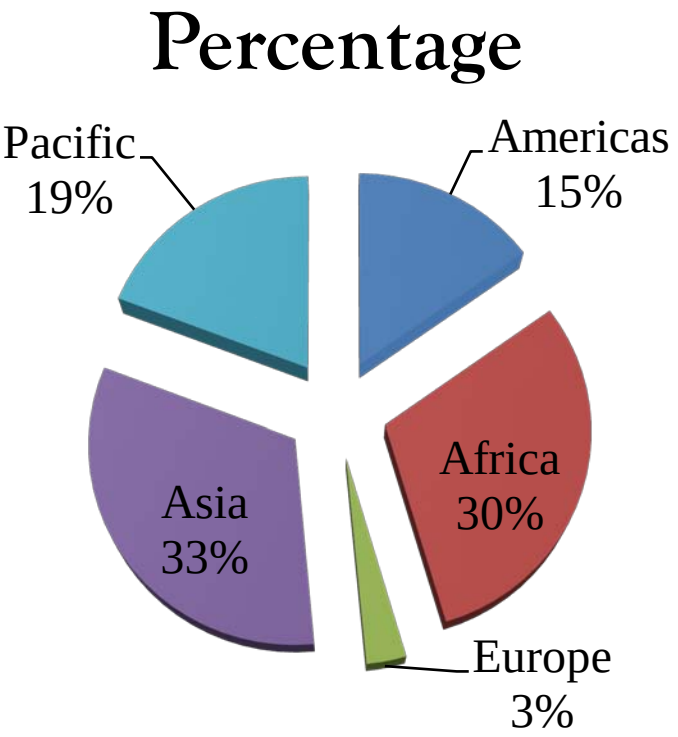
1. Languages of the World
2. Language Families
3. Language Typology

ROADMAP

1. Languages of the World
2. Language Families
3. Language Typology

How many languages in the world?

- The 1996 edition of Ethnologue listed 6,703 living languages, with their original locations divided geographically.



How many languages in the world?

- The 2005 edition lists **6,912** living languages.

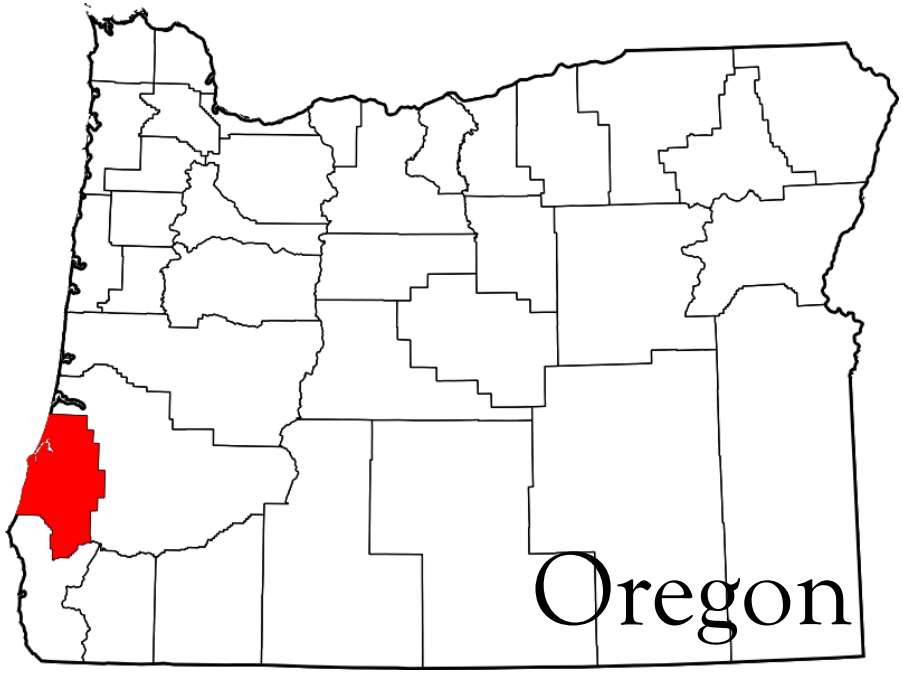
  6703 (1996) → 6912 (2005)

  Question:
  Have 209 languages been created from 1996 to 2005?

- **No!** A combination of a more complete inventory and some decisions were made as to how many speech communities to distinguish as "**languages**".

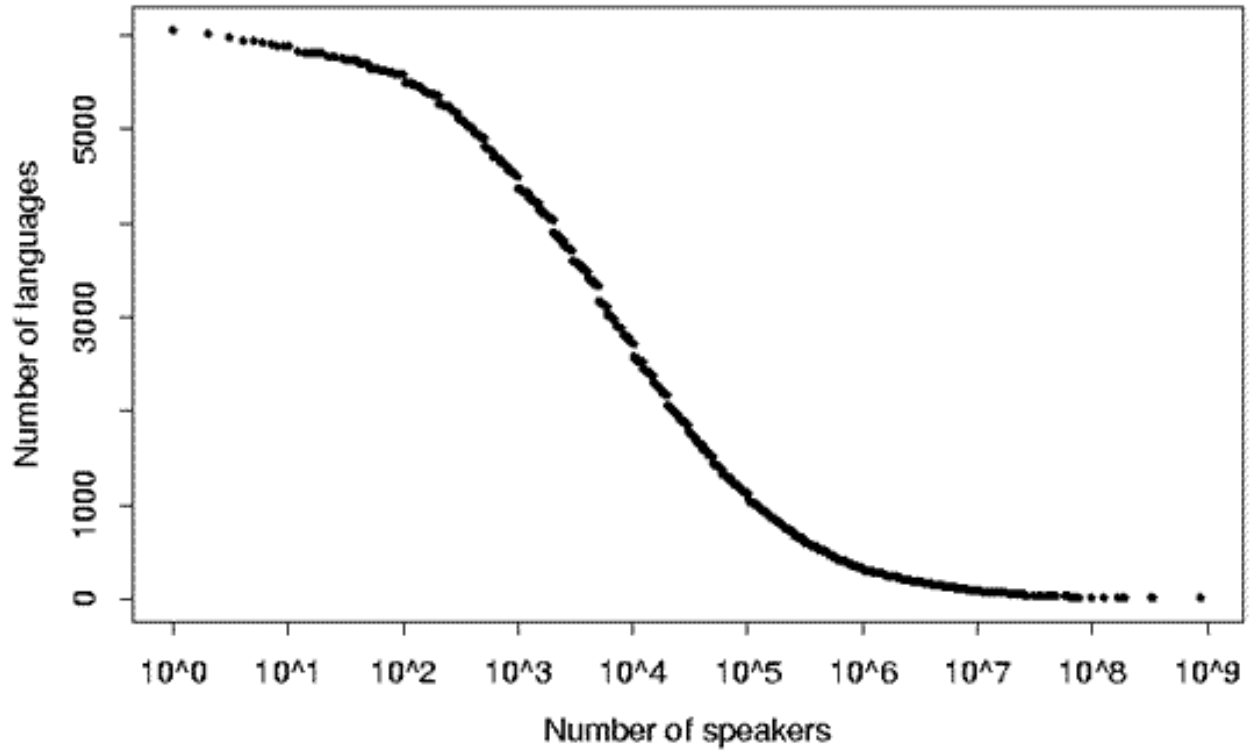
How to distinguish languages?

- Mandarin Chinese, with 885 million native speakers in 12 countries.
- Coosan, with 1 or 2 speakers in Southern Oregon.



From Wikipedia, the free encyclopedia

Speaker count vs. Language count



Data from Ethnologue (1999)

- There are 6059 languages with 1 or more speakers like Coosan.
- There are 80 languages with 10 million or more speakers.

Language extinction

- Many of the 6,000-odd "living" languages are endangered or nearly extinct.
- The language with 10,000 or fewer speakers, are especially vulnerable.

Question: Why does language extinction occur?

Language extinction

- Roughly half of the world's languages are moribund since new generations of children are not being raised to speak them.
- Thus, within a century, the number of living languages will be cut at least in half.

The "top 20" languages

Rank	Language	Native speakers (in millions)
1	Mandarin Chinese	885
2	English	322
3	Spanish	266
4	Bengali	189
5	Hindi	182
6	Portuguese	170
7	Russian	170
8	Japanese	125
9	German	98
10	Wu Chinese	77
11	Javanese	76
12	Korean	75
13	French	72
16	Cantonese	66
19	Turkish	59

All of these counts are subject to question.

Data from Ethnologue (1996)

The "top 20" languages: Issue 1

- One may question the census figures and also the question of who counts as a speaker of which language.
- The 1996 edition of Ethnologue cites 266 million native speakers of Spanish.
- The 1999 revisions increase the number of native speakers of Spanish to 332 million.
- This is a revision in who is counted as a Spanish speaker.

The "top 20" languages: Issue 2

- Another set of questions have to do with what counts as a language.

  → Where is Arabic in the table?

- Arabic is certainly one of the world's major languages.
- In fact, Arabic has 202 million speakers world-wide (#4 on the list).
- However, Ethnologue considers the local colloquial varieties of Arabic to be separate languages.

Why is Arabic not a single language?

- Different Arabic colloquials are not mutually intelligible.
- Algerian Colloquial Arabic is roughly as different from Egyptian Colloquial Arabic as Portuguese is from Spanish.

Counter-arugment

- Educated people in all the Arabic-speaking countries can speak, read and, understand "**Modern Standard Arabic**".
- Thus, an educated Egyptian in Algeria can converse easily with an educated Algerian.
- However, this relationship cannot hold between Spanish and Portuguese.

Hindi and Urdu

- Hindi and Urdu are essentially the same language.
- For historical and political reasons, they have different writing systems.
- Ordinary speakers are likely to be able to understand one another.
- Combining their counts would give us 182+57 = 239, a 30% increase for Hindi.

Serbo-Croatian

- A recent and striking example is the change in Serbo-Croatian.
- Ten years ago, Serbo-Croatian was considered a single language.
- There were two ways of writing it: with roman characters in Croatia; with cyrillic characters in Serbia.
- Due to increasing variations, now there are three languages: Serbian, Croatian, and Bosnian.

ROADMAP

1. Languages of the World
2. Language Families
3. Language Typology

Language-universality or -specificality

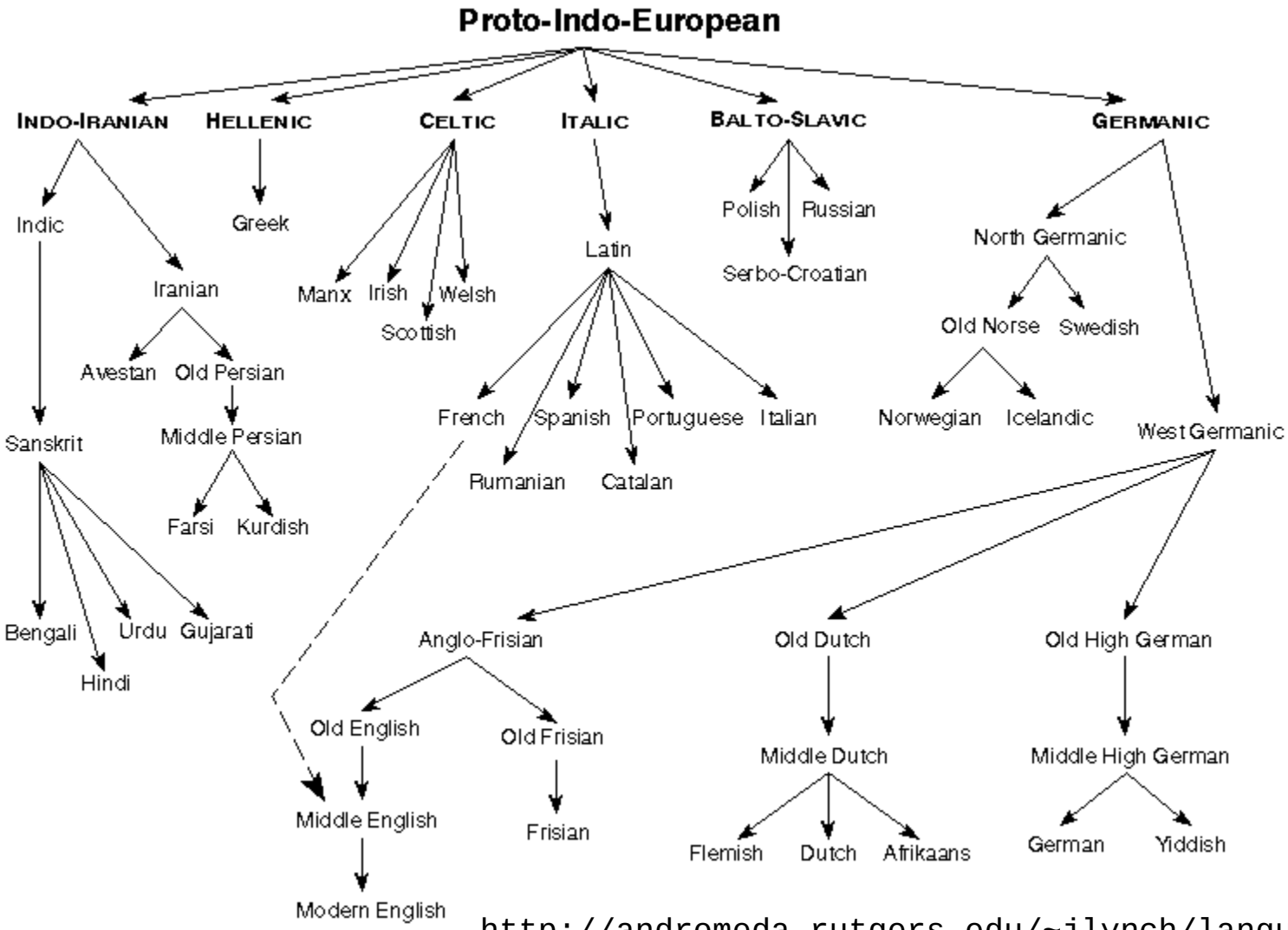
- Examining the languages of the world, we see similarities and differences.

English	German	Vietnamese
one	eins	mot
two	zwei	hai
three	drei	ba
four	vier	bon
five	fünf	nam

(Fromkin et all. 2003: 527)

Language-universality

- Languages are not a set of distinct and unrelated items, but rather a hierarchy.
- We can continue to split categories almost down to the level of the individual speaker.
- English, German, Norwegian, Danish, Swedish, Icelandic, e tc. are sisters because they descended from one parent (Germanic).
- The Romance languages (French, Spanish, and Portuguese) are also sister languages whose parent is Latin.
- Romance languages and Germanic Languages are cousins.



http://andromeda.rutgers.edu/~jlynch/language.html

- Question: How to constitute a family and the relationship among languages?

How to identify a language family?

- Two questions are generally asked to identify a language family.

    1) How many languages do you speak?
    2) How many languages are there in the world?

How many languages do you speak?

- This question has a problem.
- There are many polyglots who speak and understand several languages.

> **Charles V, the Holy Roman Emperor** (1500-1558) was a polyglot: "I speak Spanish to God, Italian to women, French to men, and German to my horse."
>
> (Fromkin et all. 2003: 532)

How many languages in the world?

- This question also has a problem.
- It's hard to ascertain the number of languages because of disagreement as to what comprises a language as opposed to a dialect.

How to categorize languages?

- In practice, the tendency is to categorize languages at level of grouping that depends on several factors:

  Mutual intelligibility

  Speaker attitudes

  Existence of a nation-state

  Geographical origins

  Historic origins

Mutual intelligibility

- How different must two dialects be before they become separate languages?
- As long as two dialects remain mutually intelligible, it is generally believed that they cannot be considered separate languages.

Mutual intelligibility

- However, mutual intelligibility is not always a symmetric relationship.
- Sometimes speakers of A can understand speakers of B, but not vice versa.
- Languages/dialects A and B are mutually intelligible, and B and C are also mutually intelligible, does not imply that A and C are mutually intelligible.

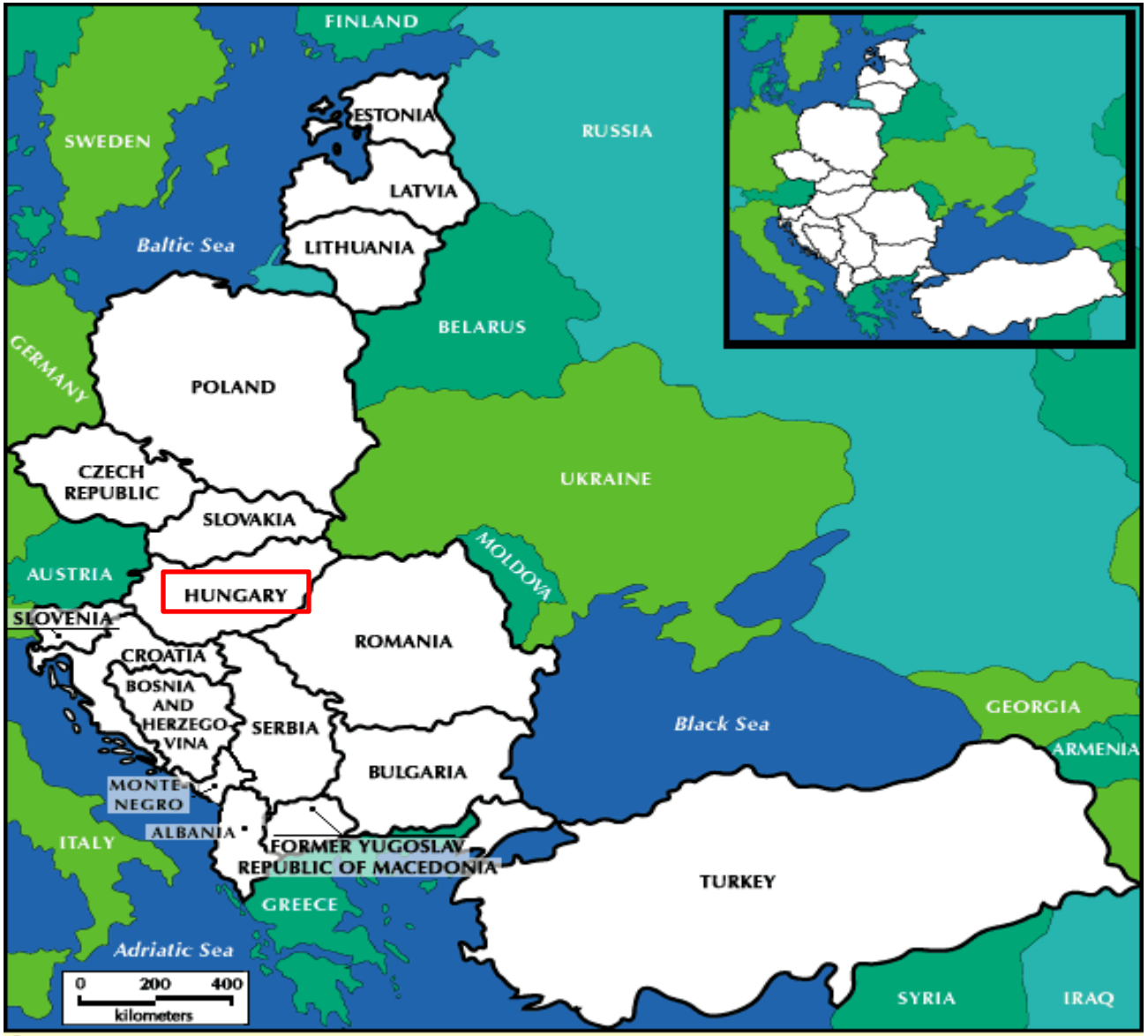
Mutual intelligibility

- The Indo-Iranian languages Hindi and Urdu are considered as separate languages since they use a different writing system even though they are mutually intelligible.
- However, mutually unintelligible languages spoken in China are often thought of as dialects because they have a common writing system and culture.

Mutually intelligibility is not a reliable factor.

Geographical origin

- It is often surprising to discover what languages are genetically related and which ones aren't.
- Nepali, the language of Nepal is an Indo-European language.
- Hungarian, surrounded on all sides by Indo-European languages is not an Indo-European language.



http://archive.rec.org/REC/Maps/map_gif/CEE_region.gif

Historic origin

Spanish and Portuguese (Iberian)

- As of 220 B.C., the southern part of the Iberian peninsula was a Carthaginian colony.
- Rome conquered southern Iberia in 206 B.C., and the rest of the peninsula somewhat later.
- Roman colonists and administrators gradually imposed the Latin language.
- The varieties of Latin spoken in various parts of the Roman Empire changed over time, giving rise to the modern "romance" languages:
  Italian, Spanish, Portuguese, French, Romanian and so on.

Roadmap

1. Languages of the World
2. Language Families
3. Language Typology

Typology

- Independent of language family associations, we can characterize languages on various dimensions where correlations among values recur.
- Typological classification can be done at any level of linguistic description, but the commonest forms are **phonological, morphological, and syntactic**.

Phonological Typology

- Phonological typologies deal with issues such as,

  Phoneme inventory
  Syllable structure
  Prosody

- We can ask several questions:

  How many distinct vowels does a language have?
  What sorts of syllable-final consonants are possible?
  Which language has word-stress?

Phonological Typology

- If a language allows stop consonants to occur in syllable-final position (English, “dip”), it will generally also allow nasal consonants in the same position (English, “dim”).
- It is often helpful to think of these patterns in terms of a hierarchy of more or less “marked” (i.e. unusual or unexpected) configurations.

Phonological Typology

A partial hierarchy of markedness for syllable structures.

Consonant Vowel >> Consonant Vowel Nasal >> Consonant Vowel Stop

- As a rule, a language that has more "marked" patterns also has less marked ones.
- To take a Japanese example, it does not allow a CVS sequence.

  Question: When a CVS sequence is imported into Japanese lexicon, what will happen?

Morphological Typology

- **Isolating**: words lack affixes, and grammatical relationships are mainly signaled by word order.
- **Inflecting**: words are marked with affixes to indicate their grammatical function.
- **Agglutinative**: words incorporate long sequences of affixal elements.
- **Polysynthetic**: whole sentences may be expressed as single words, with several stems and various functional elements expressing their relationship.

Syntatic Typology

- The most basic syntactic typology has to do with the normal order of subject, verb and object in simple sentences.
- There are six possible orders:
  SOV, SVO, VSO, VOS, OVS, and OSV.

SVO: English, French, Swahili, Hausa, Thai
VSO: Tagalog, Irish, (Classical) Arabic, (Biblical) Hebrew
SOV: Turkish, Japanese, Persian, Georgian
OVS: Apali (Brazil), Barasano (Colombia)
OSV: Apurina and Xavante (Brazil)
VOS: Cakchiquel (Guatemala), Huave (Mexico)

(Fromkin et all. 2003: 532)