Word order phenomenon in Languages

The native speakers arrange the words in the sentences to encode the information which the speakers decode to understand the indented meaning for the purpose of communication.

The words which are arranged in the sentence with loads of information have to obey the rules of grammar of the language which is commonly agreed upon by the native speakers. (formal universals)

The order of the words and the choice of re-arranging them in different orders depend on the type of languages (morphological types) and the contextual requirement of the discourse.

- The typology of languages has shown particular interest in the relative ordering of subject (S), verb (V) and object (0).
- The most prevalent distribution of these three elements in a language is referred to as the ‘WORD ORDER PHENOMENON’ in language typology.
- There could be one or more choices of arranging the words in the sentence.
- But out of these choices, there could be ONE arrangement of the words which is popularly known as ‘Basic word order’.
- We will have a full class on ‘how to determine the basic word order’ out of several choices even in languages which have flexible word order.

- However, there seems to be a theoretical problem with this terminology.
- If we look at the data from different languages, we would realize that it is not a SINGLE word out of which the subject, object or even the verb is made in all the cases in any language.
- For example,
- **John** is very funny.
- **The new student** is very funny.
- **That the typo class is held in #100** is very funny.
- **The book about lady Diana's affair with the rich French businessman** is banned.

These examples show that the SUBJECT can be of different length.

Therefore, it is agreed upon by the scholars that the term should be replaced with the '**basic constituent order**'.

What is a constituent then?

Well, a constituent is either a word or a group of words which makes the category like SUBJECT, OBJECT, VERB, ADJ, ADV etc.

However, a formal definition of the term would read like '…. a constituent is **a word or a group of words that function(s) as a single unit within a hierarchical structure**'.

- **The issue of word order:**
- There are languages in which the basic constituent order can easily be decided.
- For example, we may come across with the following sentences in English:
- 1. a. Sweet, I hate.
- b. Believe in you me, .. no way.
- c. The maid sliced the bread.
- So, we can have different sentences with different orders, such as OSV (1a) and VSO (1b) even in English,
- However, it is quite clear that these orders are "special" and that the one which has SVO order of (1c) is typically what we mostly find in English.

If we re-examine the examples, we would say that the example (1c) is **more basic** to the other examples in English.

In addition to the intuition of native speakers, there are many other reasons to claim that (1c) is the basic in English.

There is a distinct intonation-pattern in (1a), a slight pause after *sweet*.

And we know that without some discourse context, (1a) sounds odd and even ungrammatical.

The example (1b) also looks bound to some context.

Moreover, it is not considered grammatical by all speakers of English, because it is more like an idiom.

Thus, the examples (1a) and (1b) are not the representatives of basic constituent order in English.

Many other languages do not have such strict or regid word order that we just discussed in English and some other languages like Chinese, Vietnamese etc.

These languages often provide a unique challenge to typologists because their basic word order is much more difficult to decide.

For example in Hindi, there exists many options to arrange the words in a given sentence.

2. a.	lərkõ-ne	bãdər-ko	dek^{h}-a	[sov]
	Boy-3Pl-Erg	monkey-Acc	see-pst-3S	
	'Boys saw the monkey'.			
b.	bãdər-ko	lərkõ-ne	dek^{h}-a	[osv]
c.	dek^{h}-a	lərkõ-ne	bãdər-ko	[vso]
d.	lərkõ-ne	dek^{h}-a	bãdər-ko	[svo]
e.	bãdər-ko	dek^{h}-a	lərkõ-ne	[ovs]

In Sanskrit, one could have the following orders in a sentence:

3	pətrɑɲı	vr̥ksʰɑt	pətəntı	SOV
	leave-3Pl-Nom	tree-3S-Abl	fall-3Pl-Imp-Pres	
	'The leaves are falling from the tree'.			

3a	vr̥ksʰɑt	pətrɑɲı	pətəntı	OSV
	tree-3S-Abl	leave-3Pl-Nom	fall-3Pl-Imp-Pres	
	'The leaves are falling from the tree'.			

3b	pətəntı	pətrɑɲı	vr̥ksʰɑt	VSO
	fall-3Pl-Imp-Pres	leave-3Pl-Nom	tree-3S-Abl	
	'The leaves are falling from the tree'.			

3c	vr̥ksʰɑt	pətəntı	pətrɑɲı	OVS
	tree-3S-Abl	fall-3Pl-Imp-Pres	leave-3Pl-Nom	
	'The leaves are falling from the tree'.			

3d	pətəntı	vr̥ksʰɑt	pətrɑɲı	VOS
	fall-3Pl-Imp-Pres	tree-3S-Abl	leave-3Pl-Nom	
	'The leaves are falling from the tree'.			

3e	pətrɑɲı	pətəntı	vr̥ksʰɑt	SVO
	leave-3Pl-Nom	fall-3Pl-Imp-Pres	tree-3S-Abl	
	'The leaves are falling from the tree'.			

All the above orderings of the constituents are possible in Hindi and Sanskrit.

Despite the variation in word order, there is no confusion over the intended meaning of the sentences in Hindi & Sanskrit.

The nominals in Hindi and Sanskrit, unlike English, are case marked and the case markers help determining the grammatical role of the constituents in the sentence.

Therefore, moving of the order of the constituents does not pose any problem for the intended meaning in any of the sentences.

The noun declension and verb conjugation which are very rich in Hindi and Sanskrit, facilitate the moving the constituents without affecting the intended meaning.

This is not true about English. The case is very much the part of the structure and given to the constituents in the place where they occur in the sentence and thus, changing the order of the constituents changes the meaning of the sentence.

For example:

a. The lion ate the goat.

But we can't say

*b. The goat ate the lion.

Let us see why we can't have (b) as a grammatical sentence in English and other rigid word-order languages.

a.	**The lion**	**ate**	**the goat**
	Det N	V[+tr]	Det N
	Noun –Nom		noun-Acc
	Subject	Verb	Object (DO)
	Agent	V[action]	Patient

*b.	**The goat**	**ate**	**the lion**
	Det N	V[+tr]	Det N
	Noun –Nom		noun-Acc
	Subject	Verb	Object (DO)
	Agent	V[action]	Patient

There are six logically possible orders of S, V, and O.

And all of them have been claimed to serve as the basic constituent order for at least one language in the world.

An example of each is provided in (4):

4| SOV: taro ga inu o mita (Japanese [Japanese-Ryukyuan: Japan])
Taro sub dog obj saw
'Taro saw the dog'.

SVO : umugore ara-soma igitabo (Kinyarwanda [Niger-Congo: Rwanda])
woman 3S-read book
'The woman is reading a book'.

VSO: bara elohim et ha-shamayim (Biblical Hebrew [Semitic])
created God OBJ Art-heavens
'God created the heavens'.

4.

VOS: manasa lamba amin'ny savony ny lehilahy (Austronesian)
washes clothes with the soap the man
'The man washes clothes with the soap'.

OVS: toto yahosiye kamara (Hixkaryana [Carib: Brazil])
man it-grabbed-him panther
'The panther grabbed the man'. (Data from Derbyshire 1985)

OSV: pako xua u'u (Urubu [Equatorial-Tucanoan: Brazil])
banana John he-ate
'John ate bananas'. (Data from Kakumasu as cited in Derbyshire and Pullum 1981)

This is really fascinating to see that all six arrangements of the word-order are attested from the languages of the world.

So, what we saw in the last two slides is that all the six possible orders of these constituents are found in at least one language of the world.

However, they are not evenly distributed among the languages of the world.

This provides a clue to a significant cognitive principle of human language.

If the ordering of S, V, and O were random, we would expect each of the constituent order types to appear with the same frequency.

However, the facts are contrary to our assumption and some orders turn out to be relatively common, whereas others are remarkably rare.

While doing a course in language typology, we must try to understand as to why this happens!

Table-1. **Russell Tomlin, *Basic Word Order : Functional Principles*, (Croom Helm, London, 1986)**

Word Order Distribution of Languages		
Basic Word Order	Proportion of Languages	Examples
Subject-[Verb-Object]	42%	English, Indonesian
Subject-[Object-Verb]	45%	Japanese, Turkish
Verb-Subject-Object	9%	Welsh, Zapotec
[Verb-Object]-Subject	3%	Malagasy
[Object-Verb]-Subject	1%	
Object-Subject-Verb	0%	

Russell Tomlin, *Basic Word Order:Functional Principles*, (Croom Helm, London, 1986) page 22

2. Estimates of the Word Order Distribution of Languages

Word Order Type	Greenberg 1963	Ultan 1969	Ruhlen 1975	Mallison & Blake 1981	Tomlin 1979	Tomlin 1986
SVO	43.0%	34.6%	35.6%	35.0%	41.5%	41.8%
SOV	37.0%	44.0%	51.5%	41.0%	45.8%	44.8%
VSO	20.0%	18.6%	10.5%	9.0%	11.0%	9.2%
VOS	0.0%	2.6%	2.1%	2.0%	1.5%	3.0%
OVS	0.0%	0.0%	0.0%	1.0%	0.3%	0.0%
OSV	0.0%	0.0%	0.2%	1.0%	0.0%	0.0%
Unclassified	0.0%	0.0%	0.0%	11.0%	0.0%	0.0%
Number of Languages	30	75	427	100		402

In both the tables given above, the frequency of occurrence of SOV and SVO is amazing.

If basic constituent order were not governed by some principles of language and cognition, then each of the six potential orders would occur with roughly the same statistical frequency (16%).

But, we saw that SOV and SVO are found in over 40% of the languages in the sample.

However, a total of these two orders will make this figure go as high as 85%.

Thus, it is more than clear that the distribution cannot be taken as random.

And some explanation must be given to determine their (i.e. SOV and SVO) statistical dominance.

- A slightly different arrangement of the data in Table (1) comparing the relative ordering of just two of the basic constituents, Sub and Obj, reveals another striking fact which is shown in the Table (2):
- The primary concern here is to examine the fact that the subject precedes the object in the word-order.

Table (2):

Relative Frequencies of the Order of S + 0

	Languages	
Word Order	**Number**	**%**
SO	385	96
OS	17	4
Total	402	

Source: Ruhlen 1975

- The figures in the above table depict the fact that there is distinct difference amongst the languages that place the subject before the object (96%) than those that place the subject after the object (4%).
- This is not a chance similarity or any accidental outcome, rather it has some cognitive foregrounding of human thought.
- This distribution was also noted by Greenberg (1963) and he has captured this in his very first Universal:
- *Greenberg's Universal 1: In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object.*
- It is true that Greenburg didn't hypothesize anything about this dominance at the time of reporting the above universal.
- However, his reporting about the dominance gave others the hint for this dominance and later linguists made big hypothesis about this dominance.

Although this universal stipulates the linear precedence of subjects over objects, it does not explain why it should hold true for such a vast number of languages.

But later, Comrie (1989, 93) suggests that this glaring priority of subject over object in terms of the dominance has a functional explanation.

Comrie has claimed (ibid) that presumably a deeper cognitive organization of information underlies the pattern in which the subject precedes the object.

In a transitive clause, the subject generally is the initiator of the action expressed by the verb and is in control of that action, whereas the object is a mere entity on which the action is being acted on.

These properties of the subject make it more salient than the object in human cognition and this salient property of the subject is reflected in languages when they develop a constituent order that puts subjects before objects.

Table-1. **Russell Tomlin, *Basic Word Order:Functional Principles*, (Croom Helm, London, 1986)**

Word Order Distribution of Languages		
Basic Word Order	Proportion of Languages	Examples
Subject-[Verb-Object]	42%	English, Indonesian
Subject-[Object-Verb]	45%	Japanese, Turkish
Verb-Subject-Object	9%	Welsh, Zapotec
[Verb-Object]-Subject	3%	Malagasy
[Object-Verb]-Subject	1%	
Object-Subject-Verb	0%	

Russell Tomlin, *Basic Word Order:Functional Principles*, (Croom Helm, London, 1986) page 22

The statistical data in table (1) also reveal another remarkable pattern.

The data reveal(s) that languages in which O and V are contiguous are highly preferred.

That is, basic constituent orders in which V and O are not separated by S is far more common.

Specifically, they are found in 365 (91 %) of the languages.

The close bonding between O and V as opposed to V and S or O and S has been recognized by the linguists in linguistics for very long time.

- In generative school of thought, this close association is formalized in some universal phrase structure rules.
- Although, the explanation that is put forward for these rules in the ‘generative paradigm’ would require a great deal of explanation and will also divert our main focus.
- However, they can also be presented in a simplified manner as given in (4) and can be explained for the understanding of constituent order as follows:
- (4)
- S →NP; VP
- VP→V ; NP

- S →NP; VP
- VP→V ; NP
- These two simple rules of Phrase Structure Grammar explain the organization of constituents in a transitive sentence.
- The first rule can be read as ‘a sentence consists of a noun phrase (which is the subject) and a verb phrase’.
- The second rule states that ‘a verb phrase consists of a verb and a noun phrase (object)’.
- The semicolon that occurs in the right-hand side of the rules indicate that the two constituents may occur in either order.
- Together, the PS-rules only generate the four following structures:
- (5) NP_{sub} V NP_{obj} [=SVO]
- NP_{sub} NP_{obj} V [=SOV]
- V NP_{obj} NP_{sub} [=VOS]
- NP_{obj} V NP_{sub} [=OVS]
-

What we noticed in the ordering of these phrase structure rules is very interesting.

It is very useful for the analysis and explanation of the constituent order in the languages.

The phrase structure rules, which are presumed to be the part of **innate capacity of human mind**, do not generate the remaining two OSV or VSO contitunent structures.

Therefore, some exceptional linguistic property would be required to motivate these two orders and, as a consequence, they are less common in terms of word-order of languages in the world.

Let us now put both Comrie & Chomsky's hypotheses and their findings together in a table and evaluate the constituent orders and their statistical values:

Comire + Chomsky = word-order phenomenon

Ordering principles for S,V and O

Word order	Subject Saliency	PS Rules
SVO	+	+
SOV	+	+
VSO	+	-
VOS	-	+
OVS	-	+
OSV	-	-

- We have noticed that the PS-rules in (4) do produce VOS and OVS sequences, and these are very uncommon word-orders.
- How can this fact be accounted for?
- If one assumes that the order of S, V, and O is sensitive not just to the phrase structure rules but also to other principles such as Comrie's notion of subject saliency, the rarity of some of the word-orders can easily be understood and explained.
- In order to demonstrate this, assume that just the two principles that have been discussed, they seem to interact in establishing preferred constituent orders in the languages of the world.
- Those constituent orders that adhere to both principles are most common, but those adhering to one of the principles are less common, however, the order which violates both principles, is extremely rare or nonexistent.
- That's all for today☺