Word order phenomenon in Languages

Source: Whaley 1997, Comrie 1981

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 grammar of the language which is commonly agreed upon by the native speakers.

The order of the words and the choice of re-arranging them in different order 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’ for that language.
- There could be one or more choices of arranging the words in the sentence and due to the choices there could be ONE mode of arrangement of the words which is popularly known as ‘Basic word order’.

- There seems to be a theoretical problem with this terminology.
- If we look at the data of different languages, we would realize that it is not SINGLE word which makes the subject, object or even the verb in all the cases in any language. For example,
- **John** is very funny.
- **The new student** is very funny.
- **That the SALA class is held in #101** is very funny.
- **The book about lady Diana's affair with the rich man of France** is banned.

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

Therefore, it is agreed upon by the scholars that the term ‘word order’ should be replaced with the ‘**constituent order**’.

What is a constituent then?

Well, a constituent is either a word or a group of words which makes the notional categories like SUBJECT AND OBJECT (i.e. DO, IO, OblObj etc.).

If you want a formal definition, it says ‘…. either a terminal node or a set of terminal nodes is a **CONSTITUENT** if they are dominated by the same node and no other terminal nodes are dominated by that node’.

- **The issue of word order:**
- There are languages in which the basic constituent order can be easily decided.
- For example, in English we find clauses with the orders OSV (1a) and VSO (1b), but it is quite clear that these orders are "special" and that the SVO order is typical (1c).
- Let us see the examples what we mentioned:
- 1. a. Sweet, I hate.
- b. Believe in you me, hum.. no way.
- c. The maid sliced the bread.

If we 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 distinctive intonation-al pattern in (1a), a slight pause after *sweet*. Indeed, without some discourse context, (1a) sounds odd.

The example (1b) also looks bound to some context. Moreover, it is not considered grammatical by all English speakers, because it is an idiom.

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

Many other languages do not have such strict word order that we just discussed in English (also in some other languages).

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 exist many options to arrange the words in a given sentence.

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

In Sanskrit, one could have the following sentence (s):

a.	balıkah	odənəm	pəčəṇtı
	girls-3FPl-Nom	food	cook-Imp-3FPl
	‘The girls cook food’.		

b.	odənəm	balıkah	pəčəṇtı
	food	girls-3FPl-Nom	cook-Imp-3FPl
	‘The girls cook food’.		

c.	balıkah	pəčəṇtı	odənəm
	girls-3FPl-Nom	cook-Imp-3FPl	food
	‘The girls cook food’.		

d.	pəčəṇtı	odənəm	balıkah
	cook-Imp-3FPl	food	girls-3FPl-Nom
	‘The girls cook food’.		

e.	pəčəṇtı	balıkah	odənəm
	cook-Imp-3FPl	girls-3FPl-Nom	food
	‘The girls cook food’.		

f.	odənəm	pəčəṇtı	balıkah
	food	cook-Imp-3FPl	girls-3FPl-Nom
	‘The girls cook food’.		

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 (2) &(3).

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 problem for the intended meaning in any of the sentences.

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.

Do you know why can't we have this 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) [source: Whaley: 1997].

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 that have been attested in the languages 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

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 with which SOV and SVO occur is amazing.

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

SOV and SVO are found in over 40% of the languages in the sample.

However, a total of these orders will make this figure go as high as 90%.

Clearly, then, the distribution cannot be taken as random, and some explanation must be determined for their 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 pattern which is shown in the Table (2):

Table (2):

Relative Frequencies of the Order of S + 0

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

Source: Whaley 1997.

- The figures in the above table depict the fact that there is startling difference amongst the languages that place the subject before the object (96%) than those that place the subject after the object (4%).
- This distribution was also noted by Greenberg (1966) and captured by his 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.*

Although this universal stipulates the linear precedence of subjects over objects, it does not explain why it should hold true in languages.

Comrie (1989, 93) suggests that this glaring priority of subject over object has a functional explanation.

Presumably, a deeper cognitive organization of information underlies the pattern.

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 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.

The statistical fact in table (1) also reveals another remarkable pattern.

The fact reveals 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 occur far more commonly.

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

The close bonding between V and 0 as opposed to V and S or O and S has been recognized by the linguists in linguistics for very long time.

- In Government and Binding Theory, this close association is formalized in universal rules of phrase structure.
- Although, the explanation that is put forward for these rules in Government and Binding would require a great deal of explanation, however, they can also be presented as in (4) and explained for the understanding of constituent order as follows:
- (4)
- S →NP; VP
- VP→V ; NP

These two simple rules of GB Theory explain the organization of constituents in sentences of different languages.

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 semicolons that occur in the right-hand side of the rules indicate that the two constituents may occur in either order.

Together, the rules generate the 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 notice 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 **innate aspects of the human language capacity**, do not generate OSV or VSO structures.

Therefore, some exceptional linguistic property would be required to motivate these two orders and, as a consequence, they are less common.

Let us now put both the findings together in a table and evaluate the constituent orders and their statistical values:

Chomsky + Comire= 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 these orders is understandable.
- In order to demonstrate this, assume that just the two principles that have been discussed interact in establishing preferred constituent order patterns.
- Those constituent orders that adhere to both principles are most common, those adhering to one of the principles less common, and the order which violates both principles, is extremely rare or nonexistent.