

September 9, 2005

CLAUSAL WORD ORDER

For the purposes of this handout, and future presentations, I generously stole from Usama Soltan, with whose generous permission I modified his course handouts for LING 419B: *Comparative Syntax* (University of Maryland, spring semester 2005).

1. Setting the stage for comparative syntax

1.1. Some “conceptual” background you need to know

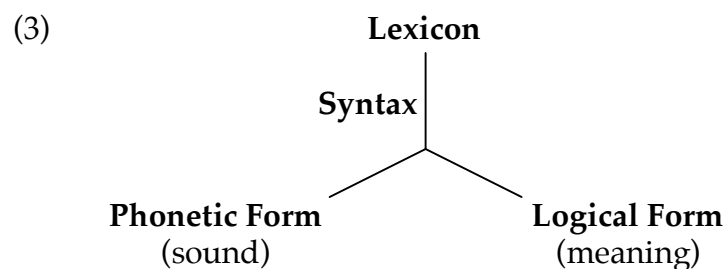
(1) *What is the object of study in theoretical linguistics?*

The simple answer is obvious: **Language**. Actually, it’s a bit more complex than that.

In principle, there are two ways of looking at language as an object of study: *externally* or *internally*. Viewed externally, the study of language makes no claims about the mind, knowledge, or mental state of the language user. Viewed internally, language is essentially a cognitive capacity, a system of knowledge in the mind of the speaker/hearer. In generative grammar, we are concerned with the study of **I-language** (for internal-language) — rather than **E-language** (for external-language).

(2) *But what constitutes an I-language?*

At least two components: A “list of words” and a device to “glue” these words together into sentences. We call the list of words a “**lexicon**” and the gluing device “**syntax**”. The syntax creates strings that are pronounceable on one end (Phonetic Form or **PF**) and interpretable on the other (Logical Form or **LF**). The model of grammar that creates this sound-meaning pairing is may be diagrammed as in 0:



Syntactic theory is then the theory that explains how lexical items are combined to form **PF- and LF-interpretable expressions** (sentences in the usual sense, but the process includes the generation of smaller chunks like phrases as well, of course).

Hypothesis 2:

A human language is the result of the interaction between universally invariant principles of grammar and a set of parameters whose (typically binary) values are fixed on the basis of the linguistic input around the learner.

The task of the linguist then is to identify these **principles and parameters (P&P)** — hence the name of the theory. Principles and parameters are what constitute (or so we think) the built-in **Universal Grammar (UG)** that we are all born endowed with.

1.2. Some “technical” background you need to know

- **Phrase structure theory:**
 - Constituency: How do we determine if a string of words is a constituent?
 - How do we represent constituency relations by phrase structure rules?
 - How do we represent constituency relations in tree structures?
 - Tree-geometric relations: dominance, sisterhood, c-command
 - X'-Theory: specifier, head, and complement; X'-schema
- **Lexical and functional categories:** NP, VP, PP, AP, IP, CP
- **θ-Theory:** predicates and arguments; θ-roles; the θ-Criterion
- **Case Theory:** The Case Filter

With a firm grasp of these principles of **syntactic analysis**, you will be well equipped to start looking at languages other than English (as we've done in previous courses). The interesting part will be to put these into action and discover parameters...

2. Patterns of word order across different phrasal categories

Historically and culturally (and in an interesting sense, linguistically) unrelated languages exhibit similar linguistic patterns and properties. For example, compare Navajo and Japanese with regard to word order within VPs, PPs, and NPs.

2.1. Position of object NP within VP

(9) ashkii [at'ééd yiyiiltsá]. *Navajo*
 boy girl saw
 'The boy saw the girl.'

(10) Jiro-ga [Akemi-o butta]. *Japanese*
 Jiro-NOM Akemi-ACC hit
 'Jiro hit Akemi.'

2.2. Position of object NP within PP

- (11) ['éé' biih] náásdzá. Navajo
 clothing into I.got.back
 'I got back into my clothes.'
- (12) Jiro-ga [Akemi to] [kuruma da] [Kobe ni] itta. Japanese
 Jiro-NOM Akemi with car by Kobe to went
 'Jiro went to Kobe by car with Akemi.'

2.3. Order of possessor and possessee within NP

- (13) [chidí bi-jáád] Navajo
 car its-leg
 'the wheel of the car'
 or 'the car's wheel'
- (14) [Jiro-no imooto-ga] sinda. Japanese
 Jiro-POSS sister-NOM died
 'Jiro's sister died.'

Now, let's compare the equally culturally and historically unrelated languages of **English and Edo** with regard to word order properties within VPs, PPs, and NPs.

2.4. Position of object NP within VP

- (15) a. Ozó [mién Adésúwá]. Edo
 Ozo found Adesuwa
 b. Ozo [found Adesuwa]. English

2.5. Positions of object NP within PP

- (16) a. Ozó rhié néné èbé [né Adésúwá]. Edo
 Ozo gave the book to Adesuwa
 b. Ozo gave the book [to Adesuwa]. English

2.6. Order of possessor and possessee within NP

- (17) a. [omó Ozó] rré. Edo
 child Ozo came
 b. A [child of Ozo's] came. / [Ozo's child] came. English

We could, of course, speculate about the **intrinsic relationships and similarities** between these languages (or lack thereof). But we'll try something more interesting.

Greenberg's Universals 2 and 4 (from Greenberg 1963):**Universal 2:**

In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it almost always precedes.

Universal 4:

With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional.

The **correlations** are interesting in themselves and they call for an **explanation** (which will discuss in detail next class), but what makes them even more significant is that **their reverse possibilities are unattested** (or at best, rare). If the correlations are random rather than rule-governed, we should expect the following sentence structures to be attested, at least with equal frequency, contrary to typological facts:

- (18) a. * Chris put the book the table on. (VO and postpositions)
 b. * Chris the book on the table put. (OV and prepositions)

Summary:

Even though human languages exhibit different word orders within clauses and phrases, the **possibilities of variation seem to be limited** and suggest the presence of underlying (or “universal” in the sense of Greenberg) principles that govern what patterns of word order can or cannot occur together in the same language. We will (might?) get back to discuss word order in detail next week.

3. Distinguishing grammatical functions cross-linguistically

Grammatical functions refer to notions such as “**subject of**” and “**object of**”. In addition to reliance on word order to distinguish subjects from objects, as the case is, for example, in the English sentence “*John loves Mary*”, human languages utilize other strategies to distinguish between subjects and objects. Two major strategies are discussed here.

Consider the following data from **Japanese and Mohawk**:

- (19) Sak Uwári **shako-núhwe**'s. *Mohawk*
 Jim Mary he.her-likes
 'Jim likes Mary.'
- (20) Sak Uwári **ruwa-núhwe**'s. *Mohawk*
 Jim Mary she.him-likes
 'Mary likes Jim.'

- (21) Jiro-**ga** Akemi-**o** butta. *Japanese*
 Jiro-NOM Akemi-ACC hit
 'John hit Akemi.'

To **distinguish grammatical functions**, Japanese uses certain markers on subject and object NPs (= **case**). By contrast, Mohawk uses certain markers on the verb itself that signal which NP is the subject and which NP is the object (= **agreement**). In technical terminology that is due to Nichols (1986), Japanese is said to be a *dependent-marking* language, whereas Mohawk is classified as a *head-marking* language.

But what is equally interesting is that the type of marking that one language uses for the verb and its dependents is the same type that it uses in **other contexts** as well. Consider how Japanese and Mohawk distinguish possessors from possessees:

- (22) Sak **rao**-wise' *Mohawk*
 Jim his-glass
 'Jim's glass'
- (23) Jiro-**no** imooto-**ga** *Japanese*
 Jiro-GEN sister-NOM
 'Jiro's sister'

And once again, **languages with mixed systems of marking are either unattested** (or rare, at best). This reinforces further our earlier conclusion that linguistic variation is subject to deep principles that — parameterized or parametrically — govern which properties may or may not occur together in natural languages.

4. Variation across some “related” languages and the status of “subject”

Now, let's consider some (historically and culturally, but also linguistically) **related languages**: English, French, and Italian (as well as Greek). One syntactic similarity between English, French, and Italian (or Greek) is that they all have SV sentences:

- (24) a. John will arrive. *English*
 b. Jean arrivera. *French*
 Jean will.arrive
 c. Gianni verrà. *Italian*
 Gianni will.come
 d. O Yiannis tha fthasi. *Greek*
 the Yiannis will arrive

Unlike English and French, however, **Italian (and Greek) also allow VS orders**:

- (25) a. * Will arrive John. *English*
 b. * Arrivera Jean. *French*
 will.arrive Jean
 c. Verrá Gianni. *Italian*
 will.come Gianni
 d. Tha fthasi o Yiannis. *Greek*
 will arrive the Yiannis

A second difference between Italian/Greek on one hand and English/French on the other is that Italian/Greek **allow the subject of a tensed sentence to be omitted** when it's understood from the discourse. This is not possible in English or French:

- (26) a. * Will arrive. *English*
 b. * Arrivera. *French*
 will.arrive
 c. Verrá. *Italian*
 will.come.
 d. Tha fthasi. *Greek*
 will arrive

A third difference between English and French on the one hand and Italian and Greek on the other has to do with the **subject-object asymmetry** with regard to *wh*-extraction out of embedded clauses (cf. (5) above):

- (27) a. Who did you say that Chris saw __ in the park? *English*
 b. * Who did you say that __ saw Chris in the park?

In this respect, **French patterns just like English**:

- (28) a. Qui veux-tu que Marie épouse __? *French*
 who want-you that Marie marries
 b. * Qui veux-tu que __ épouse Jean?

(27b) can be fixed in English by **omitting the complementizer *that***, whereas (28b) can be made grammatical in French by **changing the complementizer *que* to *qui***:

- (29) a. Who did you say __ saw Chris in the park? *French*
 b. Qui veux-tu qui __ épouse Jean? *French*
 who want-you that marries Jean

But now compare extraction of subject *wh*-phrases in **Italian**:

(30) Chi credi che __ verrà? Italian
 who you.think that will.come

(30) is perfectly grammatical in Italian. **Extraction of embedded subject *wh*-phrases does not cause any grammaticality problems in Italian**, then. (Q: *What about Greek?*)

Let's **summarize the differences** between English/French and Italian/Greek:

- Italian and Greek allow **postverbal subjects**, but English and French do not.
- Italian/Greek allow **subjects of tensed sentences to drop**, English/French do not.
- Italian/Greek freely allow **extraction of subject *wh*-phrases from embedded clauses**, but English/French only allow such extraction under certain conditions.

Main question

Is there any chance that these “surface” syntactic differences between the four languages are related? What can the P&P-framework say about this instance of cross-linguistic variation?

As it turns out, there seem to be strong evidence that **such properties are indeed interrelated**. Here's why and how.

First, Spanish and Romanian behave exactly like Italian (and Greek) with regard to the three properties above, so it cannot be an accident that such properties occur or do not occur together in natural languages.

Second, Old French was exactly like Italian, i.e. it has all the properties discussed above. At one point in its historical development, though, French became more like English and thus lost all these three properties. This makes sense only if there is a correlation in the occurrence or lack of occurrence of such syntactic properties.

Third, the properties are interrelated in that they all involve the notion “subject of the sentence” in one way or another. To see how, let's introduce a fourth difference between English/French and Italian/Spanish (leaving Greek to your reflections): the obligatoriness of a “dummy” subject (expletive) with weather verbs ('meteorological verbs') in both English and French as opposed to its absence in Italian and Spanish:

(31)	a.	It is raining.	<i>English</i>
	b.	Il pleut. it rains	<i>French</i>
	c.	Ø Piove. rain.3.SG	<i>Italian</i>
	d.	Ø Lleuve. rain.3.SG	<i>Spanish</i>

The **basic difference** between the two types of language can be expressed as follows:

In some languages (e.g., French, English, German, Edo) every tensed clause must have an overt subject. In other languages (e.g., Italian, Spanish, Romanian, Greek, Navajo, Arabic) tensed clauses need not have an overt subject. This is typically referred to as the *null subject parameter*.

Now, how does the null subject parameter help us **explain the differences** between English/French and Italian/Spanish?

First, it trivially explains why subjects can be omitted in Italian/Spanish (and Greek, among many other languages), but not in English/French (and like languages).

Second, it also explains why subjects may appear postverbally, since the preverbal position need not be filled in Italian/Spanish, but it has to in English/French, as sentences with weather verbs show.

- but now consider the following sentences from English and French:

(32) Il est arrivé **trois hommes**. *French*
 it is arrived three men
 ‘There arrived three men.’

(33) There appeared **a boat** on the horizon. *English*

The boldfaced “**logical subject**” may actually appear postverbally in both English and French, but even then the preverbal position has to be filled — in this case, by a “dummy” element: *there* in English and *il* in French (viz. “expletive-associate pair”).

Third, the subject-object asymmetry with regard to *wh*-extraction out of embedded clauses can also be explained, though perhaps further assumptions are needed here. Baker (2001) says that extraction of embedded subjects leaves behind a tensed sentence with no overt subject, an option prohibited in English/French, but tolerated in Italian/Spanish (and Greek and so on).

- this is not quite accurate — can you see why?

A **fourth** consequence of this parametric analysis is that we can also explain why object extraction does not cause ungrammaticality in any of the languages discussed here. The reason is simple: there’s no requirement on having “overt” objects in sentences. For example, English speakers don’t have to say:

(34) *It rains it. *English*

Finally, and most importantly, the parametric analysis relates four syntactic differences between English/French and Italian/Spanish (yes, and Greek...) that would seem at first blush quite unrelated:

- **occurrence versus non-occurrence** of postverbal subjects;
- **omission versus non-omission** of overt subjects of tensed clauses;
- **extraction versus non-extraction** of subject *wh*-phrases out of embedded clauses;
- **presence versus absence** of overt subjects with weather verbs.

On deeper analysis, these differences all turn out to be reducible to one fundamental parametric difference: **the setting of the null subject parameter**. The parametric analysis thus captures correlations that would otherwise be treated as accidental or remain unaccounted for.

Moral of the story

There are at least four differences we discussed here between English/French on the one hand and Italian/Spanish on the other. There's two more that Baker (2001) did not discuss. This is six in total. If these six differences are not interrelated, **we should expect 2⁶ possible Romance languages**, but actually we only find 2 such languages: French-type languages and Italian-type languages. Equally significant is the fact that Old French was just like Italian, but when it changed it became just like English. This makes sense only if the "six" surface differences are all reflexes of one "single" difference at a deeper level of linguistic analysis. The parametric approach to cross-linguistic variation does exactly that by attributing these "surface" differences to the positive or negative setting of the null subject parameter in a particular language.

5. The head parameter

Next we're going to explore some **explanations** of what we see in terms of word order variation. We will look in some depth at the head parameter (to be continued).

word order	# of languages	%
SOV	180	45
SVO	168	42
VSO	37	9
VOS	12	3
OVS	5	1
OSV	0	0

Table 1: Frequencies of basic word orders in natural languages (Tomlin 1986: 22)

5.1. The SVO-SOV contrast: English vs. Japanese

- (35) The child might think that she will show Mary's picture of John to Chris.
- (36) Jiro-ga Hiro-ga Hanako-ni zibun-no syasin-o miseta to omette iru.
 Jiro-NOM Hiro-NOM Hanako-DAT self-POSS picture-ACC showed that thinking be
 'Taro thinks / is thinking that Hiro showed a picture of himself to Hanako.'

Element A	Element B	English relation	Japanese relation
Verb	Direct object	A precedes B	A follows B
Verb	Pre/post-position phrase	A precedes B	A follows B
Verb	Embedded clause	A precedes B	A follows B
Pre-/post-position	Related noun phrase	A precedes B	A follows B
Noun	Related pre/post-phrase	A precedes B	A follows B
Complementizer	Embedded clause	A precedes B	A follows B
Auxiliary	Main verb	A precedes B	A follows B

Table 2: Word order relationships in English and Japanese (Baker 2001: 60)

- an **English-type language**: *Edo*

- (37) Ozó má tá wéé írén ghá rhièéfótò Uyì yè néné ékpétín.
 Ozo not say that he will put photo Uyi in the box
 'Ozo did not say that he will put a photo of Uyi in the box.'

- an **Japanese-type language**: *Lakota*

- (38) John wowapi k'uhe oyuke ki ohlate iyeye.
 John letter that bed the under found
 'John found that letter under the bed.'

- other SVO languages like English/Edo: Thai, Khmer, Indonesian, Zapotec, Salish
- other SOV languages like Japanese/Lakota: Turkish, Basque, Navajo, Quechua (as well as the languages of the Eskimos)

VO correlate	OV correlate
adposition – NP	NP – adposition
copula verb – predicate	predicate – copula verb
“want” – VP	VP – “want”
tense / aspect auxiliary verb – VP	VP – tense / aspect auxiliary verb
negative auxiliary – VP	VP – negative auxiliary
complementizer – S	S – complementizer
question particle – S	S – question particle
adverbial subordinator – S	S – adverbial subordinator
article – N'	N' – article
plural word – N'	N' – plural word
noun – genitive	genitive – noun
noun – relative clause	relative clause – noun
adjective – standard of comparison	standard of comparison – adjective
verb – PP	PP – verb
verb – manner adverb	manner adverb – verb

Table 3: Correlation pairs in VO and OV languages (Dryer 1992)

5.2. Towards an explanation

Main Question:

How do we account for all these correlations?

Remember the X'-schema? There lies the solution:

(39) $XP \rightarrow \text{Specifier } X'$

(40) a. $X' \rightarrow X \text{ Complement}$
 b. $X' \rightarrow \text{Complement } X$

There are two options then for the expansion of X': Either the head precedes its complement (40a), or the head follows the complement (40b). This is the so-called *head directionality parameter*.

An in-class exercise:

Draw two tree structures for the English and Japanese sentences in (35) and (36).

5.3. The position of subject: SVO and SOV languages vs. VOS and OVS languages

English and Japanese share the property of being **subject-initial** languages:

- (41) Jiro-no imooto-ga sinda. *Japanese*
 Jiro-POSS sister-NOM died
 'Jiro's sister died.'

Our formulation of the head directionality parameter in (39)-(40) does not say anything about the **position of specifiers** with regard to the head of a phrase. Shouldn't we expect similar parameterization there as well?

This is what Baker calls the *subject side parameter* — something along these lines:

- (42) *Subjects may occur initially or finally in the sentence.*

- Tzotzil (and Malagasy): **Head-initial and subject-final** (VOS)

- (43) 7i-yal-la ta te7 ti vinik-e. *Tzotzil*
 PAST-descend from tree the man
 'The man came down from the tree.'

- (44) stzek li 7antze *Tzotzil*
 skirt the woman
 'the woman's skirt'

- Hixkaryana: **Head-final and subject-final** (OVS)

- (45) kanawa yano toto. *Hixkaryana*
 canoe took person
 'The man took the canoe.'

But look at Table 1 now. **How many subject-last languages do we find there?** Any reason why such languages are so rare?

5.4. OVS reconsidered

According to Baker (2001), **OVS languages are not rare; they simply don't exist.** There is evidence from Hixkaryana that indicates that the OVS order is "*derived*" rather than "*basic*" (i.e. what we see is *not* what we get or started out from).

First, indirect objects consistently come after the subject:

- (46) otweto yimyakoni rohetxe totokomo wya. Hixkaryana
 hammock gave my.wife people to
 'My wife used to give hammocks to the people.'

Second, Hixkaryana exhibits SOV order in nonfinite embedded clauses:

- (47) ro-wy wewe yamatxhe itehe harha owo hona. Hixkaryana
 me-by tree after.felling I.go back village to
 'After I fell the tree, I will go back to the village.'

Conclusion:

OVS orders not basic; rather they are **derived by VP-movement**, along the lines suggested by Kayne (1994). But why are these languages rare? Well, if VP-movement is rare, then that's why. Are you wondering as I am?

5.5. VOS reconsidered

But now going back to Tzotzil (and Malagasy), **why do such VOS languages exist then?** Baker's answer lies in his concept of "hierarchy" of parameters, which you can see on the tree I gave to you in class last week. As you can see there, the subject side parameter is only relevant to head-initial languages, but not to head-final languages. But why are these languages rare? Can you tell from the parameter hierarchy?

Why is OSV order unattested in natural languages?

According to Baker, these are out by the **verb-object constraint**, which requires verbs to merge with their objects prior to the merge of the subject.

6. Languages with apparent "mixed" head directionality

6.1. Nupe and Amharic

- (48) a. mi kpaye gánán Musa lá ébi. Nupe
 I think COMP Musa took knife
 'I think that Musa took the knife.'
 b. èbi Musa lá o.
 knife Musa took COMP
 'It's a knife that Musa took.'
- (49) Aster wənbər-u-n lə-setiyyə-wa sat't'aciw. Amharic
 Aster chair-the-OBJ to-woman-the gave
 'Aster gave the chair to the woman.'

3.3. German and Dutch

- (50) a. Peter küsste die Frau. SVO
 Peter kissed the woman
 'Peter kissed the woman.'
- b. Küsste Peter die Frau? VSO
 kissed Peter the woman
 'Did Peter kiss the woman?'
- c. Ich denke, dass Peter die Frau geküsst hat. (C)SOV
 I think that Peter the woman kissed has
 'I think that Peter kissed the woman.'
- (51) a. De kat drinkt melk. SVO
 the cat drinks milk
 'The cat drinks milk.'
- b. Waarom drinkt de kat melk? (XP)VSO
 why drinks the cat milk
 'Why does the cat drink milk?'
- c. Ik zie dat de kat melk drinkt. (C)SOV
 I see that the cat milk drinks
 'I see that the cat drinks milk.'

The "standard" analysis:

German and Dutch are **underlyingly SOV**, with other orders being derived by movement operations.

Problems:

German and Dutch have **prepositions** (though postpositions also do occur in both languages), and their Ds occur initially in DPs:

- (52) a. auf dem Tisch German
 on the table
 'on the table'
- b. op de berg Dutch
 on the mountain
 'on the mountain'

One possible solution:

Sub-parameterize the head directionality parameter with regard to the category involved — something along the lines below:

(53) *X' is head-initial unless X is a [+V] category.*

A similar solution was suggested for **Chinese** in Huang (1994), though with the opposite directionality:

(54) *X' is head-final unless X is a [+V] category.*

(55) ta pian-le Lisi. Chinese
 he cheated Lisi
 'He cheated Lisi.'

What kind of problems will arise with this type of parameterization? See Travis' (1989) alternative solution for Chinese and Kpelle. (We will discuss the head parameter in more detail soon — and this would also make a nice class presentation.)

A more radical solution:

There is no head parameter. Kayne's (1994) antisymmetry approach to word order variation follows this line of reasoning. We will talk about that in future classes. For starters, you may want to familiarize yourself with Kayne's approach by reading the section on "A Universal word order?" (Roberts (1997: chap. 1), in the course reader).

References

- Baker, Mark C. 2001.** *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York: Basic Books.
- Chomsky, Noam. 1986.** *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.
- Dryer, Matthew S. 1992.** The Greenbergian Word Order Correlations. *Language* 68: 81-138.
- Greenberg, Joseph. 1963.** *Universals of Language*. Cambridge, MA: MIT Press.
- Huang, C.-T. James. 1994.** More on Chinese Word Order and Parametric Theory. In Barbara Lust, Margarita Suárez, and John Whitman (eds.), *Syntactic Theory and First Language Acquisition: Cross-Linguistic Perspectives*. Hillsdale, NJ: Erlbaum, 15-35.
- Kayne, Richard S. 1994.** *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
- Nichols, Johanna. 1986.** Head-Marking and Dependent-Marking Grammar. *Language* 62: 56-119.
- Roberts, Ian. 1997.** *Comparative Syntax*. London: Arnold.
- Tomlin, Russell S. 1986.** *Basic Word Order: Functional Principles*. London: Croom Helm.
- Travis, Lisa. 1989.** Parameters of Phrase Structure. In Mark R. Baltin and Anthony S. Kroch (eds.), *Alternative Conceptions of Phrase Structure*. Chicago, IL: University of Chicago Press, 263-279.