

Linguistic Theory and Terraling SSWL_1214 (Selection)

Hilda Koopman
koopman@ucla.edu

University of California, Los Angeles

UCLA/NYU at Abu Dhabi
January 5-9 2015

Language Diversity: the Code Talker's Paradox

(i) Languages must be similar enough that **any normal child** can learn any one of them...

(... and that the Code Talkers were able to translate!)

(ii) ... and yet can substantially differ

(... so much that the Japanese couldn't break the Code!)

Universal Grammar based Diversity

■ Language develops in the framework of **innate learning biases, Universal Grammar, a blueprint** which guides the acquisition of language by children. All human languages obey UG.

■ But there are all sorts of **structural** differences between languages.

→ Universal Grammar includes **parameters**, whose value is fixed on the basis of the linguistic environment which each control a variety of phenomena.

+ From the point of view of UG, the **vocabulary** is arbitrary.

So far

- 1 The object of study: why and how do we study it? Reverse engineering problem: what are the design specifications?
- 2 We know rules.. (recursion, question formation etc)
- 3 linear order is not relevant for rules. What counts is abstract hierarchical structure
 - example 1: yes no question formation in English Point: simple rule based on linear order does not work.
poverty of the stimulus and language acquisition
- 4 Inside the black box: Words form constituents
 - concrete illustration: (a few) constituency tests.
- 5 Structure encodes meaning.
Probing structural ambiguity. → structure encodes meaning: different meanings imply different structures.

Roadmap *continued*

- 1 The recipe for structure building: how to form phrases?
 - finite state grammar? (*problems..*)
 - phrase structure grammar? (*yes, more like it*), but + movement, and beyond:
 - X-bar theory? *parameters*(head, complement, subject/specifier)
- 2 REMAINING:)
Structure and referential dependencies (plays a role in Baker chapter 4))

Where we are at: Rules of syntactic combination (UG and parameters)

- 1 manipulate words
- 2 words (or subword morphemes) are assembled into complexes (molecules)
- 3 are recursive (complexes can be further combined into bigger units and produce infinitely long strings).
- 4 produce tree-like structures that encode constituency
- 5 interpretation is based on tree structure (compositionality)
- 6 Internal organization of phrases obeys X-bar schema
- 7 use notions like head, complement, "subject"/specifier
- 8 Head: determines what must be present in its phrase
- 9 Belongs to a category (label) (V, N, P, T, C...) (determines the distribution of the phrase)
- 10 *see next page*

...continued

- Reordering: Movement. a particular head or phrase can satisfy more than one property
- head movement T to C. (English yes/no questions) (N to V in Mohawk)
- wh- question formation. TO COME UP:
- Constituent structure constrains referential dependencies (anaphors (himself), pronouns(him) and names/referential expressions (Jacques, the woman) (Baker Chapter 4)
- and parameters

Parameters and parameter clusters

- null subject parameter
- word orders parameters: headedness, . and exploration of Terraling

The Null subject parameter

- Some languages allow null subjects, some languages don't:

A. Null Subjects

- (1)
- a. parla 'he/she speaks (Italian)
 - b. parla (Catalan)
 - c.* parle (French)
 - d.*speaks (English)

- Italian and Catalan allow null subject (pro-drop languages); French and English don't.

- Italian and Catalan allow null subjects (pro-drop languages); French and English don't.

A. Null S?

Italian	Yes
Catalan	Yes
French	No
English	No

- Italian and Catalan allow null subjects (pro-drop languages); French and English don't.

A. Null S?

Italian	Yes
Catalan	Yes
French	No
English	No
Navajo	yes
Mohawk	yes
Korean	yes

Person marking on V?

- Person Morphology

	1S	2nd sing	3rd sing	1st plu	2ndplu	3rd plu
It.	parl-o	parl-i	parl-a	parl-iamo	parl-ate	parl-ano
Cat	parl-o	parl-es	parl-a	parl-em	parl-eu	parl-en
Fr	parl	parl	parl	parl-ō	parl-é	parl
Eng	speak	speak	speaks	speak	speak	speak

- Italian and Catalan have rich person morphology
French and English have an impoverished person morphology.
Navajo?, Mohawk?, Korean?, Kinyarwanda, Kasakh,
Romanian, Arabic, Afrikaans?, Zulu, ..?

	A. Null subjects?	B. Rich pers morph?
It	Yes	Yes
Cat	Yes	Yes
Fr	No	No
Eng	No	No
Navajo	yes	yes
Mohawk	yes	yes
Mandarin	no	no
Korean	yes	no

Allows subjects to follow the Verb (VS) ?

- (2)
- a. ha telefonato Gianni (Italian)
 - b. ha telefonat en Joan (Catalan)
 - c. *a téléphoné John (French)
 - d. *telephoned John (English)
-
- 4 languages are SV.
 - but only in Catalan and English may the subject also follow (VS)

C. Allows VS (in addition to SV)

	A. Null subjects?	B. Rich pers morph	C. V S? (SV and)
It	Yes	Yes	Yes
Cat	Yes	Yes	Yes
Fr	No	No	No
Eng	No	No	No

All 4 languages allows fronting of an object *who* from an embedded that-clause

Fronting of an object *who*:

- (3)
 - a. I think that Mary will call someone
 - b. *who* do you think that Mary will call *who*
 - c. *who* do you think that Mary will call _

- No changes between decl and questions in: *...think that Mary will call ...*

But extraction the subject *who* from an embedded that clause

Fronting of subject *who*:

- (4) a. I think [that [someone will call Mary]]
- b. who do you think [that [who will call Mary]]
- c. *who do you think [that [_ will call Mary]]

something special needs to happen! (that must disappear)

- (5) * who do you think [~~that~~ [_ will call Mary]]

- (6) Chi credi che telefonerà?
who believ.2 that call.fut? Italian S
- (7) Qui creus que telefonarà?
who believ.2 that call? Catalan S
- (8) qui tu crois que Marie appellera _?
who you believe that Marie call.fut
who do you believe Marie will call? French O
- (9) *qui tu crois que appellera Marie?
who you believe that meet.fut Marie? French S
- (10) qui tu crois *qui* appellera Marie?
who you believe *qui* will call Mari? Special *qui*

	Null subjects?	Rich pers morph	V S? (SV and)	fronting emb S who?
Italian	Yes	Yes	Yes	Yes
Catalan	Yes	Yes	Yes	Yes
French	No	No	No	No
English	No	No	No	No

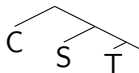
	Null subjects?	Rich pers morph	V S? (SV and)	fronting emb S who?
Italian	Yes	Yes	Yes	Yes
Catalan	Yes	Yes	Yes	Yes
French	No	No	No	No
English	No	No	No	No
Navajo	yes	?	no	?

	Null subjects?	Rich pers morph	V S? (SV and)	fronting emb S who?
Italian	Yes	Yes	Yes	Yes
Catalan	Yes	Yes	Yes	Yes
French	No	No	No	No
English	No	No	No	No
Navajo	yes	?	no	?
Mohawk	yes	yes	yes	?yes

- but why?

V [that t: [* C _]]

- Hypothesis: Extraction of the preverbal subject position of a complement clause is disallowed (UG?).



- Languages must use other resources to allow subjects to be questioned
- Extraction from the post verbal subject (V S) is OK (Italian, Catalan)
- or doing something to the C. (English: that deletion, French: use the relative pronoun qui)
- Vata: (Kru)– putting in a pronoun in subject position. ?other ways?

- (11) Current state of our knowledge about the null subject parameter? Does it hold up? Does headedness matter? It needs to be put to the test to as many cases as we can. Find a way to code the variation systematically: SSWL.

Traveling and doing this for as many languages as we can

Testing theories! Building a database..Terraling

- How do kids do it?
What's in the black box
We need a theory!
- ... we need to build a database that codes linguistic variation (both historical and synchronic) in the relevant dimensions and fine detail to support theories.
such a database does not currently exist, so we are developing one. (genomics)
rightarrow SSWL/Terraling and explore the tools.

Correlations? implications? Comparisons?

- Besides cataloging the rich patterns that exist in each language, we need to find out if combinations that are predicted to be excluded do indeed fail to occur. (Need tools to explore the database).
- Could similarities be accidental? Or are they reflection of the system?
- How should we conceive of mixed headedness languages? How mixed are languages? Are these exceptional or frequent?
- Functionalists versus Formalists
we don't know, but if we don't ask questions we will never find out.
As a native speaker of a mixed language: nothing feels unstable about it.
(in fact, I could give you an idea on how to turn English into Dutch or vice versa)

The linguistic explorer database: SSWL

- original prototype: (as of January 2nd)
SSWL: Site Statistics
Number of Languages: 252
Number of Languages over 90%: 22
Number of Contributors: 363
Number of Properties: 112
Number of Examples: 3757
Number of Property:Value Pairs: 16001

snapshot of SSWL on terraling" SSWL1214

- Languages below 50% removed, "expert crowd sourcing" project.
- Open ended and build to grow:
- in build search functionality
- Data.. Generated by "property definitions" ((ideally) written by community) . (Property Value: Yes/ No/ NA).
- How to.? Codes data that linguists base their theories (word order).
codes: existing variation (no notion of dominant order used by typologists, WALS)
not: head parameter: H Compl Yes/No
not: O in VP
not: polysynthetic?
yes: anaphor in subject position can have a non-local antecedent, etc.
- values: set by "linguistic experts"
illustrated with examples, comments (where appropriate), references.
- Download; simple searches, combining searches, compare, implications, cross, visualize (similarity tree),...

The Head parameter(s): parameters of X-bar theory closely connected to Word order typology

- Greenberg 1966 45 Universals of Language (sample of 30 languages)
 - From looking for shared histories, or shared cultures to searching for Universals of Languages.
 - Syntactic Typology. (and the WAL database (World Atlas of Language Structures)). suggestive of the found diversity, but not usable for formal linguistics purposes. (variation is hidden, data entry).
 - studied surface patterns in languages. (E-language: examples of patterns in Language X, or I-language: the recipes underlying language X)
not the recipes of how these are formed (simply not known at the time).

I language and E language

- (12) a. samples of even numbers 2,4,6,8,...: Extensional characterization. E-languages
b. for making even numbers $x : 2y$, *yaninteger* Intentional characterization of the set. I-language

Chomsky (1986): E-language and I language.

E for External (and suggestive of Extensional), and I for Internal, Individual (and suggestive of intentional).

I language: Knowledge of a Language X for an individual speaker (UG, and parameters), with slight differences between speakers. E language: sentences that result from (many different) I languages.

I language and E language

- (12) a. samples of even numbers 2,4,6,8,...: Extensional characterization. E-languages
- b. for making even numbers $x : 2y$, *yaninteger* Intentional characterization of the set. I-language

Chomsky (1986): E-language and I language.

E for External (and suggestive of Extensional), and I for Internal, Individual (and suggestive of intentional).

I language: Knowledge of a Language X for an individual speaker (UG, and parameters), with slight differences between speakers. E language: sentences that result from (many different) I languages.
differences in input.

Every speaker creates his own I language.

Linguists

- **Typologists** (broadly construed): E language, do not assume Chomsky's I-language/UG/movement
- **Formalists**: I language.
- **Functionalist** explanations (historical accidents, ease of use, ease of understanding, better for communication, simpler system..., general cognitive strategies,..) versus formal explanations (independently motivated ingredients taken together account for the puzzles (modular accounts)).

Fields are determined by the questions you ask about them.

Basic word orders: Patterns of S, V, O in SSWL1214

Search: Cross (up to 6) (not a balanced sample)

05 SVO	1	06 SOV	0	07 VSO	0	08 VOS	0	09 OSV	0	10 OVS	0	57
05 SVO	1	06 SOV	1	07 VSO	0	08 VOS	0	09 OSV	0	10 OVS	0	24
05 SVO	0	06 SOV	1	07 VSO	0	08 VOS	0	09 OSV	0	10 OVS	0	15
05 SVO	1	06 SOV	1	07 VSO	1	08 VOS	1	09 OSV	1	10 OVS	1	14
05 SVO	1	06 SOV	0	07 VSO	1	08 VOS	1	09 OSV	0	10 OVS	0	8
05 SVO	0	06 SOV	0	07 VSO	1	08 VOS	0	09 OSV	0	10 OVS	0	6
05 SVO	0	06 SOV	1	07 VSO	0	08 VOS	0	09 OSV	1	10 OVS	0	5
05 SVO	0	06 SOV	0	07 VSO	1	08 VOS	1	09 OSV	0	10 OVS	0	5
05 SVO	1	06 SOV	0	07 VSO	0	08 VOS	1	09 OSV	0	10 OVS	0	4
05 SVO	1	06 SOV	0	07 VSO	1	08 VOS	1	09 OSV	0	10 OVS	1	3
05 SVO	1	06 SOV	1	07 VSO	1	08 VOS	0	09 OSV	0	10 OVS	0	3
05 SVO	1	06 SOV	1	07 VSO	0	08 VOS	0	09 OSV	1	10 OVS	1	3
05 SVO	1	06 SOV	1	07 VSO	0	08 VOS	1	09 OSV	1	10 OVS	1	2
05 SVO	1	06 SOV	1	07 VSO	1	08 VOS	0	09 OSV	0	10 OVS	1	2
05 SVO	1	06 SOV	1	07 VSO	0	08 VOS	0	09 OSV	1	10 OVS	0	1
05 SVO	0	06 SOV	0	07 VSO	0	08 VOS	0	09 OSV	0	10 OVS	1	1
05 SVO	1	06 SOV	1	07 VSO	0	08 VOS	1	09 OSV	0	10 OVS	0	1
05 SVO	1	06 SOV	0	07 VSO	1	08 VOS	0	09 OSV	0	10 OVS	1	1
05 SVO	1	06 SOV	0	07 VSO	0	08 VOS	0	09 OSV	0	10 OVS	1	1
05 SVO	1	06 SOV	1	07 VSO	0	08 VOS	0	09 OSV	0	10 OVS	1	1
05 SVO	1	06 SOV	1	07 VSO	0	08 VOS	0	09 OSV	0	10 OVS	1	1
05 SVO	1	06 SOV	0	07 VSO	1	08 VOS	0	09 OSV	0	10 OVS	0	1
05 SVO	1	06 SOV	1	07 VSO	1	08 VOS	0	09 OSV	1	10 OVS	1	1
05 SVO	1	06 SOV	0	07 VSO	0	08 VOS	0	09 OSV	1	10 OVS	0	1
05 SVO	1	06 SOV	0	07 VSO	0	08 VOS	0	09 OSV	0	10 OVS	1	1

(13) 160 languages. (finds only those with all values given)

- a. Bias for SO over O(..)S.
- b. Potential candidates for Polysynthetic languages
Greek (Ancient), Bardi, Digor Ossetic, Eastern Armenian, Georgian, Iron Ossetic, K'iche'. Kiyaka, Latin, Pima, Tlingit, Turkish, Warlpiri, Western Armenian, *Mohawk is entered, but not found by this algorithm (it lacks an iso code, and latlong?, but Nahuatl (classical) is missing! (not complete, SOV was hard to decide, and anyway OSV is no . Compare Nahuatl (classical), with Nahuatl (Central Huasteca).*

Greenberg Universal 1 "In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object."

(14) bias SO over OS.

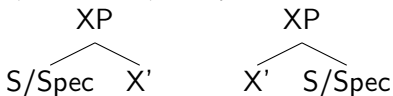
why? expected?

Greenberg Universal 1 "In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object."

(14) bias SO over OS.

why? expected?

- (15) a. S combines with VO/OV.
b. This bias is NOT expected under X-bar theory (Subject/Spec: left/right).



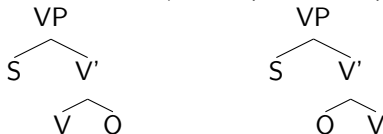
functionalists/ formalists.
which languages are in the database

How many types of languages: a simplified example

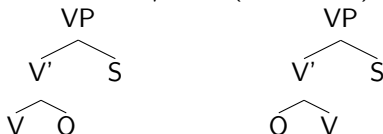
How many S,O, V language types can we generate with just the X-bar schema and its parameters? 4/6

(16) hit (agent, theme). UG: theme combines with V before the agent does

Subject initial with V initial/final: (SVO, SOV)



Subject final with V initial/final: (VOS, OVS)



(17) What about VSO and OSV? Do they occur? (yes!)

VSO and OSV

05 SVO	1	06 SOV	1	07 VSO	1	08 VOS	1	09 OSV	1	10 O
05 SVO	1	06 SOV	0	07 VSO	1	08 VOS	1	09 OSV	0	10 O
05 SVO	0	06 SOV	0	07 VSO	1	08 VOS	0	09 OSV	0	10 O
05 SVO	0	06 SOV	1	07 VSO	0	08 VOS	0	09 OSV	1	10 O
05 SVO	0	06 SOV	0	07 VSO	1	08 VOS	1	09 OSV	0	10 O
05 SVO	1	06 SOV	0	07 VSO	1	08 VOS	1	09 OSV	0	10 O

VSO and OSV?

- (18) but VSO and OSV cannot be generated by the UG rules so far.
WHY.

VSO and OSV?

- (18) but VSO and OSV cannot be generated by the UG rules so far.
WHY.

How to resolve the problem?

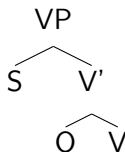
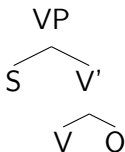
VSO and OSV?

- (18) but VSO and OSV cannot be generated by the UG rules so far.
WHY.

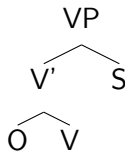
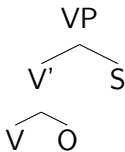
How to resolve the problem?

Reordering (aka *Movement*) must be involved. But how?

How to get VSO by movement?



Subject final with V initial/final: (VOS, OVS)



(19) Input Candidates?

Rules? S moves; V moves, O moves?

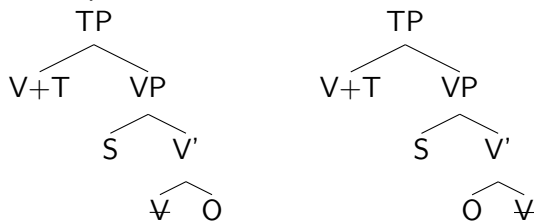
Eliminate one by one... Subject initial with head initial/final: (SVO, SOV)

- (20) Candidates? Rules? V moves to the left (V to T or C)
T is a head, heads move to head positions
S moves downward

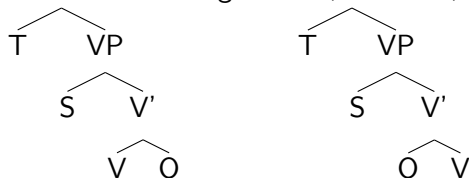
- (20) Candidates? Rules? V moves to the left (V to T or C)
T is a head, heads move to head positions
S moves downward
No good: S moves down downward movement. only
upwards movement, S is final

Two candidate structures, and the Head parameter

VSO can be derived from a structure with SVO or SOV, by movement of V to a head that has the VP as its complement. (head movement recipe, different effects and environments from T to C)

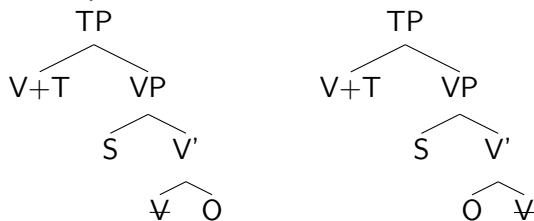


If V moves to T, and S is contained in the complement of T: what is the setting of the head parameter for T, for X? T VP: X Comp
Two candidates: we would get T VP, and V O, or T VP and O V:

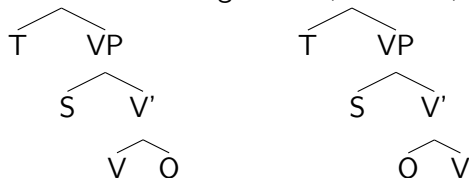


Two candidate structures, and the Head parameter

VSO can be derived from a structure with SVO or SOV, by movement of V to a head that has the VP as its complement. (head movement recipe, different effects and environments from T to C)



If V moves to T, and S is contained in the complement of T: what is the setting of the head parameter for T, for X? T VP: X Comp
Two candidates: we would get T VP, and V O, or T VP and O V:



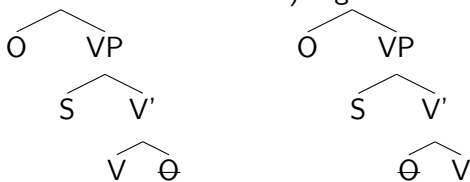
Greenberg Universal 3

Greenberg Universal 3. Languages with *dominant* VSO order are always prepositional (P NP)

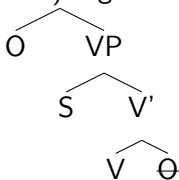
- True on Terraling? Check Universal 3.
- Search for implications in the database: VSO implies P NP?
See below.

The 6 pattern. How to get OSV

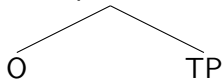
The Object must have moved or appear in some position (properties must be worked out!) higher than the subject.



the Object must have moved or appear in some position (properties must be worked out!) higher than the subject.



This movement requires our bi (Navajo bi-forms).



How to check Universal 3 in Terraling

To start with:

double implication: select all basic orders and P N N P: no

interesting results!

Property Name	Property Value	Property Name	Property Value	Count
12 NP P	No	11 P NP	Yes	91
11 P NP	No	12 NP P	Yes	31
05 SVO	Yes	09 OSV	Yes	2
11 P NP	NA	05 SVO	Yes	2
11 P NP	NA	06 SOV	Yes	2
11 P NP	NA	09 OSV	Yes	2
11 P NP	NA	10 OVS	Yes	2
11 P NP	NA	12 NP P	NA	2
12 NP P	NA	05 SVO	Yes	2
12 NP P	NA	06 SOV	Yes	2
12 NP P	NA	09 OSV	Yes	2
12 NP P	NA	10 OVS	Yes	2
12 NP P	NA	11 P NP	NA	2

Property Name	Property Value	Property Name	Property Value	Count
12 NP P	No	11 P NP	Yes	91
11 P NP	No	12 NP P	Yes	31
05 SVO	Yes	09 OSV	Yes	2

Property Name	Property Value	Property Name	Property Value	Count
12 NP P	No	11 P NP	Yes	91
11 P NP	No	12 NP P	Yes	31
05 SVO	Yes	09 OSV	Yes	2

Why not?

VSO yes, concurs with with many other types of orders, including free word order languages (like Mohawk, and some of these are NP P languages)

Cross P NP and NP P

Property Name	Property Value	Property Name	Property Value	C
11 P NP	Yes	12 NP P	No	91
11 P NP	Yes	12 NP P	Yes	44
11 P NP	No	12 NP P	Yes	31
11 P NP	NA	12 NP P	NA	2

Map:



Universal 3 VSO and P NP

Property Name	Property Value	Property Name	Property Value	Count
12 NP P	No	11 P NP	Yes	91
11 P NP	No	12 NP P	Yes	31
05 SVO	Yes	09 OSV	Yes	2

Universal 3 VSO and P NP

Property Name	Property Value	Property Name	Property Value	Count
12 NP P	No	11 P NP	Yes	91
11 P NP	No	12 NP P	Yes	31
05 SVO	Yes	09 OSV	Yes	2

Why no implications?

VSO yes, coocurs with with many other types of orders, including free word order languages (like Mohawk, and some of these are NP P languages)

Two further searches

Search:

- (24) 1. VSO yes; P NP yes; SOV no, OSV no;
should show all the languages in the database that have
this properties

(list and map)

- (25) VSO yes *and* NP P no, (SOV no, OSV no) should show
0 results.

✓

Your search query returned no results.

One could imagine a language that is halfway between Navajo (OV and N P) and English (VO and P N) VO and NP P (postpositions).
Baker p.40.

- (26) a. Chris put the book the table on (*not in Greenberg's sample*)
b. Chris the book on the table put (*rare in Greenberg's sample*)

Do these exist: perform a search

Basic Order

In generative grammar, word order variation of a given set of items (e.g. V, O, S) is derived by postulating a **basic order** from which all orders are derived by word order permutation functions (aka **movement**).

Basic Order

- Empirically, in a given language (in a given proposition), this is based on **sets of fine grained observable patterns**
- Theoretically, it is coded as basic order + movement because two properties need to be explained (coded in the theory of movement):
 - not all patterns are possible
 - different existing patterns have different syntactic properties (constituent structures),

Fine grained Word Order Patterns: Examples

Verb Object order

English: VO → Classified as VO

John **saw this movie**

...that John **saw this movie**

Dutch: both VO and OV — >reported as lacking dominant order
in the typological literature

Jan **zag deze film** = John **saw this movie**

.. dat Jan **deze film zag** = ...that John **this movie saw**

The importance of syntactic structure

In Dutch

V O is a super pattern of V X O

O V is a super pattern of O Y V

X's and Y's are completely different:

X: (subjects), weak object pronouns, modal particles, temporal adverbs, etc...

Y: negation, modal particles (NOT subjects, NOT weak pronouns, etc...

(27) Marie **gaf** hun ieder jaar toch maar een kadootje
Mary gave them each year PART PART a present

(28) dat Marie hun iedere jaar toch maar een kadootje
Mary gave them each year PART PART a present
gaf
gave

Conclusion:

How many languages in terraling have V movement? are like Dutch? X moving to Y.

(29) Search for languages which are both SOV and SVO.
Languages could also be X V be

X= could be V, N, A, P..
as long as Y is a head that selected XP.