

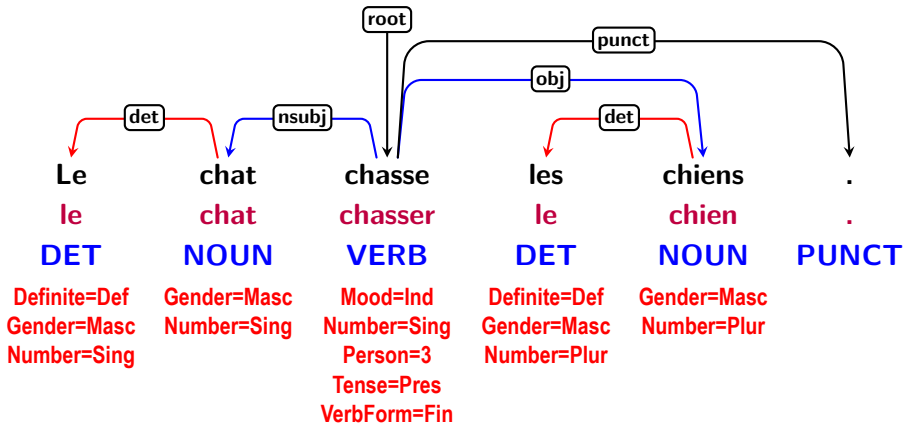
Indirect Objects across Languages: A Trap in Universal Dependencies?

Daniel Zeman

📅 September 18, 2025



Morphological and Syntactic Annotation



Current UD Languages

Information about language families (and genera for families with multiple branches) is mostly taken from [WALS Online](http://wals.info/) (IE = Indo-European).

▶		Abaza	1	<1K		Northwest Caucasian
▶		Abkhaz	1	9K		Northwest Caucasian
▶		Afrikaans	1	49K		IE, Germanic
▶		Akkadian	2	25K		Afro-Asiatic, Semitic
▶		Akuntsu	1	1K		Tupian, Tupari
▶		Albanian	2	4K		IE, Albanian
▶		Alemannic	1	1K		IE, Germanic
▶		Amharic	1	10K		Afro-Asiatic, Semitic
▶		Ancient Greek	3	456K		IE, Greek
▶		Ancient Hebrew	1	90K		Afro-Asiatic, Semitic
▶		Apurina	1	<1K		Arawakan
▶		Arabic	3	1,042K		Afro-Asiatic, Semitic
▶		Armenian	2	94K		IE, Armenian
▶		Assyrian	1	<1K		Afro-Asiatic, Semitic
▶		Azerbaijani	1	<1K		Turkic, Southwestern
▶		Bambara	1	13K		Mande
▶		Basque	1	121K		Basque
▶		Bavarian	1	15K		IE, Germanic
▶		Beja	1	11K		Afro-Asiatic, Cushitic
▶		Belarusian	1	305K		IE, Slavic
▶		Bengali	1	<1K		IE, Indic
▶		Bhojpuri	1	6K		IE, Indic
▶		Bokota	1	2K		Chibchan, Guaymíic
▶		Bororo	1	160K		Bororoan
▶		Breton	1	10K		IE, Celtic
▶		Bulgarian	1	156K		IE, Slavic
▶		Buryat	1	10K		Mongolic
▶		Cantonese	1	13K		Sino-Tibetan, Chinese
▶		Cappadocian	2	4K		IE, Greek
▶		Catalan	1	553K		IE, Romance
▶		Cebuano	1	1K		Austronesian, Central Philippine
▶		Chinese	7	309K		Sino-Tibetan, Chinese
▶		Chukchi	1	6K		Chukotko-Kamchatkan
▶		Classical Armenian	1	99K		IE, Armenian
▶		Classical Chinese	2	433K		Sino-Tibetan, Chinese
▶		Coptic	2	89K		Afro-Asiatic, Egyptian
▶		Croatian	1	199K		IE, Slavic
▶		Czech	6	4,284K		IE, Slavic
▶		Danish	1	100K		IE, Germanic
▶		Dutch	2	506K		IE, Germanic
▶		Egyptian	1	21K		Afro-Asiatic, Egyptian

- 35 language families
- 179 languages
- 319 treebanks
- 712 contributors
- 2.2 million sentences
- 37 million words

(as of release 2.16)

UD as a Resource for Typological Studies

- Alzetta, C., Dell'Orletta, F., Montemagni, S. & Venturi, G. (2018). Universal Dependencies and quantitative typological trends. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 4540–4549, Miyazaki: ELRA.
- de Marneffe, M.-C., Nivre, J. & Zeman, D. (2024). Function words in Universal Dependencies. *Linguistic Analysis* 43(3–4), 549–588.
- Futrell, R., Mahowald, K. & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. In *Proceedings of National Academy of Sciences*, 112(33), 10336–10341.
- Gerdes, K., Kahane, S. & Chen, X. (2021). Typometrics: From Implicational to Quantitative Universals in Word Order Typology. *Glossa: a journal of general linguistics* 6(1): 17.
- Levshina, N. (2019). Token-based typology and word order entropy: A study based on Universal Dependencies. *Linguistic Typology*, 23(3), 533–572.

INTERCORP

projekt paralelních korpusů
Filozofické fakulty Univerzity Karlovy v Praze

kon text

Query Corpora Save Concordance Filter Frequency Collocations View Help

Corpus: InterCorp v16ud - Czech | Query: dům, ház, ev (29,453 hits) ~ Details

Hits: 29,453 | i.p.m.: 56.99 (related to the whole corpus) | ARF: 11,768.94 | Result is sorted

Line selection: simple

InterCorp v16ud - Czech

InterCorp v16ud - Hungarian

InterCorp v16ud - Turkish



Co to je za dům/NOUN/Acc, kde nikdo nežije ?



tenhle dům/NOUN/Nom nemá ducha verandy ?



Tenhle dům/NOUN/Nom fakt nemá žádný duchy ?



Mám zapálit ten dům/NOUN/Acc i s ní ?



Mimochodem, co je v tom podivném pustém domě/NOUN/Loc ?



Proč jste se vyčural na našem domě/NOUN/Loc ?



To je Yuneho dům/NOUN/Nom .



Tohle není náš dům/NOUN/Nom .



Ale omylem jsem je přinesla do domu/NOUN/Gen pana Cha .



Strýčku, Eun Chae sedí před naším domem/NOUN/Ins .



Na schodech u domu/NOUN/Gen Moo Hwuka

Egy ilyen házban/NOUN/Ins miért nincs senki ?

Ez furcsa... a háznak/NOUN/Gen nem lenne veranda szelleme ?

Ebben a házban/NOUN/Ins egyáltalán nincsenek szellemek ?

Gonosz boszorkány ... rá kellene gyűjtanom az egészet házát/NOUN/Acc !

Ámde mi legyen azzal a gyanús elhagyatott házzal/NOUN/Ins ?

Miért pisíltél le egy idegen házát/NOUN/Acc ?

Ez Yoon-ék háza/NOUN/Nom , a mi házunk hátrébb van .

Nem ez a mi házunk/NOUN/Nom .

De véletlenül vittem azokat , Mr . Cha házába/NOUN/Ins .

Eun Chae , itt van a házunk/NOUN/Nom előtt , bácsikám .

Moo Hwuk abiussi házánál/NOUN/Acc

Böyle bir evde/NOUN/Loc nasıl kimse olmaz ?

Bu evde/NOUN/Loc veranda hayaleti bile yok mu ?

Bu evde/NOUN/Loc yaşayan hiç hayalet yok mu ?

Evini/NOUN/Acc başına yıksam mı ?

Bu arada o issiz ev/NOUN/Nom de neyin nesi ?

Sen niye evimizin/NOUN/Gen duvarına işedin ki ?

Bu Yune'un evi/NOUN/Nom .

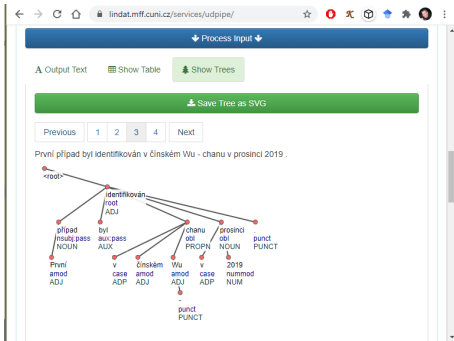
Bayım , bu bizim evimiz/NOUN/Nom değil .

Yanlışlıkla sizin evde/NOUN/Loc unutmuşum .

Eun - chae evin/NOUN/Gen önünde öylece oturup

- Moo - hwuk'un evindeyim/NOUN/Loc

<https://lindat.mff.cuni.cz/services/udpipe/>



UD as a Resource for Typological Studies

Quite useful...

... but **one has to be careful**. Sometimes people differ in their interpretation of the guidelines.

Indirect objects are a prime example.

Dependents of Clauses (Verbal or Not)

	Nominal	Clausal	Modifier	Function
Core	nsubj	csubj		
Non-Core	obl vocative dislocated expl	advcl	advmod discourse	aux cop mark

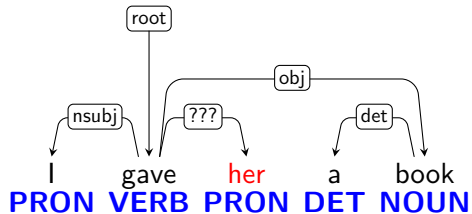
Dependents of Verbs, Adjectives and Adverbs

	Nominal	Clausal	Modifier
Core	obj	ccomp	
	iobj	xcomp	
Non-Core	obl expl	advcl	advmod

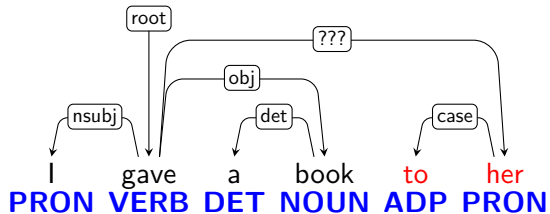
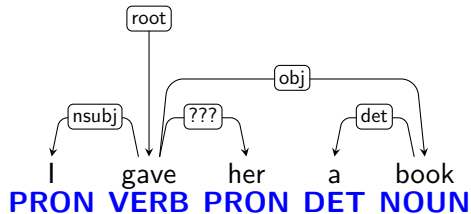
Dependents of Nominals

Nominal	Clausal	Modifier	Function
nmod	acl	amod	det
appos		nummod	case

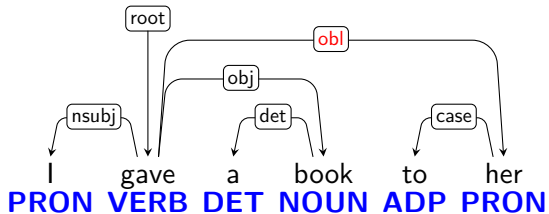
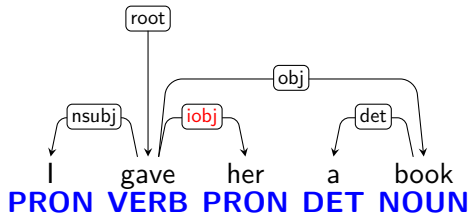
Information Packaging and 'Indirect Object'



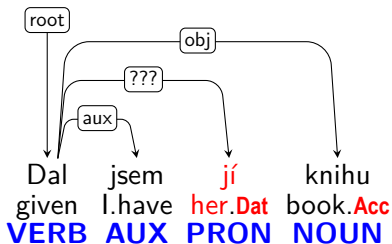
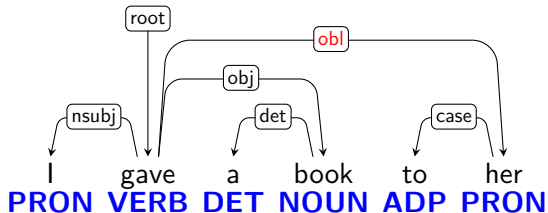
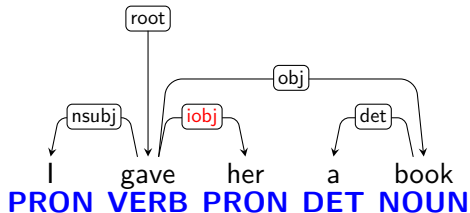
Information Packaging and 'Indirect Object'



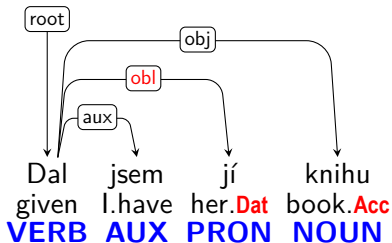
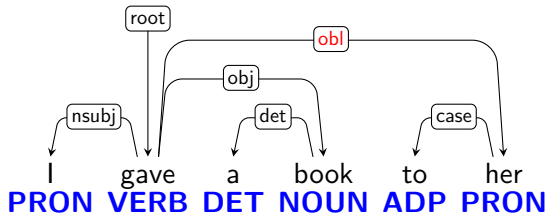
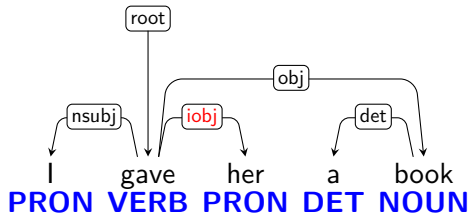
Information Packaging and 'Indirect Object'



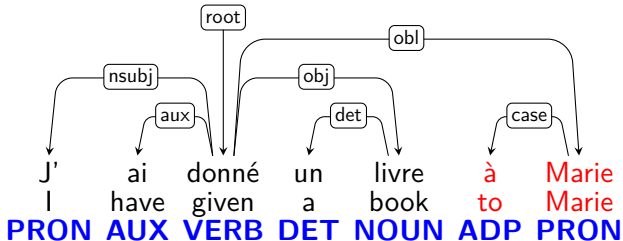
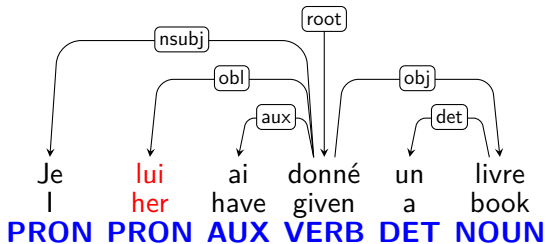
Information Packaging and 'Indirect Object'



Information Packaging and 'Indirect Object'



Information Packaging and 'Indirect Object'



Indirect Object

Merriam-Webster

<https://www.merriam-webster.com/dictionary/indirect%20object>

*a noun, pronoun, or noun phrase that occurs in addition to a direct object after some verbs and indicates the person or thing that **receives** what is being **given** or done : the person or thing that the action of a verb is performed for or **directed to***

Indirect Object

Lawless French

<https://www.lawlessfrench.com/grammar/indirect-objects-2/>

*In French, an **indirect object** is a person found after a **preposition** (usually *à* or *pour*).*

Ditransitive Constructions

Ditransitive Constructions

Martin Haspelmath. 2015. Ditransitive Constructions. *Annual Review of Linguistics* 2015.1:19–41.

- Three arguments: **agent** (A), **theme** (T), **recipient** (R)
- Possessive transfer (*give, lend*) or cognitive transfer (*tell, show*).

Ditransitive Constructions

Martin Haspelmath. 2015. Ditransitive Constructions. *Annual Review of Linguistics* 2015.1:19–41.

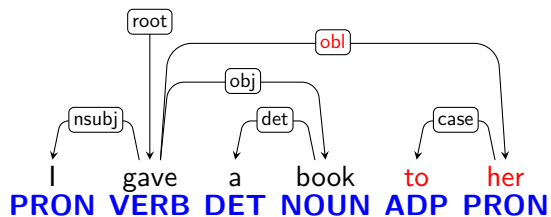
- Three arguments: **agent** (A), **theme** (T), **recipient** (R)
 - Possessive transfer (*give, lend*) or cognitive transfer (*tell, show*).
-
- “*Ditransitive constructions have often been said to involve an “indirect object” in the past, but this notion has not proved helpful in recent research. In this overview, the focus is on broadly cross-linguistic research on ditransitive constructions.*”

Ditransitive Constructions

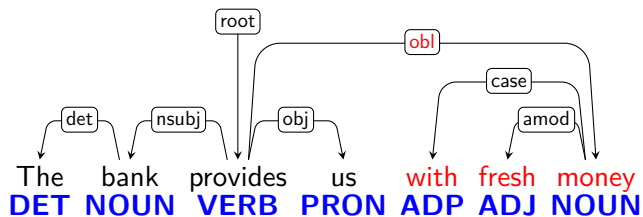
Martin Haspelmath. 2015. Ditransitive Constructions. *Annual Review of Linguistics* 2015.1:19–41.

- Three arguments: **agent** (A), **theme** (T), **recipient** (R)
- Possessive transfer (*give, lend*) or cognitive transfer (*tell, show*).
- Indirective alignment (R distinct from monotransitive P)
- Secundative alignment (T distinct from monotransitive P)
- Neutral alignment (double object construction)
- *“Ditransitive constructions have often been said to involve an “indirect object” in the past, but this notion has not proved helpful in recent research. In this overview, the focus is on broadly cross-linguistic research on ditransitive constructions.”*

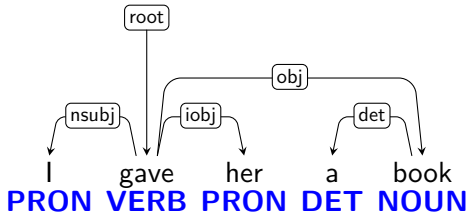
Indirective Alignment



Secundative Alignment



Neutral Alignment



- The two objects are **almost** equally 'core'
 - *She was given a book* — **A book was given her ...* $R > T$
 - *What did I give her* — **Who did I give a book ...* $T > R$

UD Guidelines Requirements

- iobj in UD must be a **core** argument (not oblique)
 - $\Rightarrow$ typically in neutral alignment
 - $\Rightarrow$ many “ditransitive clauses” look like monotransitive in UD

UD Guidelines Requirements

- iobj in UD must be a **core** argument (not oblique)
 - $\Rightarrow$ typically in neutral alignment
 - $\Rightarrow$ many “ditransitive clauses” look like monotransitive in UD
- At most one obj per clause
- No limit on the number of iobj
 - More core objects $\Rightarrow$ must select one of them as obj, all others end up as iobj
 - Also allowed: iobj without obj present

UD Guidelines Requirements

- iobj in UD must be a **core** argument (not oblique)
 - $\Rightarrow$ typically in neutral alignment
 - $\Rightarrow$ many “ditransitive clauses” look like monotransitive in UD
- At most one obj per clause
- No limit on the number of iobj
 - More core objects $\Rightarrow$ must select one of them as obj, all others end up as iobj
 - Also allowed: iobj without obj present
- But some treebanks ignore these guidelines and go with “traditional” definition of indirect object 😞

Core vs. Oblique

Dependents of Clauses (Verbal or Not)

	Nominal	Clausal	Modifier	Function
Core	nsubj	csubj		
Non-Core	obl vocative dislocated expl	advcl	advmod discourse	aux cop mark

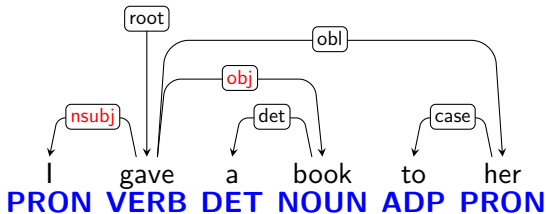
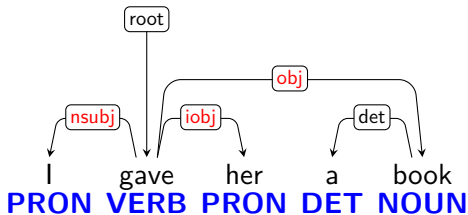
Dependents of Verbs, Adjectives and Adverbs

	Nominal	Clausal	Modifier
Core	obj iobj	ccomp xcomp	
Non-Core	obl expl	advcl	advmod

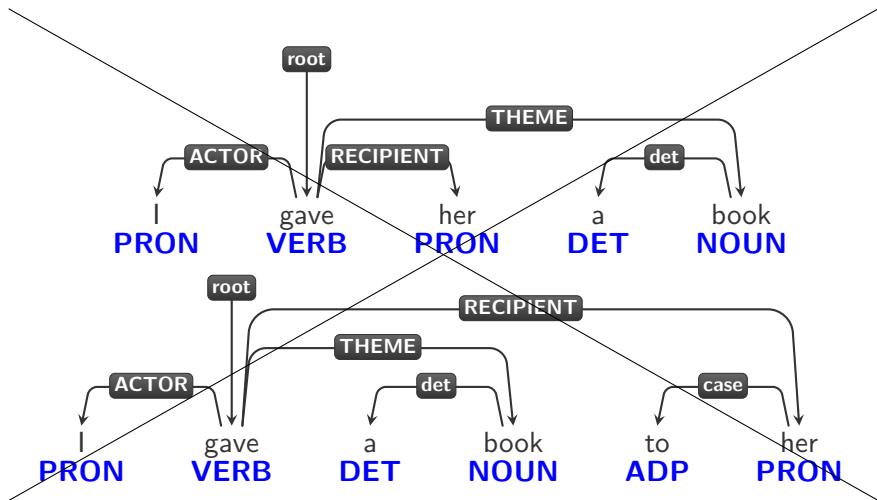
Dependents of Nominals

Nominal	Clausal	Modifier	Function
nmod appos	acl	amod nummod	det case

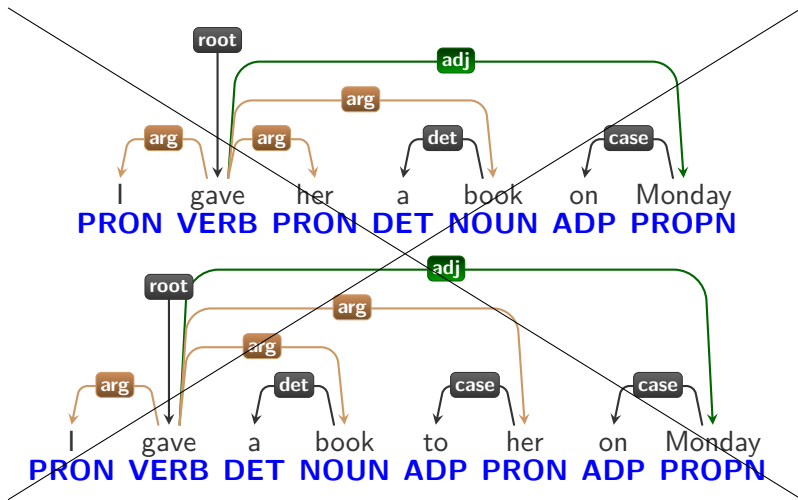
Information Packaging: Core Arguments vs. Oblique Dependents



UD is NOT about Semantic Roles!



UD Avoids Argument-Adjunct Distinction!



So What Is Core and Why?



Language-specific Coding Strategy

- Idea:
 - **Oblique** arguments are marked **similarly to adjuncts** (prepositions, certain morphological cases...)
 - Core arguments are marked differently
 - $\Rightarrow$ easy for annotators and non-linguists!
- Why are core arguments special?
 - They tend to be **targeted by grammatical rules**
 - Passivization
 - Control verbs
 - Reflexives
 - ...

Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?
- Andrews, 2007 (In Shopen: Language Typology)
 - Identify **primary transitive predicates**
 - We need semantic roles for this! (One-time only.)

Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?
- Andrews, 2007 (In Shopen: Language Typology)
 - Identify **primary transitive predicates**
 - We need semantic roles for this! (One-time only.)
 - Actor/agent = function **A**
 - Undergoer/patient = function **P**

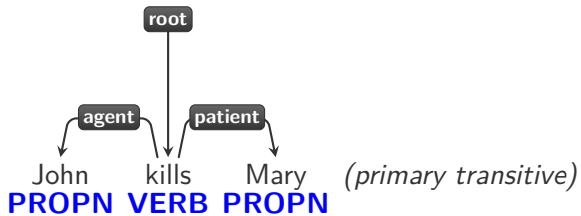
Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?
- Andrews, 2007 (In Shopen: Language Typology)
 - Identify **primary transitive predicates**
 - We need semantic roles for this! (One-time only.)
 - Actor/agent = function **A**
 - Undergoer/patient = function **P**
 - Note the way they are coded
 - Note other grammatical rules that target them
 - Generalize to other predicates with same coding and rules

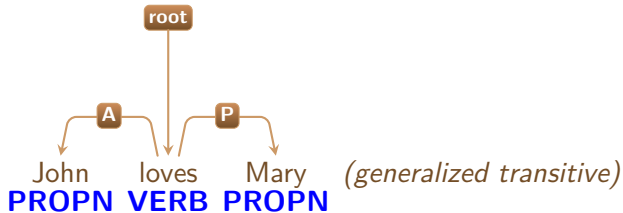
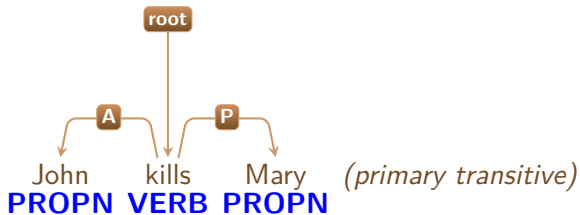
Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?
- Andrews, 2007 (In Shopen: Language Typology)
 - Identify **primary transitive predicates**
 - We need semantic roles for this! (One-time only.)
 - Actor/agent = function **A**
 - Undergoer/patient = function **P**
 - Note the way they are coded
 - Note other grammatical rules that target them
 - Generalize to other predicates with same coding and rules
- Then define:
 - function A $\Rightarrow$ **nsubj**
 - function P $\Rightarrow$ **obj**

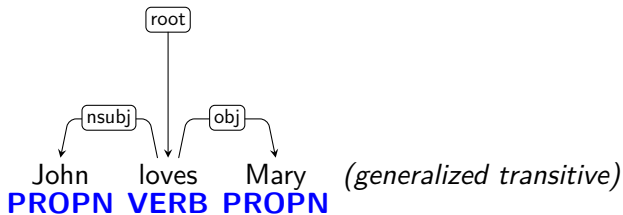
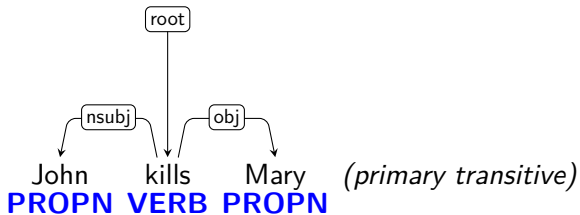
Transitive Predicates in English



Transitive Predicates in English



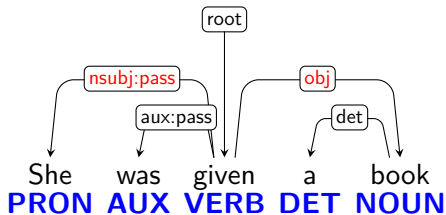
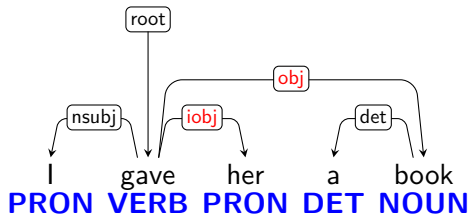
Transitive Predicates in English





Ditransitive Predicates

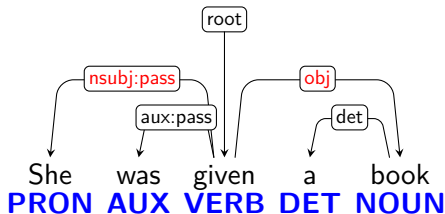
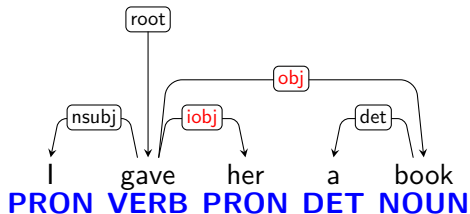
- Three core arguments: `nsubj`, `obj`, and **`iobj`**
- The third argument is 'similar enough' to the second argument (**P**)
- We do **not** refer to semantic roles **R** and **T**





Ditransitive Predicates

- Three core arguments: `nsubj`, `obj`, and **`iobj`**
- The third argument is 'similar enough' to the second argument (**P**)
- We do **not** refer to semantic roles **R** and **T**
- Which one is `iobj`? Language-specific rules. English: the one closer to the verb (i.e., in English it is the **RECIPIENT** semantic role).



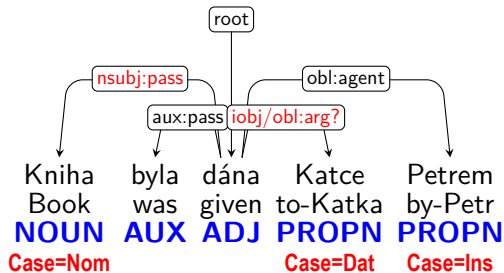
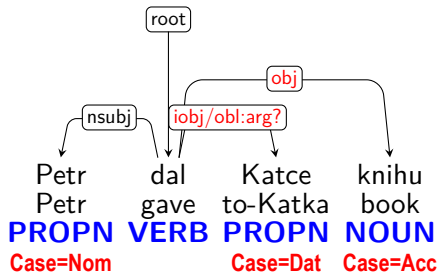


-
- The image displays two syntax trees for the sentence "I gave her a book" and "She was given a book".
- Left Tree (Active Voice):**
- root** branches into **nsubj** and **obj**.
 - nsubj** branches into the word **I**.
 - obj** branches into **iobj** and **det**.
 - iobj** branches into the word **her**.
 - det** branches into the word **a**.
 - The word **gave** is the main verb, connected to the root.
 - Below the words, their parts of speech are listed: **PRON** (I), **VERB** (gave), **PRON** (her), **DET** (a), **NOUN** (book).
- Right Tree (Passive Voice):**
- root** branches into **nsubj:pass** and **obj**.
 - nsubj:pass** branches into the word **She**.
 - obj** branches into **aux:pass** and **det**.
 - aux:pass** branches into the word **was**.
 - det** branches into the word **a**.
 - The word **given** is the main verb, connected to the root.
 - Below the words, their parts of speech are listed: **PRON** (She), **AUX** (was), **VERB** (given), **DET** (a), **NOUN** (book).

- 26

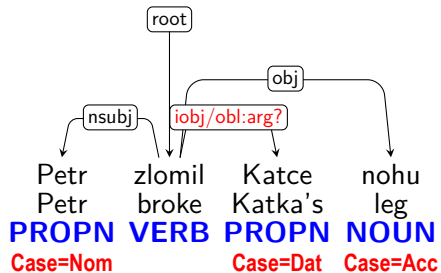
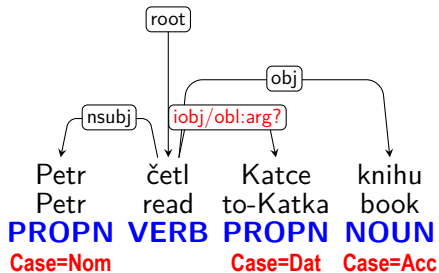


Czech Ditransitive Clauses



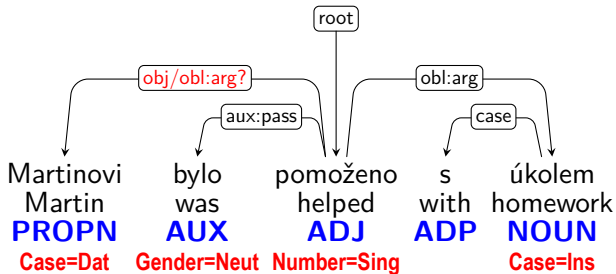
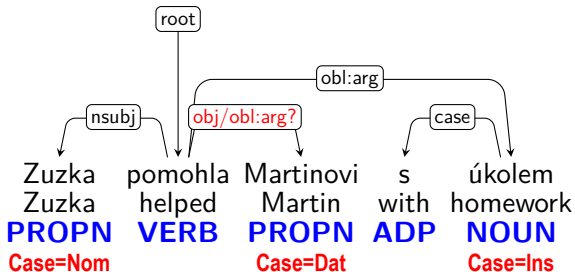


Dative: Recipient vs. Beneficiary



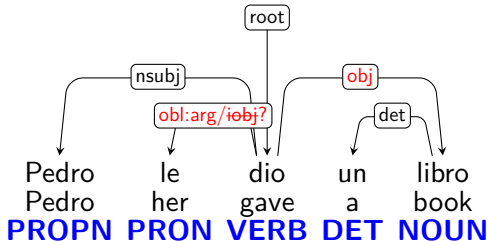
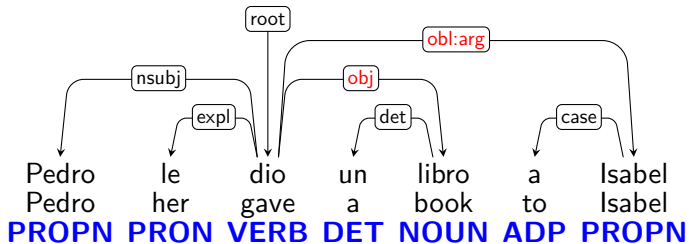


Monotransitive with Dative?



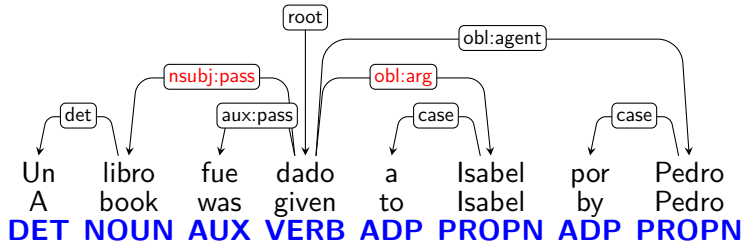
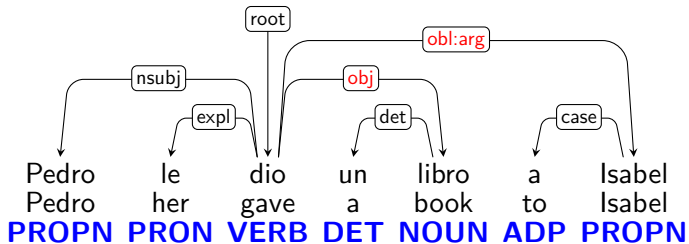


Spanish Ditransitive Clauses



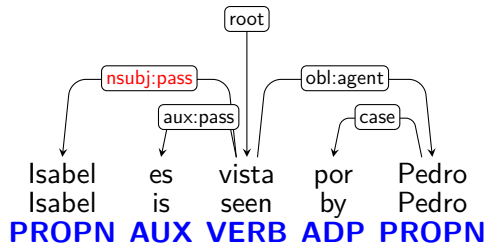
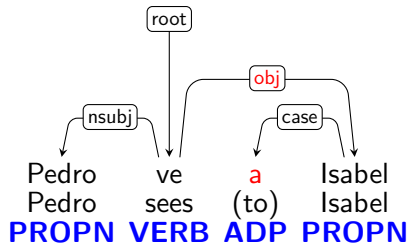


Spanish Ditransitive Clauses



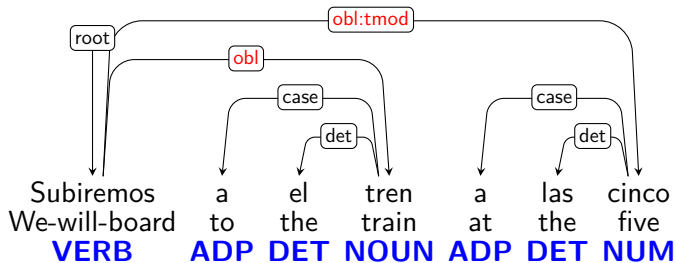


Problem: Animate Direct Objects with Preposition



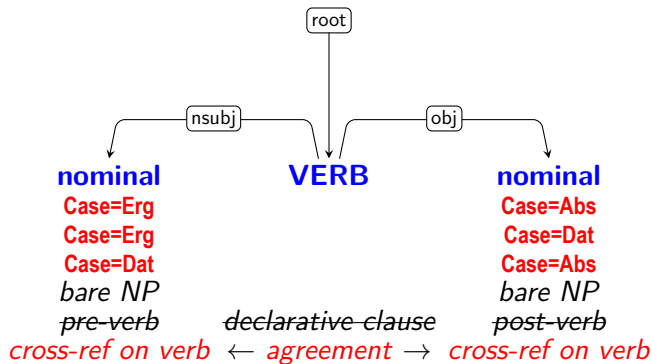


Spanish Oblique/Adjuncts



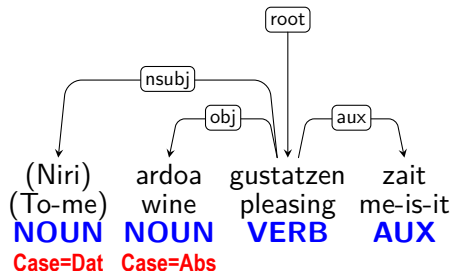
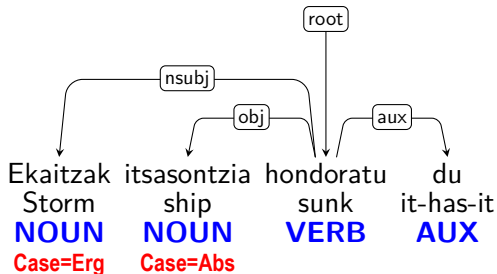


Basque Transitive Clauses



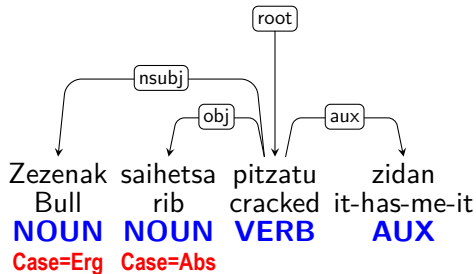
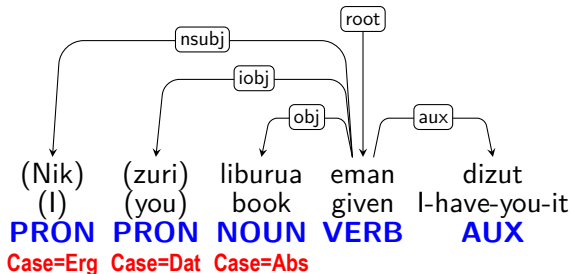


Basque Transitive Clauses



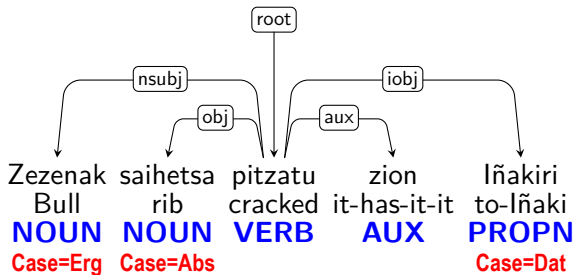
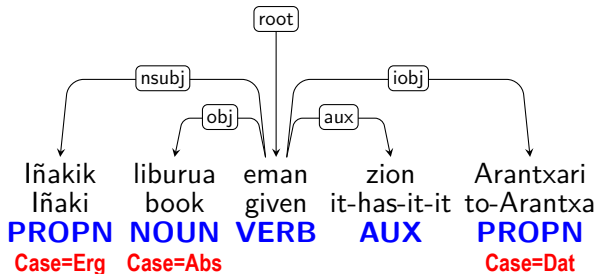


Basque Ditransitive Clauses



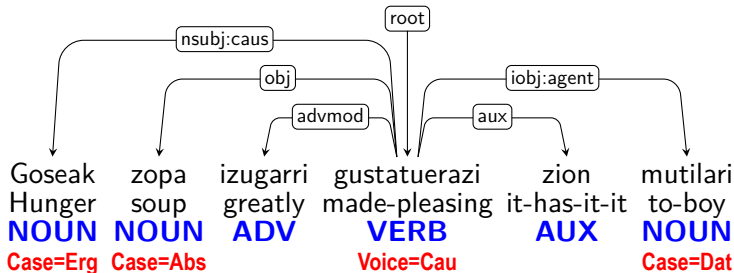
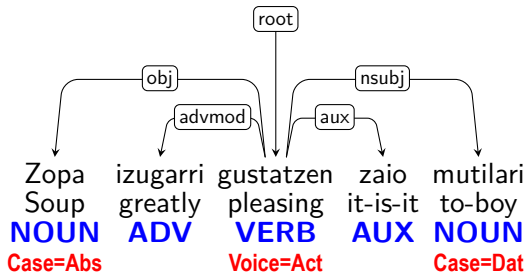


Basque Ditransitive Clauses



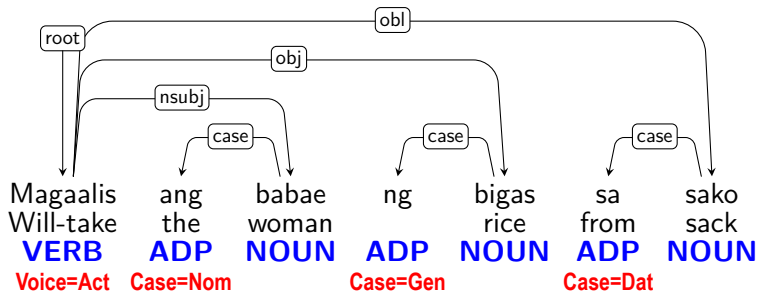


Basque Causative Applied to Dative Subject



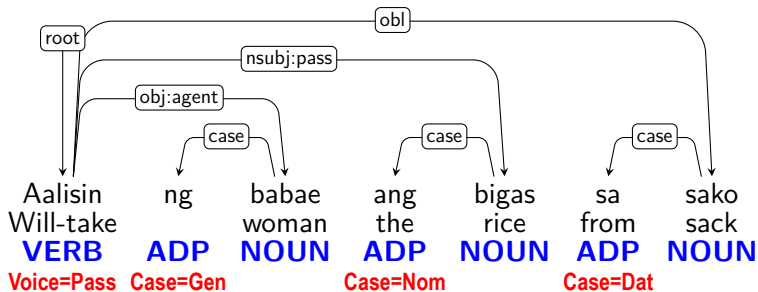


Tagalog Actor Voice



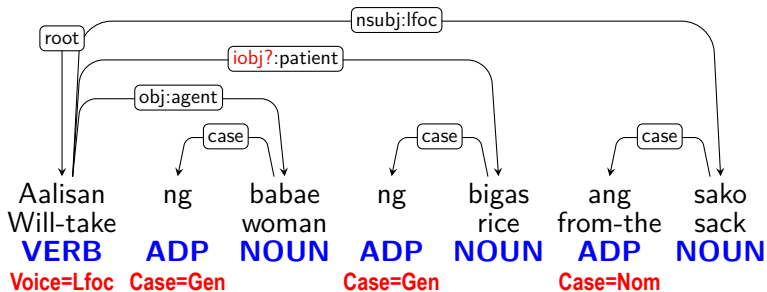


Tagalog Patient/Undergoer Voice



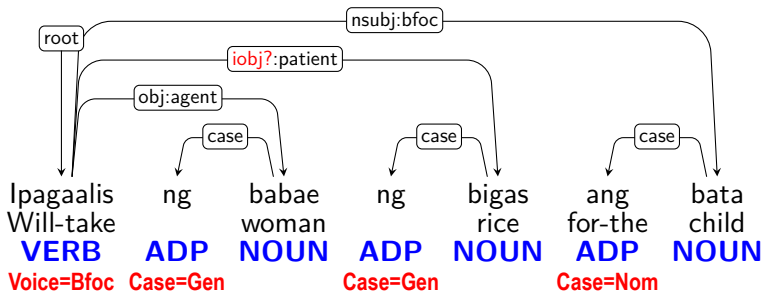


Tagalog Locative Voice $\Rightarrow$ Ditransitive!

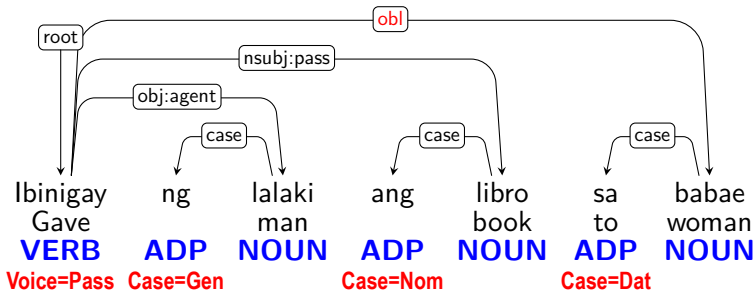




Tagalog Benefactive Voice $\Rightarrow$ Ditransitive!



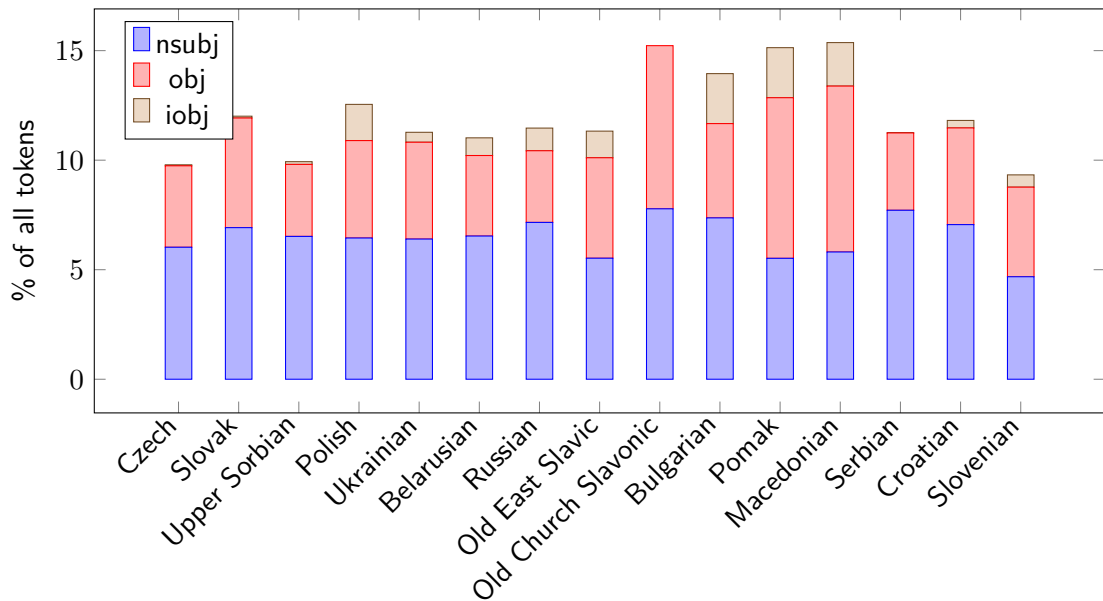
Tagalog: Giving Not Ditransitive



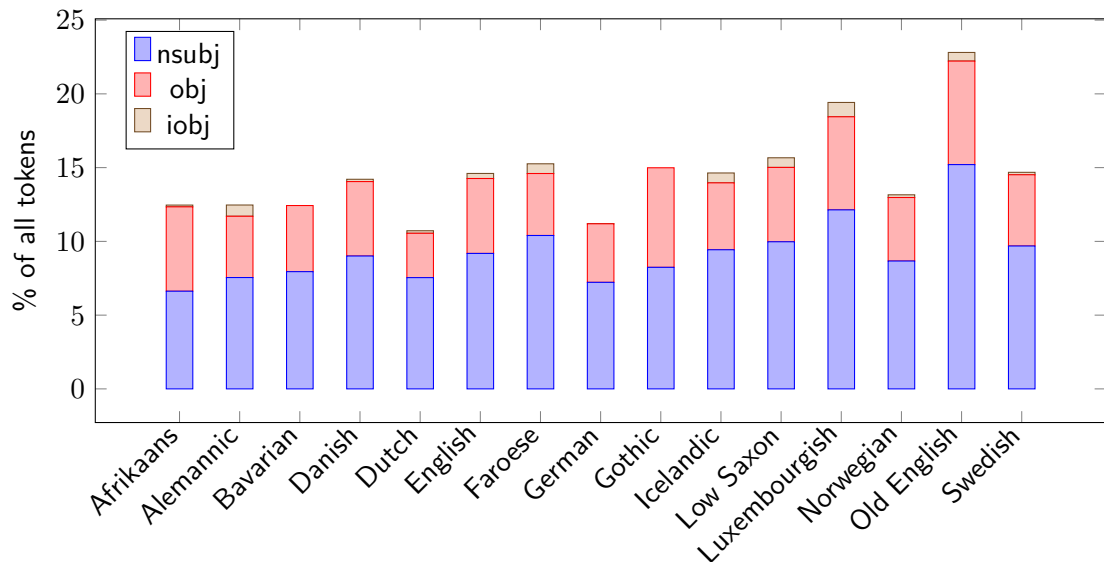
A Peek to the Data

- iobj is rare
 - Usually less than 1% tokens, except Phrygian (7%) and Umbrian (3%) (if they are annotated correctly)
 - Always the least frequent nominal core argument
- But present in most language families
 - Prominent exception: **Uralic** languages. Missing from all languages except Hungarian
 - Seven other languages are only representatives of their families: Japanese, Chukchi, Yupik, Kiche, Bororo, Bambara, Esperanto

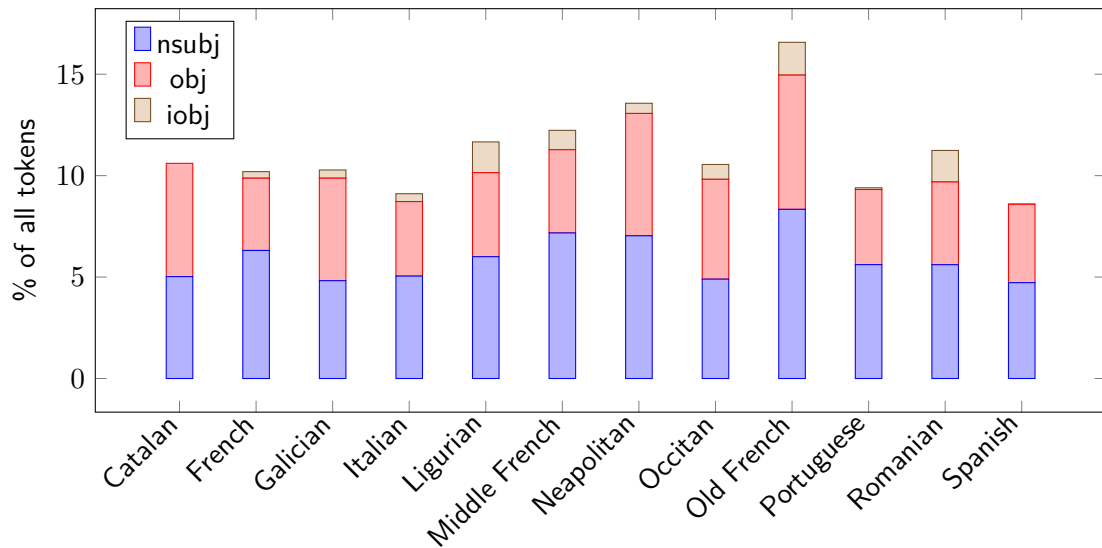
Slavic Languages



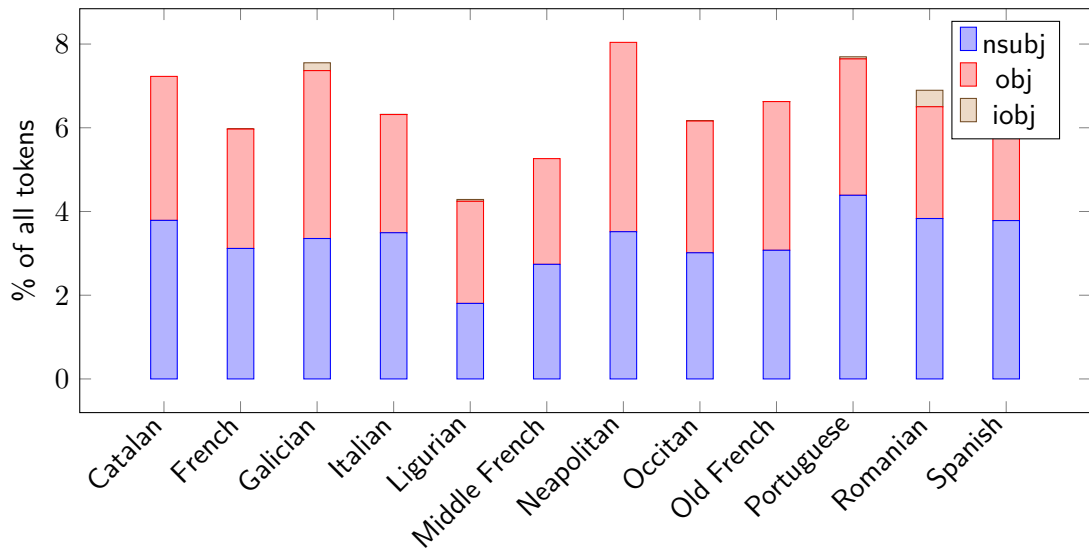
Germanic Languages



Romance Languages



Romance Languages – Only Non-Pronominal Arguments



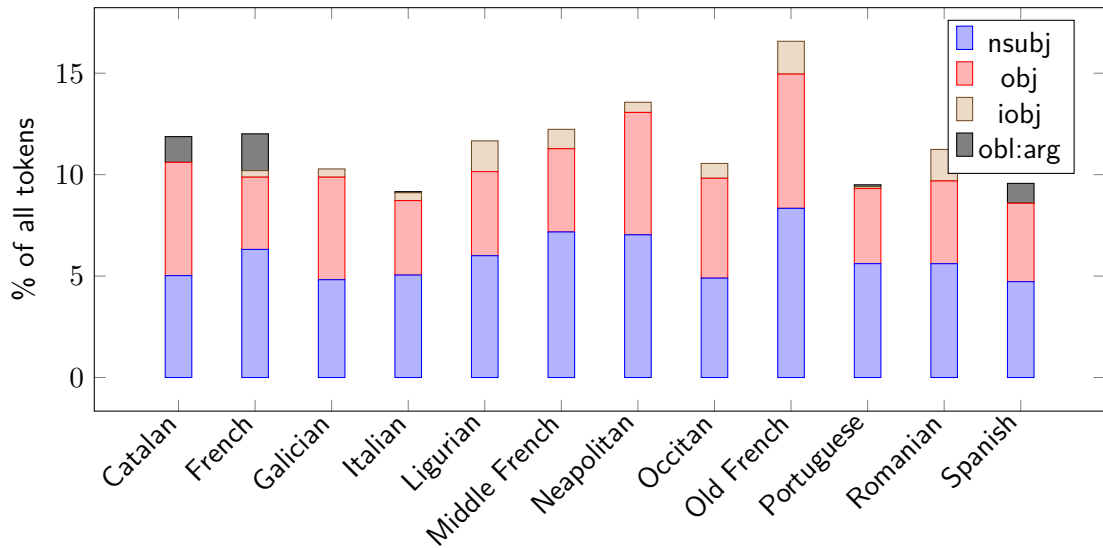
iobj in Romance Languages by Parent Lemma

-  French: *dire* 'say' 317, *demander* 'ask' 116, *faire* 'do' 111, *donner* 'give' 102, *permettre* 'permit' 72
-  Italian: *fare* 'do' 358, *dire* 'say' 309, *dare* 'give' 183, *piacere* 'please' 153, *parere* 'seem' 106
-  Romanian: *da* 'give' 1579, *zice* 'say' 1264, *face* 'do' 742, *spune* 'say' 375, *fi* 'be' 304
-  Portuguese: *dar* 'give' 188, *dizer* 'say' 75, *fazer* 'do' 31, *pedir* 'ask' 30, *oferecer* 'offer' 25
-  Galician: *ter* 'have' 53, *dar* 'give' 48, *facer* 'do' 28, *permitir* 'permit' 23, *corresponder* 'correspond' 22
-  Spanish: *give* (English) 1

obl:arg in Romance Languages by Parent Lemma

-  French: *naître* 'be born' 591, *situer* 'locate' 439, *être* 'be' 424, *faire* 'do' 260, *à* 'to' 253 (noise; occurs under *à partir de*)
-  Italian: *dire* 'say' 39, *fare* 'do' 38, *dare* 'give' 20, *rispondere* 'respond' 15, *avere* 'have' 13
-  Portuguese: *associar* 'associate' 152, *corresponder* 'correspond' 94, *depende* 'depend' 70, *resultar* 'result' 70, *relacionar* 'relate' 56
-  Spanish: *dar* 'give' 572, *hacer* 'do' 308, *decir* 'say' 244, *pedir* 'ask' 205, *permitir* 'permit' 171
-  Catalan: *fer* 'do' 194, *anar* 'go' 162, *comptar* 'count' 161, *donar* 'give' 155, *participar* 'participate' 149

Romance Languages with obl:arg



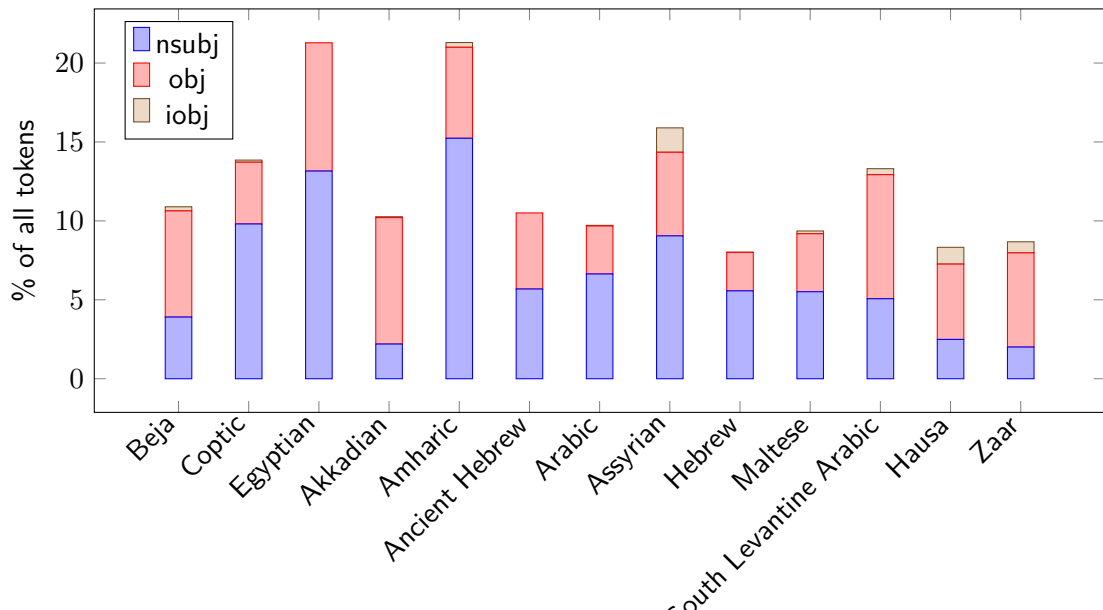
iobj in Slavic Languages by Parent Lemma

-  Czech: *mít* 'have' (noise), *požádat* 'ask', *dát* 'give', *řici* 'say', *vyzvat* 'invite'
-  Slovak: *poprosiť* 'ask', *spýtať* 'ask', *presvedčiť* 'convince', *vyzvať* 'invite', *požiadať* 'ask'
-  Upper Sorbian: *pasć*, *přewostajić*, *dowěrjeć*, *přistać*, *spřistupnić*
-  Polish: *mieć* 'have', *dotyczyć* 'concern', *powiedzieć* 'say', *dać* 'give', *mówić* 'talk'
-  Ukrainian: *дати* 'give', *надаватися* 'be given', *передати* 'pass', *треба* 'needed', *сказати* 'say'
-  Russian: *дать* 'give', *казаться* 'seem', *нужный* 'needed', *сказать* 'say', *давать* 'give'
-  Slovenian: *povedati* 'say', *dati* 'give', *reči* 'say', *pomagati* 'help', *biti* 'be'
-  Croatian: *dati* 'give', *davati* 'give', *reći* 'say', *moći* 'be able', *pružiti* 'offer'
-  Bulgarian: *дам-(се)* 'give', *говоря* 'say', *участвам* 'participate', *отида-(си)* 'go', *вляза* 'enter'

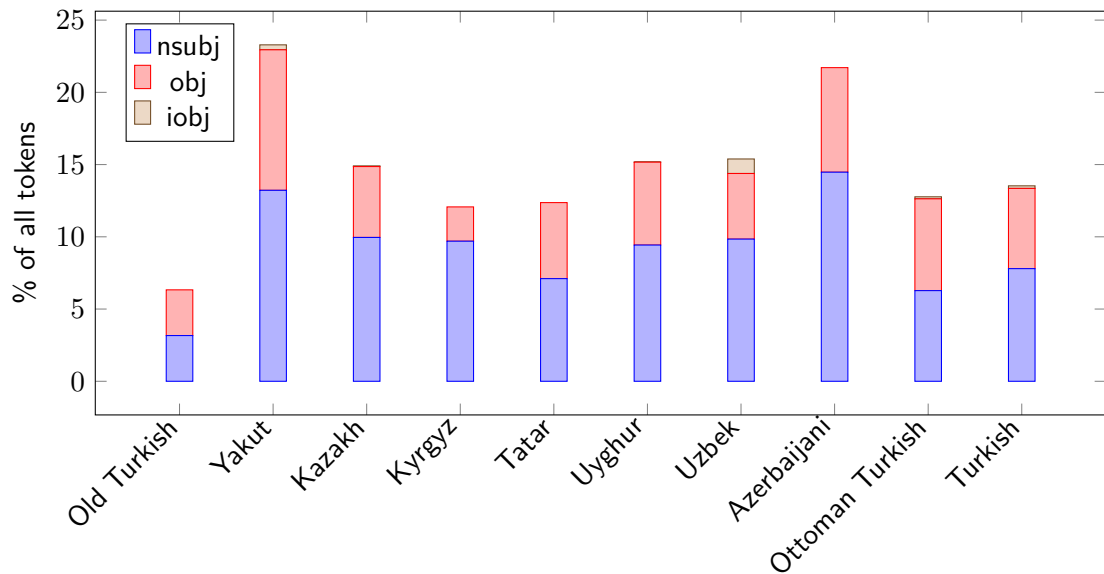
iobj in Germanic Languages by Parent Lemma

-  English: *show, give, tell, ask, get*
-  Afrikaans: *gee* 'give', *verleen* 'granted', *skenk* 'donate', *bied* 'offer', *stuur* 'send'
-  Dutch: *geven* 'give', *vragen* 'ask', *vertellen* 'tell', *lijken* 'seem', *zeggen* 'say'
-  Low Saxon: *geaven* 'give', *seggen* 'say', *vortellen* 'tell', *maken* 'make', *doon* 'do'
-  German: *kosten* 'cost'
-  Danish: *give* 'give', *vise* 'show', *tænke* 'think', *fortælle* 'tell', *gøre* 'do'
-  Swedish: *ge* 'give', *lära* 'learn', *ta* 'take', *göra* 'do', *skaffa* 'get'
-  Norwegian: *la* 'let', *gi* 'give', *gje* 'give', *sikre* 'secure', *be* 'pray'
-  Icelandic: *segja* 'say', *gefa* 'give', *veita* 'provide', *gera* 'do', *bjóða* 'offer'

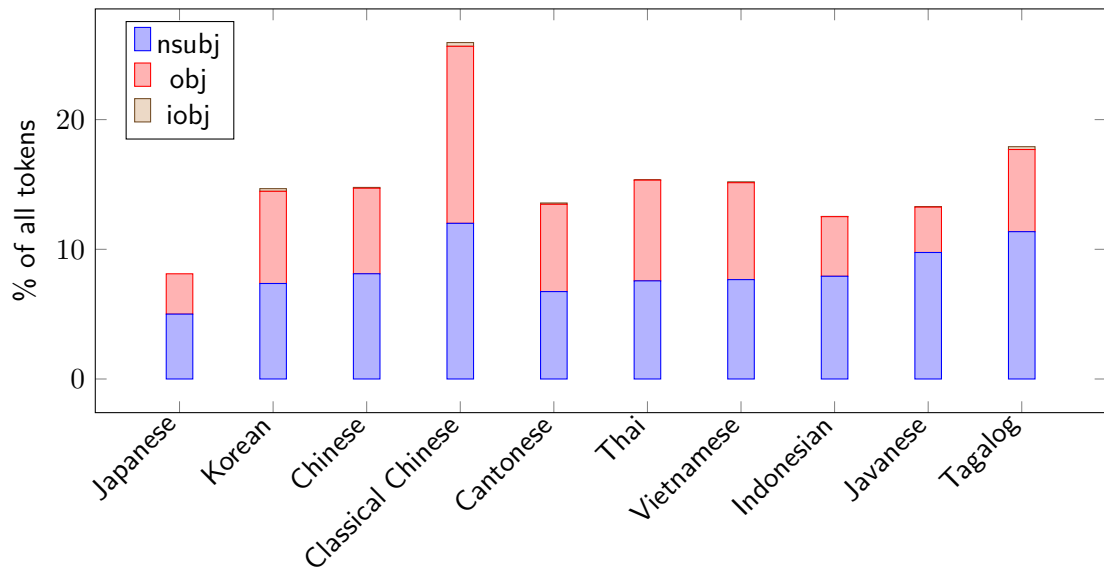
Afro-Asiatic Languages



Turkic Languages



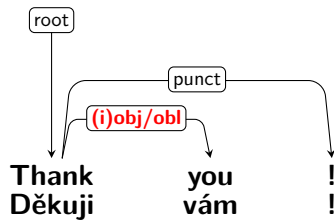
Selected East Asian Languages



Summary

- iobj may not be what you think it is
- A change in UD would have to be v3 of the guidelines (not soon)
- Be careful with contrastive studies involving iobj

<https://universaldependencies.org/>



<https://universaldependencies.org/>