EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Syntax & Semantics WiSe 2022/2023

Lecture 17: Summary Syntax

10/01/2023, Christian Bentz

Overview

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Basic Concepts

Basic Concepts

- **Constituency** (Lecture 2)
- **Parts of Speech** (Lecture 2)
- **Headedness** (Lecture 3)
- **Valency** (Lecture 3)
- **Grammatical Functions** (Lecture 3)

Definition: Constituents

Both the **basic elements/units** of a sentence – often orthographic words – as well as **combinations of those**, i.e. **phrases**, count as constituents.

Most basic constituents:
[Kim] [sees] [a] [big] [tree]

Higher level constituents:
[big[tree]], [a[big[tree]]], etc.

Müller (2019). Grammatical theory, p. 7.

What is a word anyways?

> *The general distinction between morphology and syntax is widely taken for granted, but it crucially depends on a cross-linguistically valid concept of '(morphosyntactic) word'. I show that there are no good criteria for defining such a concept. I examine ten criteria in some detail [...] and I show that none of them is necessary and sufficient on its own, and no combination of them gives a definition of 'word' that accords with linguists' orthographic practice.*

Haspelmath (2011). The indeterminacy of word segmentation and the nature of morphology and syntax, p. 31.

Word Criterion: Free Occurrence

"Bloomfield (1933: 160) called utterance segments that can occur on their own **free forms**, and he famously defined the word as "a free form which does not consist entirely of (two or more) lesser free forms; in brief, a word is a minimum free form"."

Haspelmath (2011), p. 39 citing Bloomfield (1933).

Example

(1) Where are you? - *Here*.
What do you need? - *Money*.

Word Criterion: Free Occurrence

"But this definition does not single out forms that correspond to our intuition of grammatical words. On the one hand, it is too strict, because by this definition compounds [...] would not be words, but phrases, because they have constituents that are themselves free forms. On the other hand, it is much too loose, because many phrases [...] would count as words [...]"

Haspelmath (2011), p. 39-40.

Example

(2) *firewater* (two separate free forms): fire water

(3) *the flower* (one single free form): *the

Table 1. Nine studies that examine wordhood using test batteries

	Zwicky & Pullum 1983	Kanerva 1987	Bresnan & Mchombo 1995	Ackema & LeSourd 1997	Monachesi 1999	Harris 2000	Milićević 2005	Lieber & Scalise 2006	Bickel et al. 2007
Free occurrence				+			+		
External mobility and internal fixedness	+			+	+	+			
Uninterruptibility				+					+
Non-selectivity	+	+			+	+	+		+
Non-coordinatability			+	+	+		+	+	+
Anaphoric islandhood			+					+	
Nonextractability			+					+	
Morphophonological idiosyncrasies	+	+			+	+	+		
Deviations from biuniqueness									+

Haspelmath (2011), p. 60.

Tests for Constituency

- **Substitution Test**
  he knows [the man] → he knows [a woman] ✓
- **Pronominalization Test**
  he knows [the man] → he knows [him] ✓
- **Question Formation Test**
  Whom does he know? – [The man]. ✓
- **Permutation Test**
  he knows [the man] → [the man] he knows ✓
  he knows [the man] → he [the man] knows x
- **Fronting Test**
  he knows [the man] → [the man] he knows ✓
- **Coordination Test**
  he knows [the man] → he knows [the man] and [the woman] ✓

Problems with Constituency Tests

> “It would be ideal if the tests presented here delivered clear-cut results in every case, as the empirical basis on which syntactic theories are built would thereby become much clearer. Unfortunately, this is not the case. There are in fact a number of problems with constituent tests, [...]”

Müller (2019). Grammatical theory, p. 11.

Universality of Constituency (?)

Thalanyji (?, Pama-Nyungan(?))

(4) Kupuju-lu **kaparla-nha** yanga-lkin **wartirra-ku-nha**
child-ERG dog-ACC chase-PRES woman-DAT-ACC
"The child chases the woman's dog."

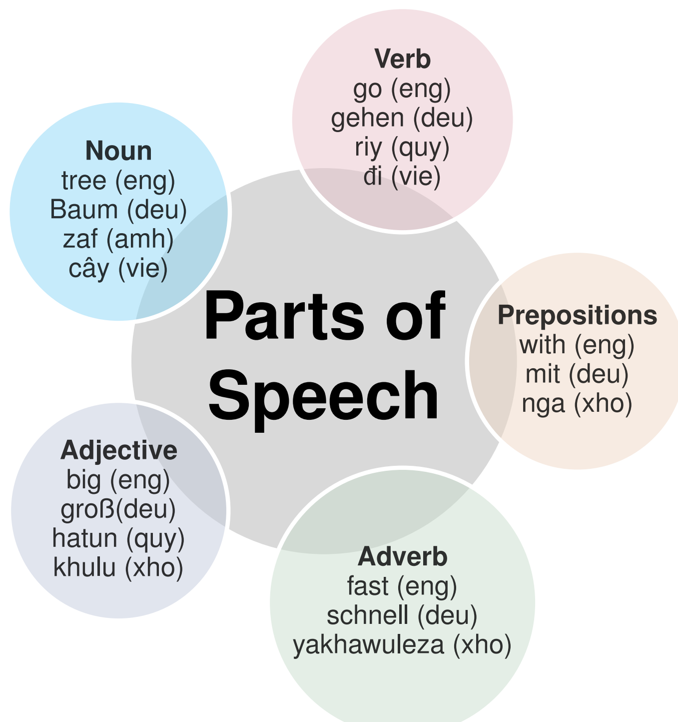
"Note how possessive modifiers – coded by a special use of the dative case – additionally pick up the case of the noun they modify, as with the accusative -nha on "dog" and "woman-Dat" [...] It is this **case-tagging**, rather than **grouping of words into constituents**, which forms the basic organizational principle in many Australian languages."

Evans & Levinson (2009), p. 441.

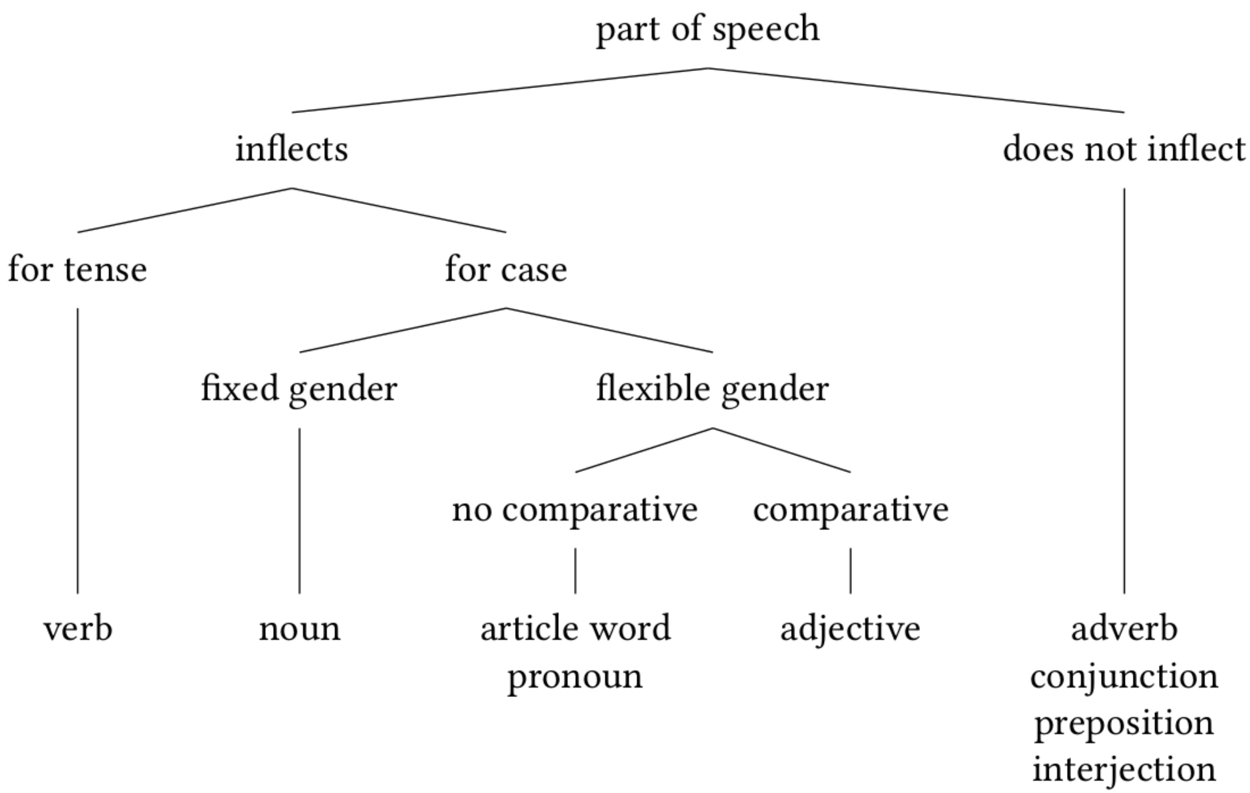
Note however: We don't know what the different constituent tests above would say about the constituency of *kaparla-nha wartirra-ku-nha*. This is only possible with a detailed knowledge of how the language is used.

Definition: Parts of Speech

Parts of Speech are classes of words that each lexical item is assigned to according to its *morphosyntactic* properties. According to Müller (2019: 18) the basic POS are *Verb*, *Noun*, *Adjective*, *Adverb*, *Prepositions*.



Decision Tree



Müller (2019). Grammatical theory, p. 24.
Based on Duden Grammar by Eisenberg et al. (2005).

Summary: Problems with POS

- **Problem 1**: The number of basic POS can differ according to the framework any particular researcher adheres to (e.g. Interjection, Conjunction, etc. might be seen as additional POS).
- **Problem 2**: It is controversial whether all languages even have the basic POS mentioned above.
- **Problem 3**: The abbreviations used for POS can also differ across frameworks.
- **Problem 4**: Isolating languages have very little or no inflections. According the the Decision Tree all words in these languages would be in the class of adverbs, conjunctions, etc.

Headedness

"The **head** of a constituent/phrase is the element which determines the *most important properties* of the constituent/phrase. At the same time, the head also determines the *composition of the phrase*. That is, the head requires certain other elements to be present in the phrase."
Müller (2019). Grammatical theory, p. 28.

Examples:

(5) This man *dreams* in his sleep.
(6) this *man*
(7) *in* his sleep
(8) his *sleep*

The heads are here indicated in *italics*.

Overview: Heads and Phrase Types

Example	**Head**	**Phrase Type**
she **knows** the man	knows (V)	VP
he **is smart**	smart (A) or is (V)	AP or VP
smart **woman**	woman (N)	NP
the **woman**	woman (N)	NP
the man's **cat**	cat (N)	NP
very **beautiful**	beautiful (A)	AP
very **quickly**	quickly (Adv)	AdvP
in the library	in (P)	PP

Valency according to Tesnière

> *“Nous avons vu qu’il y avait des verbes sans actant, des verbes à un actant, des verbes à deux actants et des verbes à trois actants.”*
>
> *Tesnière (1959). Éléments de syntaxe structurale, p. 238.*

Verb	V	V	V	V
Arguments	_	A	A A	A A A
Sentence Type:	impersonal sentence	intransitive sentence	transitive sentence	ditransitive sentence
Valency:	avalent (0)	monovalent (1), one-place predicate	bivalent (2), two-place predicate	trivalent (3), three-place predicate

Note: Müller states that the pronouns in expletives (e.g. *it rains*) should be considered obligatory arguments of the verb, while Tesnière explicitely calls them “sans actant”.

Grammatical Functions: Subject and Object

"In some theories, grammatical functions such as **subject** and **object** form part of the formal description of language (see Chapter 7 on Lexical Functional Grammar, for example). [...] it is by no means a trivial matter to arrive at a definition of the word subject which can be used cross-linguistically."

Müller (2019). Grammatical theory, p. 35.

Grammatical Functions: Subject

The following syntactic properties defining a subject are cited by Müller:

- agreement of the finite verb with it
- nominative case in non-copular clauses
- omitted in infinitival clauses
- optional in imperatives

Müller (2019). Grammatical theory, p. 35.

Valency and Grammatical Functions

> "If we can be clear about what we want to view as a subject, then the definition of *object* is no longer difficult: objects are all other arguments whose form is directly determined by a given head. [...] it is commonplace to talk of *direct objects* and *indirect objects*. The direct object gets its name from the fact that – unlike the indirect object – the referent of a direct object is directly affected by the action denoted by the verb."
>
> Müller (2019), p. 38.

Verb	V	V	V	V
Arguments	_	A	A A	A A A
Gramm. Functions:	None or SUBJ	SUBJ	SUBJ, OBJ	SUBJ, DOBJ, IOBJ
Valency:	avalent (0)	monovalent (1)	bivalent (2)	trivalent (3)

Notation: DOBJ (direct object), IOBJ (indirect object)

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

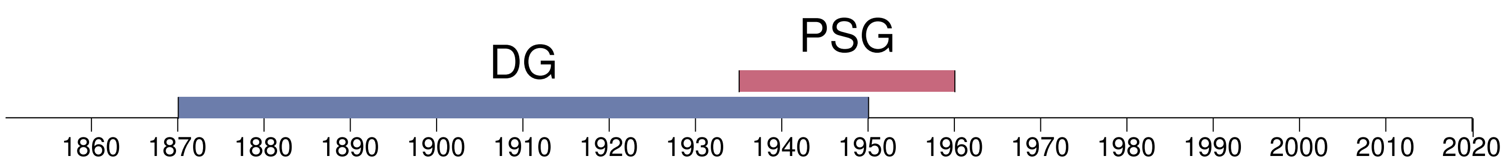
Faculty of Philosophy
General Linguistics

Syntactic Frameworks: Overview

Historical Perspective

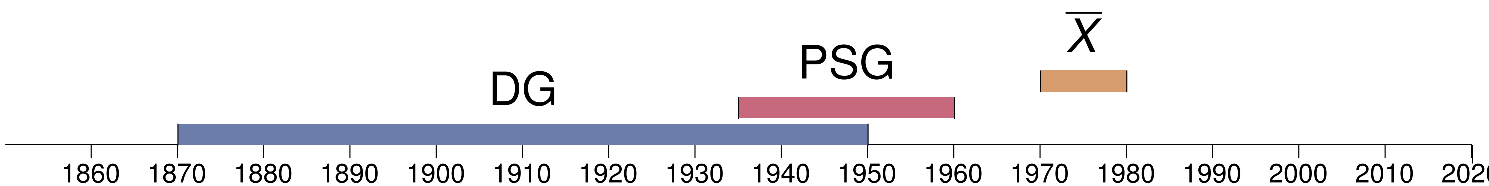
Most **basic syntactic concepts** (headedness, valency, POS, grammatical functions) were already relevant in **Dependency Grammar (DG)**.

Phrase Structure Grammar (PSG) added a strong constituency component via **re-write rules**. This also gave rise to **tree and bracket representations**.



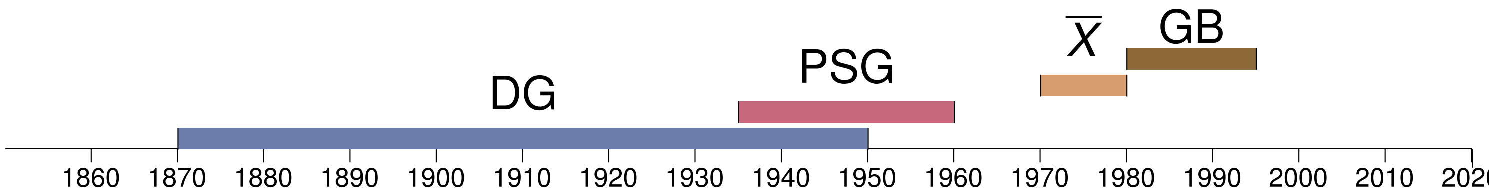
Historical Perspective

$\overline{X}$**-Theory** took PSGs to a **higher level of abstraction** by introducing $\overline{X}$-rules. Remember that the X is a variable representing all kinds of phrase types (AP, NP, PP, etc.)



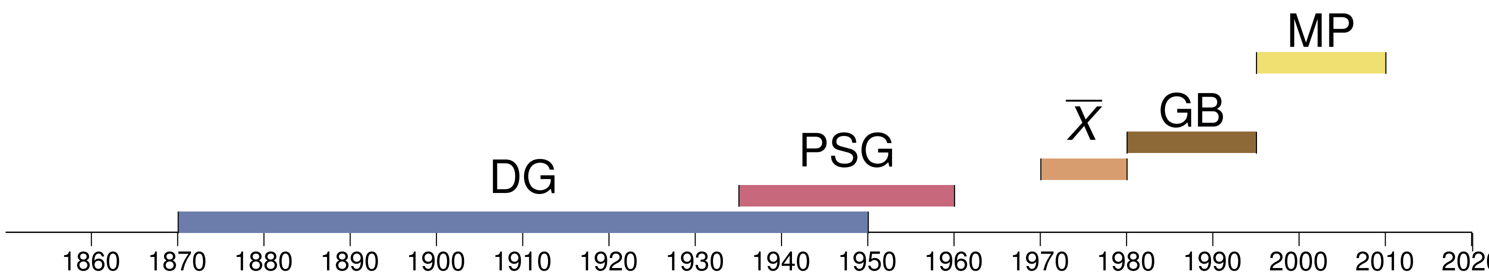
Historical Perspective

This tendency of further abstracting away from **surface structure** to understand **deep structure** was followed up on by **Government & Binding (GB)**. The principle of *government* is introduced to deal with case assignment, while *binding* deals with anaphora resolution. The field quickly fragmented into different definitions of such principles.



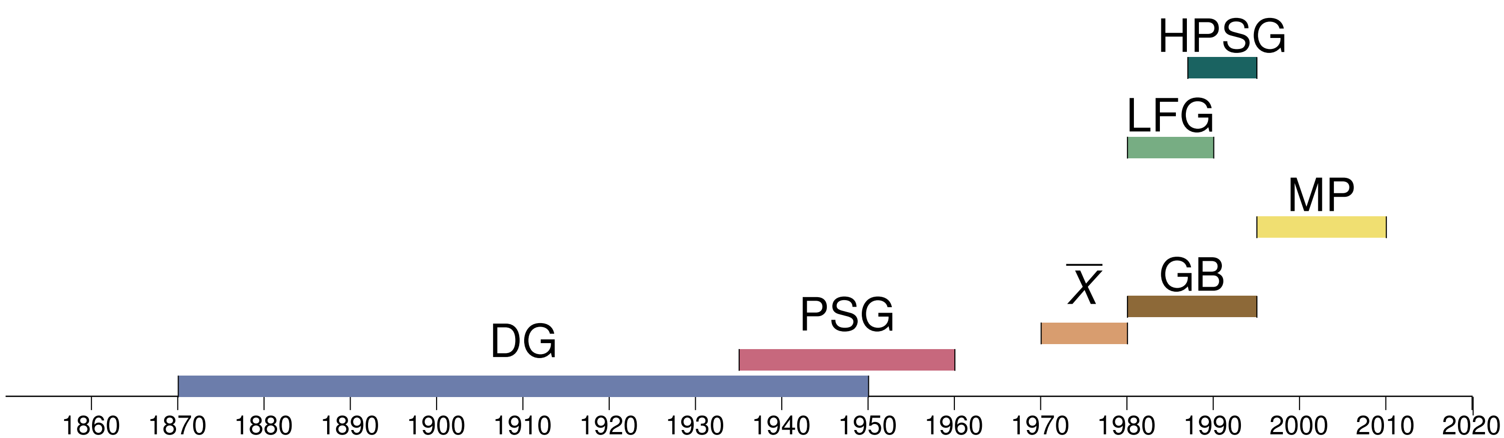
Historical Perspective

The **Minimalist Program (MP)** then strongly reduces the GB aparatus in order to base syntactic theory on a few core operations (i.e. merge and move). Another divergence from GB and X-bar theory is that it uses features for structure building (rather than phrase structure rules).



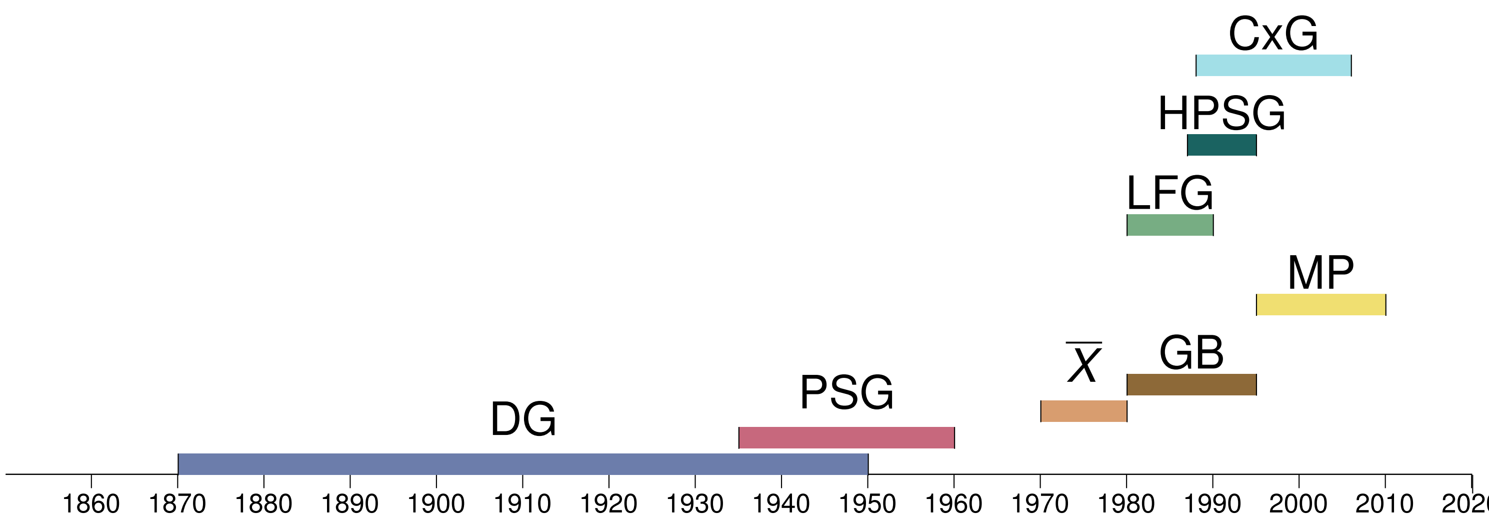
Historical Perspective

Lexical Functional Grammar (LFG) and **Head-Driven Phrase Structure Grammar (HPSG)** rather focus on **lexicalization of syntactic structure** by introducing **feature descriptions in matrix form**. This also rendered tree/bracket notations rather marginal.

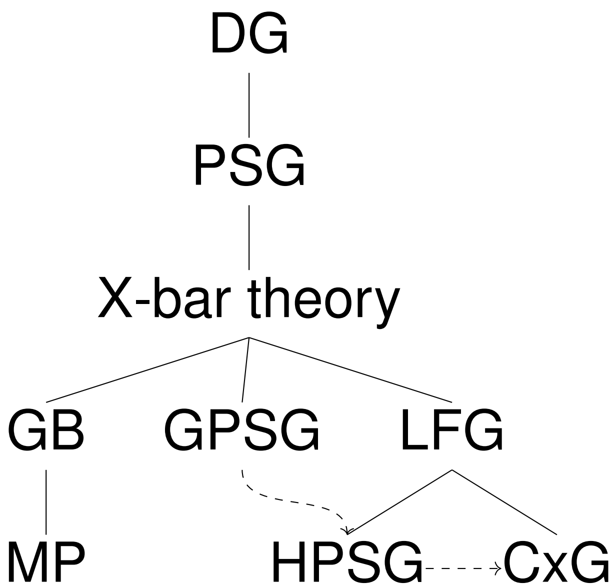


Historical Perspective

Construction Grammar breaks with a core concept of syntax, and promotes moving away from **compositionality** towards **holistic patterns**, i.e. constructions, which are learned and stored if sufficiently frequent.



Syntactic Framework Tree



DG: Dependency Grammar
PSG: Phrase Structure Grammar
GB: Government & Binding
GPSG: Generalized Phrase Structure Grammar
LFG: Lexical Functional Grammar
HPSG: Head-Driven Phrase Structure Grammar
CxG: Construction Grammar
MP: Minimalist Program

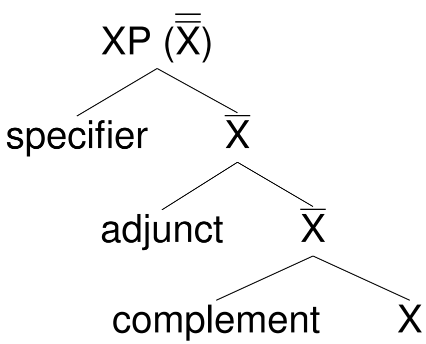
Basic Concepts in Syntactic Frameworks

	Const.	POS	Heads	Valency	Gram. Functions
DG	x	✓	✓	✓	✓
PSG	✓	✓	✓	✓	x
X-bar	✓	✓	✓	✓	✓
G&B	✓	✓	✓	✓	✓
MP	✓	✓	✓	✓	✓
LFG	✓	✓	✓	✓	✓
HPSG	✓	✓	✓	✓	✓
C&G	✓	✓	✓	x	✓

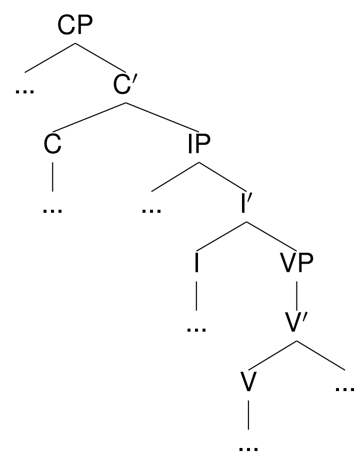
Transformational Frameworks

The core idea of **transformational frameworks** is that there is some **underlying template** (i.e. deep structure) which is adapted by transformations and movements to give rise to the full variety of sentence structures encountered in linguistic production (except for noise such as misspronunciations, etc.).

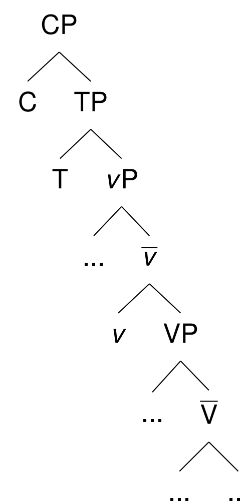
X-bar Theory



G&B



Minimalist Program



Constraint-Based Frameworks

The core idea of **constraint-based frameworks**[1] is to capture syntactic relationships by **structural frames** (e.g. feature matrices, constructions) which constrain how elements can be combined and slots are filled.

LFG

$$\begin{bmatrix} \text{PRED} & \text{'devour}\left\langle \text{SUBJ,OBJ} \right\rangle\text{'} \\ \text{TENSE} & \textit{past} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'david'} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{SPEC} & \text{a} \\ \text{PRED} & \text{'sandwich'} \end{bmatrix} \end{bmatrix}$$

HPSG

$$\begin{bmatrix} \textit{head-specifier-phrase} & \\ \text{PHON} \left\langle \textit{Kim sleeps} \right\rangle & \\ \textbf{SYNSEM|LOC|CAT} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \boxed{1} \\ \text{SPR} & \left\langle \boxed{2} \right\rangle \\ \text{COMPS} & \left\langle \right\rangle \end{bmatrix} \\ \text{HEAD-DTR} \begin{bmatrix} \ldots \end{bmatrix} & \\ \text{NON-HEAD-DTRS} \left\langle \ldots \right\rangle & \end{bmatrix}$$

C&G

- **[N-s]** (regular plurals)
- send **<someone>** to the cleaners
- the **Xer** the **Yer**
- **Subj V Obj$_1$ Obj$_2$**

[1] Also sometimes called *model theoretic*.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Syntax:
Current Research and Applications

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Faculty of Philosophy
General Linguistics

Dependency Grammar

Universal Dependencies

Universal Dependencies (UD) is a framework for consistent annotation of grammar (parts of speech, morphological features, and syntactic dependencies) across different human languages. UD is an open community effort with over 200 contributors producing more than 100 treebanks in over 70 languages. If you're new to UD, you should start by reading the first part of the Short Introduction and then browsing the annotation guidelines.

- Short introduction to UD
- UD annotation guidelines
- More information on UD:
  - How to contribute to UD
  - Tools for working with UD
  - Discussion on UD
  - UD-related events
- Query UD treebanks online:
  - SETS treebank search maintained by the University of Turku
  - PML Tree Query maintained by the Charles University in Prague
  - Kontext maintained by the Charles University in Prague
  - Grew-match maintained by Inria in Nancy
  - INESS maintained by the University of Bergen
- Download UD treebanks

If you want to receive news about Universal Dependencies, you can subscribe to the UD mailing list. If you want to discuss individual annotation questions, use the Github issue tracker.

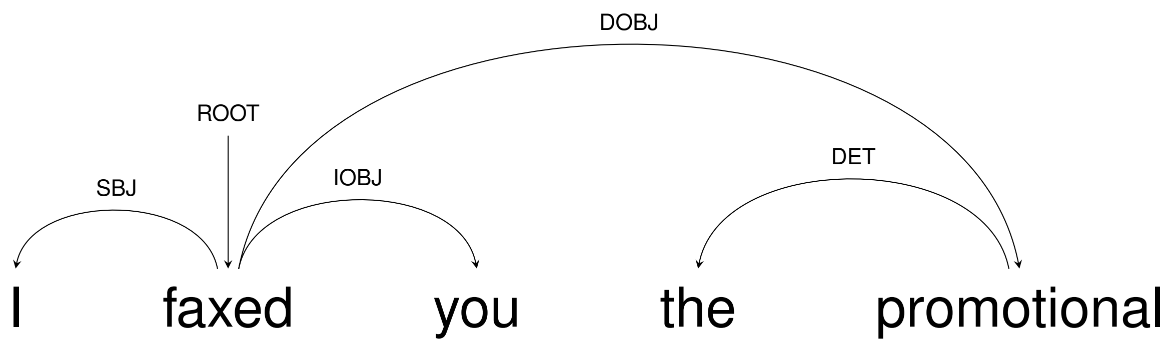
Current UD Languages

Information about language families (and genera for families with multiple branches) is mostly taken from WALS Online (IE = Indo-European).

Afrikaans	1	49K		IE, Germanic
Akkadian	1	1K		Afro-Asiatic, Semitic
Amharic	1	10K		Afro-Asiatic, Semitic
Ancient Greek	2	416K		IE, Greek
Arabic	3	1,042K	W	Afro-Asiatic, Semitic
Armenian	1	36K		IE, Armenian
Assyrian	1	<1K		Afro-Asiatic, Semitic
Bambara	1	13K		Mande
Basque	1	121K		Basque
Belarusian	1	13K		IE, Slavic
Breton	1	10K	W	IE, Celtic
Bulgarian	1	156K		IE, Slavic
Buryat	1	10K		Mongolic
Cantonese	1	13K		Sino-Tibetan
Catalan	1	531K		IE, Romance

Example Sentence

Lecture Notation:



Universal Dependencies Notation:

sent_id = email-enronsent00_02-0047
text = I faxed you the promotional [...]

ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	DEPREL	DEPS
1	I	I	PRON	PRP	Case=Nom|Number=Sing|Person=1|PronType=Prs	2	nsubj	2:nsubj
2	faxed	fax	VERB	VBD	Mood=Ind|Tense=Past|VerbForm=Fin	0	root	0:root
3	you	you	PRON	PRP	Case=Acc|Person=2|PronType=Prs	2	iobj	2:iobj
4	the	the	DET	DT	Definite=Def|PronType=Art	5	det	5:det
5	prom.	prom.	NOUN	NN	Number=Sing	2	obj	2:obj

Quantitative Typology of Word Order

> *In most typological research, languages have been treated as single data points with a categorical value (e.g. OV or VO, prepositional or postpositional). [...] By using **continuous** variables instead of **categorical** ones, it is possible to capture **intra-linguistic variation**, which is ubiquitous in language, at the same time avoiding the existing bias towards a restricted set of linguistic patterns [...]*

Levshina (2019). Token-based typology and word order entropy: A study based on Universal Dependencies.

Word Order in WALS

Hawaiian: VSO
Maori: VSO

Mandarin: SVO
Hakka: SVO

Kara-Kalpak: SOV
Kumyuk: NA

Word Order in Parallel Texts

SOV	SVO	OSV	OVS	VSO	VOS
Polynesian (Hawaiian, Maori)					
3	31	2	2	70	3
6	26	5	4	76	18
Sinitic (Mandarin, Hakka)					
54	235	6	0	3	5
18	84	1	2	5	3
Turkic (Kara-Kalpak, Kumyk)					
114	2	8	7	0	0
89	1	12	11	4	1

Table 1: Number of transitive clauses with a given order of subject/object/verb, according to our algorithm, for six languages (from three families).

Östling (2015). Word order typology through multilingual word alignment.

Case Marking vs. Word Order

There is a trade-off between **marking of core cases on subject and object** (x-axis) and the **ordering entropy** (a measure of predictability) of subject and object (y-axis). In other words, languages which mark subject and object explicitely) by trend allow more flexibility in the ordering.

Levshina (2019).

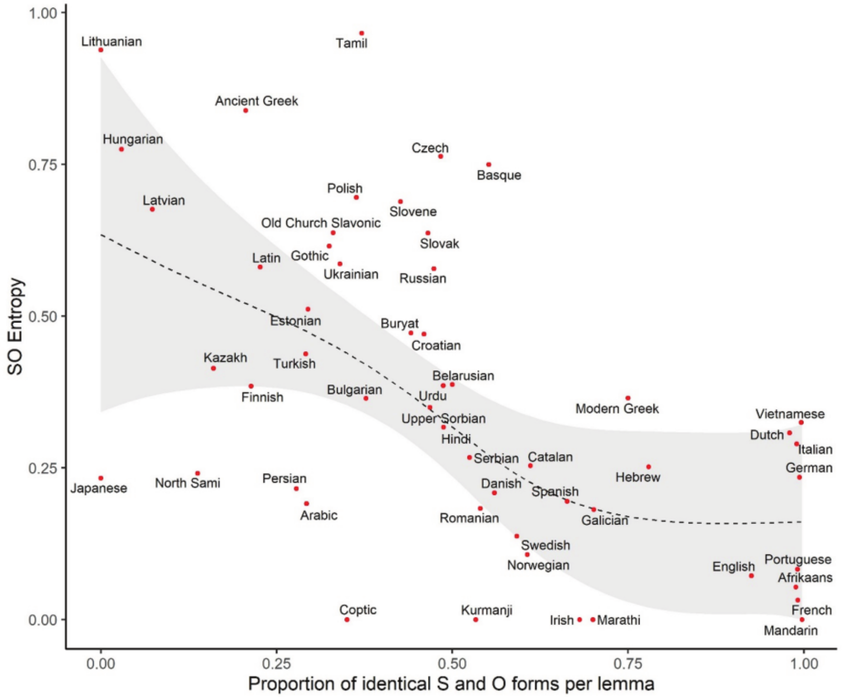


Figure 3: Negative correlation between SO entropy and the proportion of identical S and O forms.

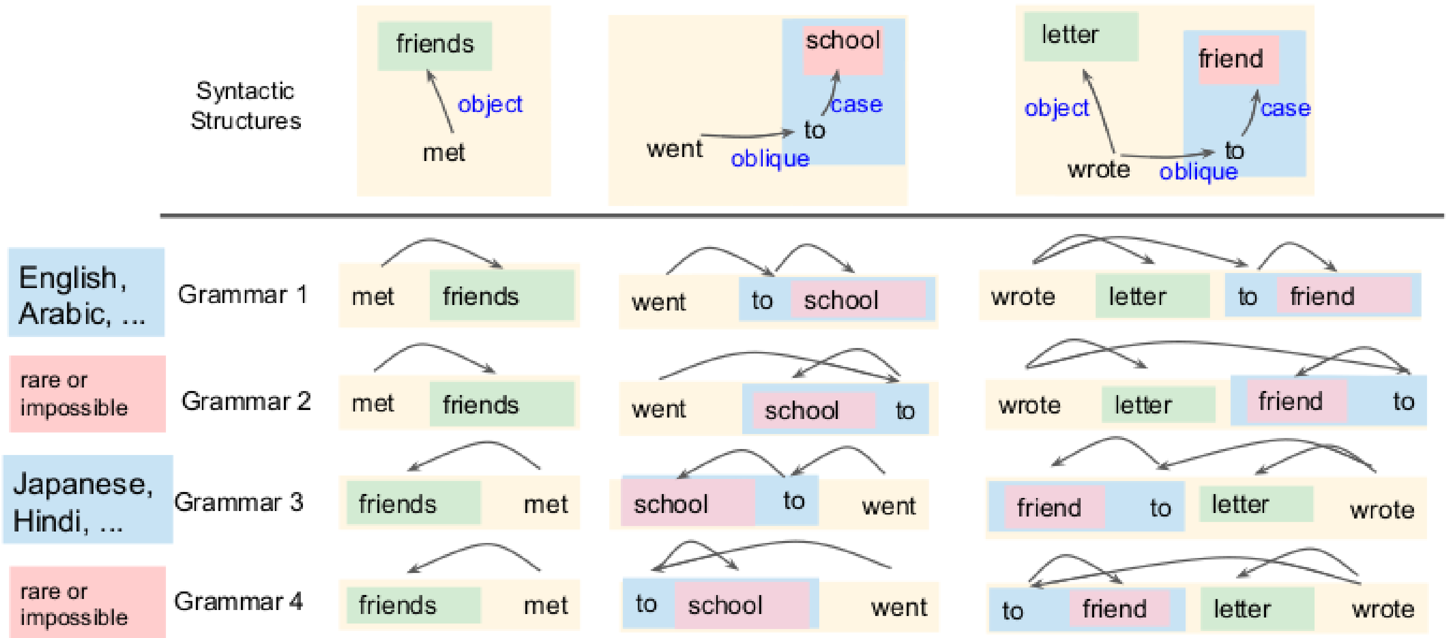
Greenbergian Word Order Universals

Table 1. Greenberg word-order correlations, exemplified by Arabic (left) and Japanese (right) examples

	Arabic (English, . . .)		Japanese (Turkish, . . .)	
Correlation no.	Correlates with verb	Correlates with object	Correlates with object	Correlates with verb
	kataba	risāla	tegami-o	kaita
	wrote	*letter*	*letter*	*wrote*
①	li	ṣadīq	tomodachi	ni
	to	*a friend*	*friend*	*to*
②	kāna	ṣadīq	tomodachi	datta
	was	*a friend*	*friend*	*was*
③	sawfa	yaktub	kak-	-udesho
	will	*write*	*write*	*will*
④	ṣadīq	John	John no	tomodachi
	friend	*of John*	*John of*	*friend*
⑤	kutub	taqra'uhā	anata-ga yonda	hon
	books	*that you read*	*that you read*	*book*
⑥	'an	tuṣil	toochaku suru	koto
	that	*she arrives*	*arrives*	*that*
⑦	dhahabt	'ilā lmadrasa	gakkoo ni	itta
	went	*to school*	*school to*	*went*
⑧	'urīd	'an 'ughādir	ik-	-itai
	wants	*to leave*	*to go*	*want*

Across the world, the orders of different constituents are strikingly correlated with that of verb and object. Selection is based on a more recent typological study by Dryer (13), restricted to those correlations that are annotated in available corpus data. See *SI Appendix*, section S1 for more on Greenberg correlations.

Hahn, Jurafsky, and Futrell (2020). Universals of word order reflect optimization of grammars for efficient communication.



Hahn, Jurafsky, and Futrell (2020).

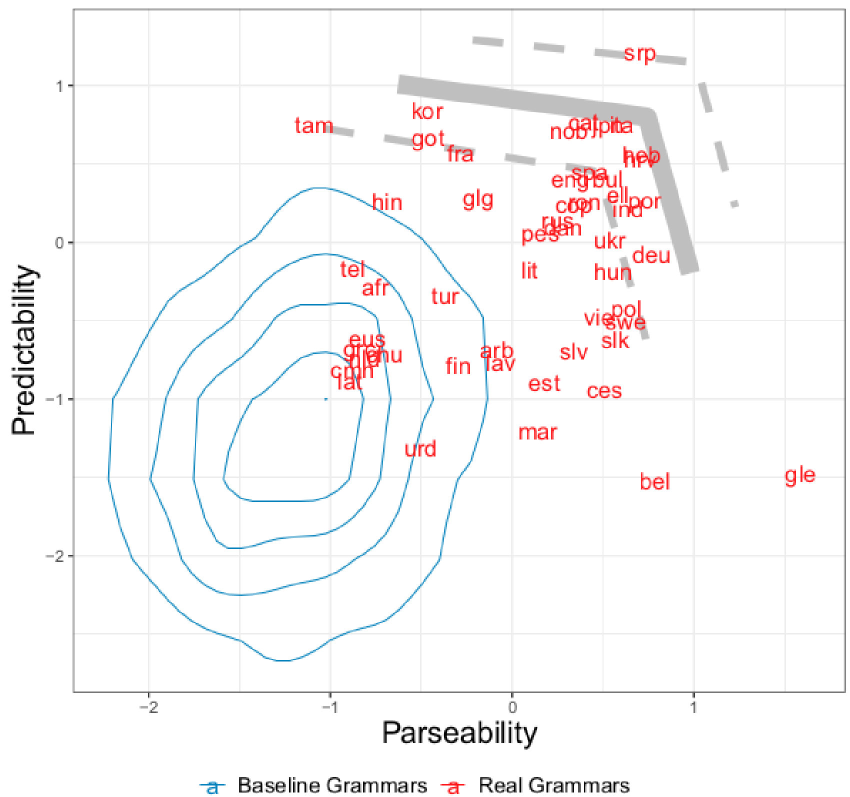
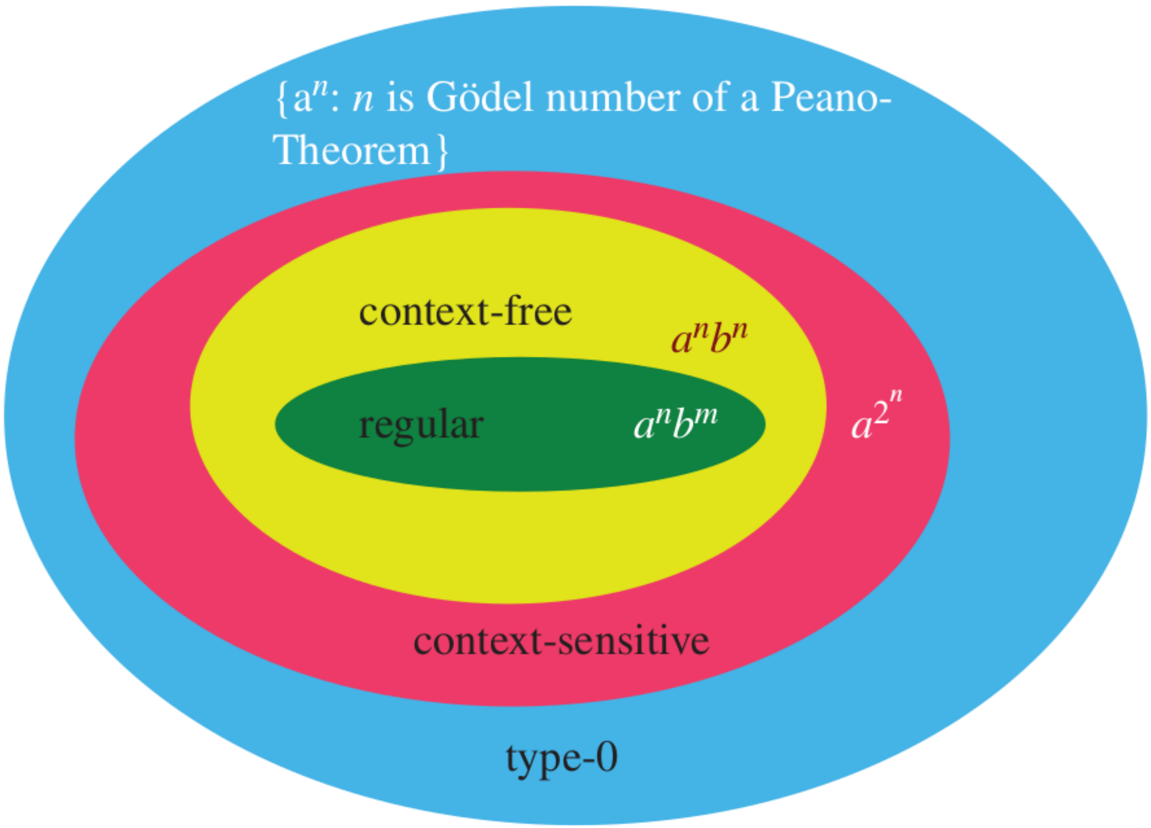


Fig. 4. Predictability and parseability of the real word-order grammars of 51 languages (red), indicated by International Organization for Standardization codes, compared to baseline word-order grammars (blue distribution). Predictability and parseability scores are *z*-scored within language, to enable comparison across languages. The gray curve indicates the approximate Pareto frontier of computationally optimized grammars, averaged over the 51 languages, with dashed SDs.

Hahn, Jurafsky, and Futrell (2020).

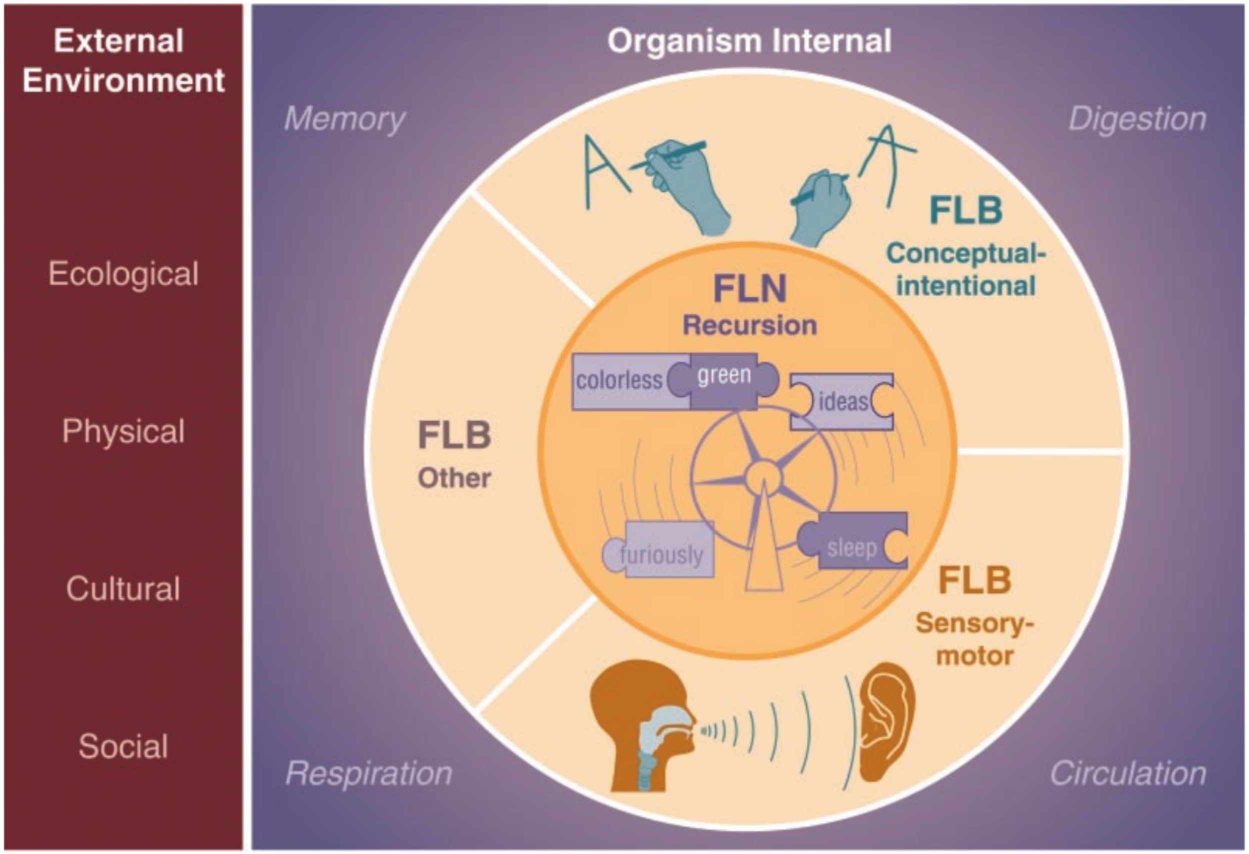
The Chomsky Hierarchy

The Chomsky Hierarchy



Chomsky (1956). Three models for the description of language.
Jäger & Rogers (2012). Formal language theory: refining the Chomsky hierarchy.

What is in the Faculty of Language (Narrow Sense)?



Hauser, Chomsky & Fitch (2002). The faculty of language: What is it, who has it, and how did it evolve?

Is the Human Language Capacity **Domain-Specific?**

"Here, we explored human pattern-processing capabilities in the **visual domain** by generating abstract visual sequences [...] Participants successfully acquired all three grammars after only minutes of exposure [...] cognitive mechanisms with the computational power to process linguistic syntax are **not specific to the domain of language** [...]"

Westphal-Fitch et al. (2018).

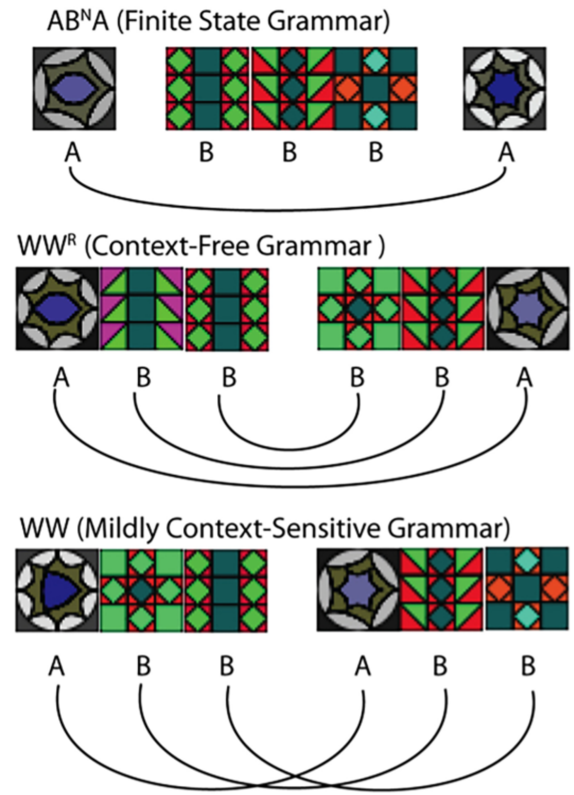
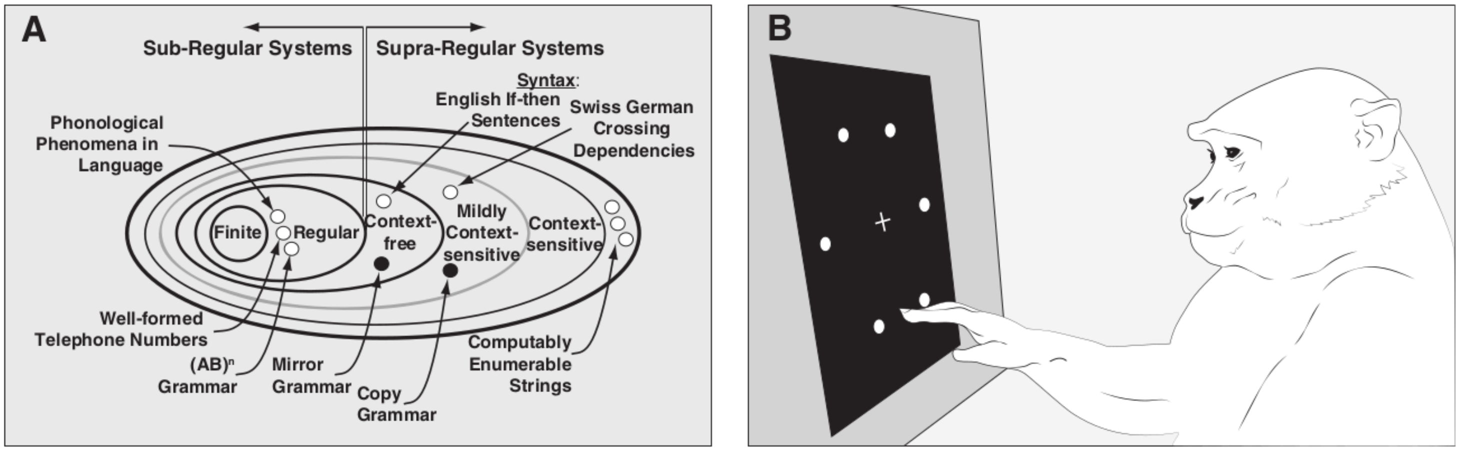


FIGURE 2 | Examples of stimuli sequences for all three grammars with $N = 3$.

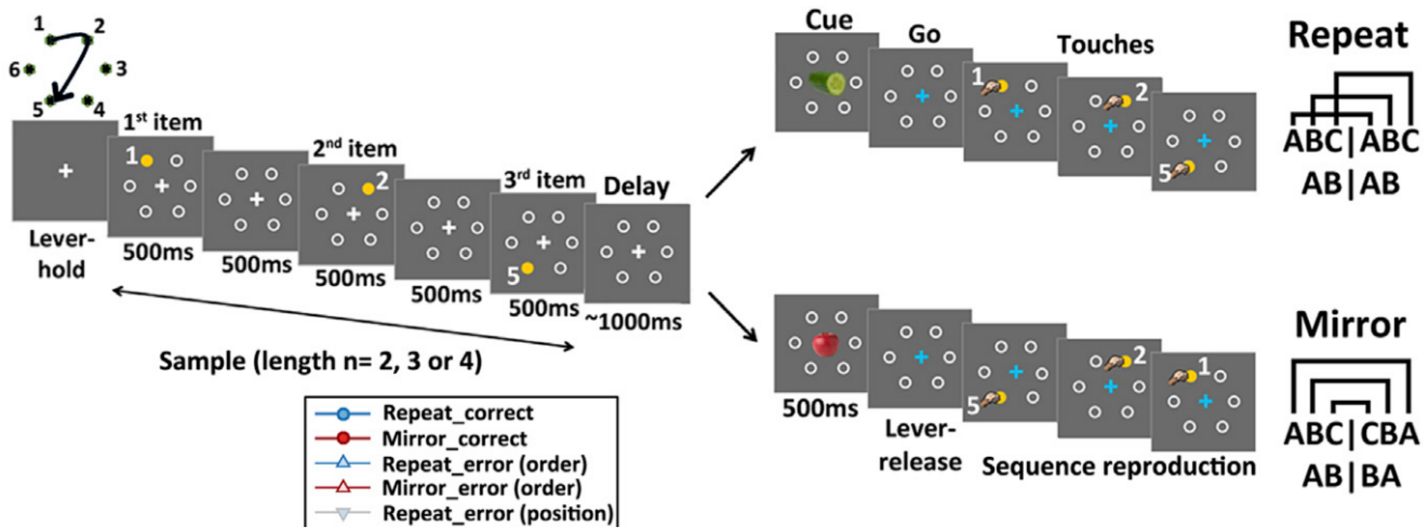
Is the Human Language Capacity Species-Specific?



“Macaque monkeys can be trained to produce complex spatial sequences beyond the simplest levels of grammar previously known from animal studies. This indicates cognitive capabilities in the spatial-motor domain that approach the computational complexity level of human syntax.”

Fitch (2018). Bio-linguistics: Monkeys break through the syntax barrier.
Jiang et al. (2018). Production of supra-regular spatial sequences by Macaque monkeys.

Is the Human Language Capacity Species-Specific?



"Using a production task, we show that macaque monkeys can be trained to produce **time-symmetrical embedded spatial sequences** whose formal description requires **supra-regular grammars** or, equivalently, a push-down stack automaton."

Jiang et al. (2018). Production of supra-regular spatial sequences by Macaque monkeys.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

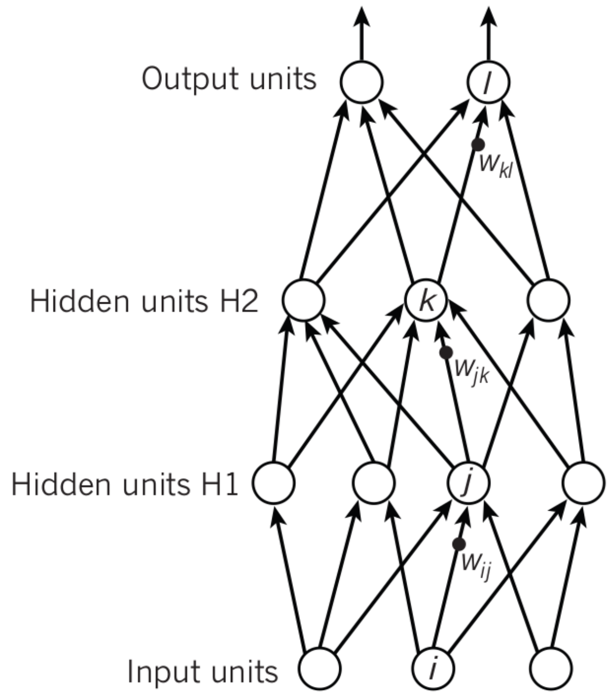
Faculty of Philosophy
General Linguistics

Syntax in Neural Networks

The Advent of Deep Learning

"The hidden layers of a multilayer neural network learn to represent the network's inputs in a way that makes it easy to predict the target outputs. This is nicely demonstrated by training a multilayer neural network to predict the next word in a sequence from a local context of earlier words [...]"

LeCun et al. (2015). Deep learning.



$$y_l = f(z_l)$$
$$z_l = \sum_{k \, \varepsilon \, H2} w_{kl} y_k$$

$$y_k = f(z_k)$$
$$z_k = \sum_{j \, \varepsilon \, H1} w_{jk} y_j$$

$$y_j = f(z_j)$$
$$z_j = \sum_{i \, \varepsilon \, \mathrm{Input}} w_{ij} x_i$$

How much Syntax do Neural Nets Learn?

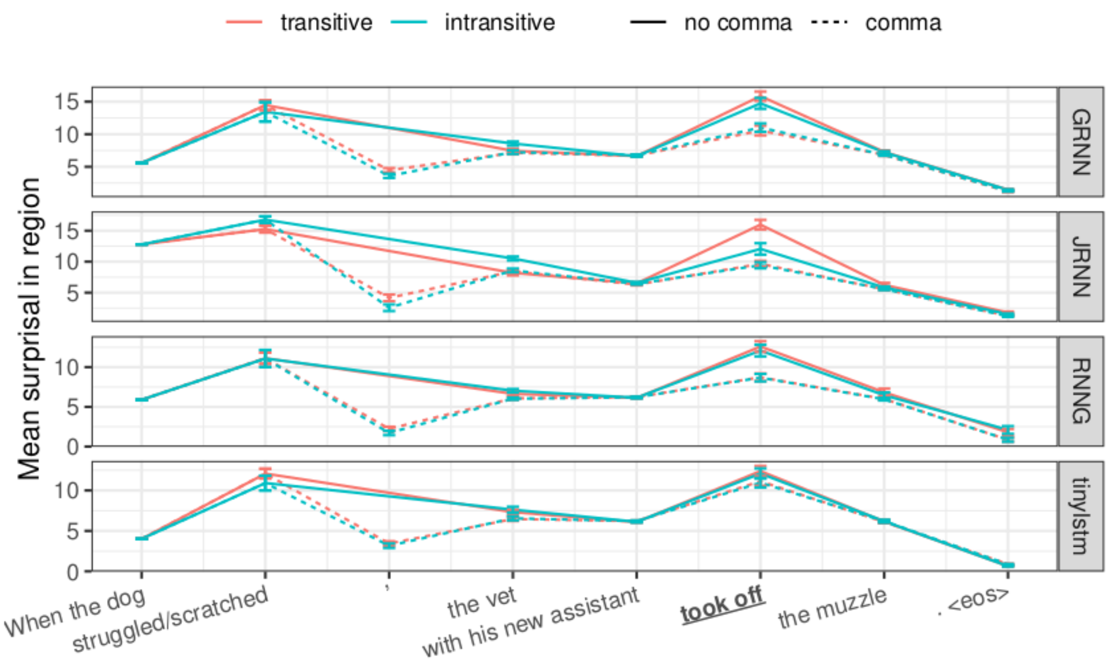
"From a linguist's perspective, the applied success of deep neural networks (DNNs) is striking because, unlike the systems that were popular in NLP a decade ago [...], DNNs' input data and architectures are not based on the symbolic representations familiar from linguistics, such as **parse trees** or **logical formulas**. Instead, DNNs learn to encode **words and sentences as vectors** (sequences of real numbers); these vectors [...] are then transformed through a series of simple arithmetic operations to produce the network's output.

[...]

In this review, we focus on work that directly evaluates DNNs' syntactic knowledge [...] this body of work suggests that contemporary DNNs can learn a **surprising amount about syntax** but that this ability falls short of human competence [...]

Linzen & Baroni. (2020). Syntactic Structure from Deep Learning.

Syntactic Phenomena in Neural Nets



“All models show NP/Z garden path effects, indicating that they are sensitive to some cues indicating end-of-clause boundaries. However, only the large LSTMs [GRNN, JRNN] appear to use **verb argument structure** [transitivity] information as a cue to these boundaries.”

Futrell et al. (2019). Neural language models as psycholinguistic subjects.

Do Syntactic Trees help NLP Tasks?

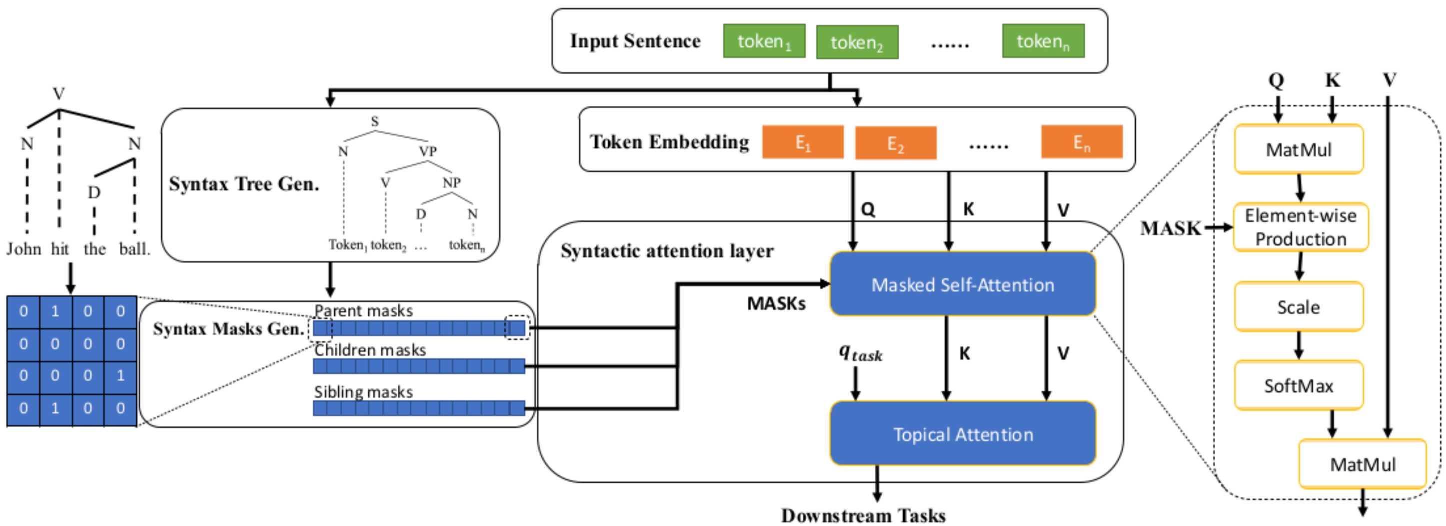


Figure 1: The Overall Architecture of Syntax-BERT. Note that the leftmost part shows an example of syntax tree and its corresponding parent syntax mask ($d = 1$).

"The most commonly-used syntax trees are constituency trees (Chen and Manning, 2014) and dependency trees (Zhu et al., 2013), and we use both of them in our experiments unless notified."

Bai et al. (2021). Syntax-BERT: Improving Pre-trained Transformers with Syntax. Trees

Do Syntactic Trees help NLP Tasks?

Model	Avg	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m/-mm	QNLI	RTE	WNLI
Transformer	66.1	31.3	83.9	81.7/68.6	73.6/70.2	65.6/84.4	72.3/71.4	80.3	58.0	65.1
Syntax-Transformer (Ours)	**68.8**	**36.6**	**86.4**	**81.8/69.0**	**74.0/72.3**	**65.5/84.9**	**72.5/71.2**	**81.0**	**56.7**	65.1
BERT-Base	77.4	51.7	93.5	87.2/82.1	86.7/85.4	71.1/89.0	84.3/83.7	90.4	67.2	65.1
Syntax-BERT-Base (Ours)	**78.5**	**54.1**	**94.0**	**89.2/86.0**	**88.1/86.7**	**72.0/89.6**	**84.9/84.6**	**91.1**	**68.9**	65.1
BERT-Large	80.5	60.5	94.9	89.3/85.4	87.6/86.5	72.1/89.3	**86.8/85.9**	92.7	70.1	65.1
Syntax-BERT-Large (Ours)	**81.8**	**61.9**	**96.1**	**92.0/88.9**	**89.6/88.5**	**72.4/89.5**	86.7/86.6	**92.8**	**74.7**	65.1
RoBERTa-Base	80.8	57.1	95.4	90.8/89.3	88.0/87.4	72.5/89.6	86.3/86.2	92.2	73.8	65.1
Syntax-RoBERTa-Base (Ours)	**82.1**	**63.3**	**96.1**	**91.4/88.5**	**89.9/88.3**	**73.5/88.5**	**87.8/85.7**	**94.3**	**81.2**	65.1
RoBERTa-Large	83.9	63.8	96.3	91.0/89.4	72.9/90.2	72.7/90.1	89.5/89.7	94.2	84.2	65.1
Syntax-RoBERTa-Large (Ours)	**84.7**	**64.3**	**96.9**	**92.5/90.1**	**91.6/91.4**	**73.1/89.8**	**90.2/90.0**	**94.5**	**85.0**	65.1
T5-Large	86.3	61.1	96.1	92.2/88.7	90.0/89.2	74.1/89.9	89.7/89.6	94.8	87.0	65.1
Syntax-T5-Large (Ours)	**86.8**	**62.9**	**97.2**	**92.7/90.6**	**91.3/90.7**	**74.3/90.1**	**91.2/90.5**	**95.2**	**89.6**	65.1

Table 4: Comparison with state-of-the-art models without pre-training on GLUE benchmark.

“Experiments on various datasets of natural language understanding verify the effectiveness of syntax trees and achieve consistent improvement over multiple pre-trained models, including BERT, RoBERTa, and T5.”

Bai et al. (2021). Syntax-BERT: Improving Pre-trained Transformers with Syntax Trees

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Faculty of Philosophy
General Linguistics

References

References

Adger, D. (2003). *Core Syntax. A minimalist approach.* Oxford: Oxford University Press.

Bai et al. (2021). Syntax-BERT: Improving Pre-trained Transformers with Syntax Trees. *arXiv*.

Chomsky, Noam (2015). *The minimalist program*. 20th anniversary edition. Cambridge, Massachusetts: MIT Press.

Fitch, Tecumseh (2018). Bio-Linguistics: Monkeys break through the syntax barrier. *Current Biology*.

Futrell, Richard et al. (2019). Neural language models as psycholinguistic subjects: Representations of syntactic state. *Proceedings of NAACL-HLT*.

Goldberg, Adele (2006). *Constructions at work*. The nature of generalization in language. Oxford: Oxford University Press.

Hahn, M. , Jurafsky, M., and Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *PNAS*, 117 (5), 2347-2353..

Jiang, Xinjian et al. (2018). Production of supra-regular spatial sequences by Macaque monkeys. *Current Biology*.

LeCun, Yann, Bengio, Yoshua & Hinton, Geoffrey (2015). Deep learning. *Nature*.

Levshina, Natalia (2019). Token-based typology and word order entropy: A study based on Universal Dependencies. *Linguistic Typology*.

References

Linzen, Tal, & Baroni, Marco (2020). Syntactic structure from deep learning. *Annual Review of Linguistics*.

Müller, Stefan. 2019. *Grammatical theory: From transformational grammar to constraint-based approaches. Third revised and extended edition.* ***Volume I.*** Berlin: Language Science Press.

Östling (2015). Word order typology through multilingual word alignment. *Proceedings of ACL*.

Velupillai, Viveka (2012). *An introduction to linguistic typology*. Amsterdam/Philadelphia: John Benjamins.

Westphal-Fitch, G. et al. (2018). Artificial grammar learning capabilities in an abstract visual task match requirements for linguistic syntax. *Frontiers in Psychology*.

Thank You.

Contact:

Faculty of Philosophy
General Linguistics
Dr. Christian Bentz
SFS Wilhelmstraße 19-23, Room 1.24
chris@christianbentz.de
Office hours:
During term: Wednesdays 10-11am
Out of term: arrange via e-mail