Heritage grammars and linguistic complexity: A view from grammatical gender

Cambridge Linguistic Society

Terje Lohndal & Mike Putnam

NTNU/UiT & Penn State University

January 28, 2021

Heritage languages as a window onto mental grammars

- Linguistic theory should contain a 'specification of the class of potential grammars' (Chomsky 1965: 24)
- This also includes the range multilingual speakers, even though formal work has mostly focused on the 'ideal speaker-listener' of Chomsky (1965) (Lohndal 2013, Lohndal et al. 2019).
- '[H]eritage speakers constitute an outcome often assumed to be impossible outside of pathology or trauma: children exposed to a language from birth who nevertheless appear to deviate from the expected native-like mastery in pronounced and principled ways' (Polinsky & Scontras 2019: 2).

The dynamic nature of heritage grammars

- Morphosyntax is one of the areas in heritage grammars that is often subject to change compared with a given baseline (e.g., Montrul 2016, Polinsky 2018, Lohndal in press, Putnam et al. in press).
- The dynamic nature of this area makes is a fertile domain for investigating how mental grammars change across the lifespan of an individual speaker and across generations of speakers.
- It is often argued that heritage speakers reduce the complexity of their heritage grammars, e.g., loss of morphological richness (Benmamoun et al. 2013, Montrul 2016, Polinsky 2018) or employing fewer syntactic rules (see e.g., Polinsky 2011, Westergaard & Lohndal 2019).
- However, how do we understand **complexity**?

Some overall questions regarding complexity

- Q_1: How can we define, or even approximate, **complexity** in a mental grammar?
- Q_2: Why should we apply elements of a formal grammar system here to help us out, and which one should we select?
- Q_3: What sort of heuristic in combination with formal mechanisms can we make use of to measure **complexity**?

Outline

What we hope to achieve in this talk:

- Establish a working definition of **complexity** of linguistic structure and accompanying operations
  - Do this through the lens of late-insertion **exoskeletal** grammar, where structural properties are independent of lexical properties.
- Apply this heuristic to **heritage grammars**, specifically to the phenomenon of grammatical gender.

Complexity

- The nature of complexity is not a new question, cp. information science (Shannon, 1947).
- More recently, there has been a renewed interest in understanding complexity from a linguistic perspective, cf. DeGraff (2001), McWhorter (2001), Dahl (2004), Miestamo (2006, 2008), Sampson, Gil & Trudgill (2009), Newmeyer & Preston (eds. 2014), Audring (2014, 2017, 2019), and Di Garbo & Miestamo (2019).
- Some research provides descriptive generalizations rather than predictive models, though there are also some recent formal attempts with predictive power (e.g., Jakubowicz 2005, Jakubowocz and Strik 2008, Biberauer et al. 2014).

Defining complexity: An example

- Miestamo (2006, 2008): The number of elements in a system and connections between these
- From an information-theoretic point of view, complexity can be reduced to description length
  - The simpler entity can be compressed into a smaller space without losing information
- From a linguistic point, the notions of cost and difficulty of processing and learning are often associated with complexity.

Measuring grammatical complexity

- Miestamo (2006, 2008) proposes two principles (cf. Di Garbo 2014, 2016):
  - **The Principle of Fewer Distinctions**
    Fewer semantic/pragmatic distinctions are grammatically encoded (focus on grammatical meaning)
  - **The Principle of One-Meaning-One-Form**
    Each meaning is expressed by one form, and each form corresponds to only one meaning (focus on the meaning-form relationship)

Application: Grammatical gender

- Fewer Distinctions: 2 is less complex than 3, 4, ...
- One-Meaning-One-Form: More complex if the expression of gender involves:
  - fusion/multiple exponence (gender + other features)
  - fission (discontinuous/multiple morphemes)
  - allomorphy
  - syncretism (Di Garbo & Miestamo 2019: 17-18)
- A third principle: **The Principle of Independence**:
  Systems that are independent of other systems and structures are less complex. (Di Garbo 2014, 2016, Audring 2019)
  - A gender system whose formal realization is dependent on number distinctions is more complex than a system where this is not the case.

Larger questions about complexity

- A useful point of departure could be Miestamo's perspective: The number of elements in a system and connections between these.
- However, operationalizing this is less trivial:
  - Is a grammar with more functional heads more complex than one with fewer?
  - What about the realization of those heads - if one morpheme realizes many heads as opposed to one head, is the former more complex than the latter? What is the 'total' complexity of such a system?
- Where is complexity located?
  - The computational system?
  - Number of functional projections?
  - Overt realization of functional structure?
  - Mapping to the interfaces?

An exoskeletal approach to grammar

- Emphasis on the way in which syntactic structure determines both the grammatical properties and 'the ultimate fine-grained meanings of lexical items themselves' (Borer 2003: 33).
- A syntax-driven approach to structure building, in contrast to a lexicon-driven approach.
- All exoskeletal approaches adopt some version of a realizational approach to morphology, where morphosyntactic properties license inflectional exponents (cf. Stump 2001 for an overview).

(1) $[\ \alpha\ \beta\ \gamma\ ] \leftrightarrow /\mathrm{X}/$
synsem features — exponent (Embick 2015: 9)

Delimiting **complexity** in exoskeletal grammars

- The exoskeletal architecture allows us to separate the functional features from associated exponents.
- Complexity can be measured in terms of the number of functional features, but more importantly,
- core differences between languages reside in the mapping from synsem features to exponents.
    - This mapping is often not transparent (i.e., 1:1), and non-transparent mappings can be considered more complex (they are e.g., harder to acquire, they require more complicated rules, etc.).

Criteria for complexity

- The computational mechanisms/functions are assumed to be identical across all languages (Merge, Agree, etc.), thus there are no differences in complexity as far as the computational system is concerned
- However, what the functions operate over may differ:
  - More functional features in the syntax yields a more complex grammar.
  - One-Form-One-Meaning mappings are simpler than alternatives.
- Investigating the underlying structures of vulnerable and non-vulnerable areas in heritage grammars offers a way to move forward in terms of trying to develop a metric for complexity.

Some difficulties in operationalizing complexity

- What is vulnerable is not necessarily complex and what is non-vulnerable is not necessarily simple (think of Verb Second).
- Needlessly, complexity will have to be a relative notion.
  - L2 speakers of English struggle to acquire subject-verb agreement, which is salient and transparently encoded.
  - In heritage speakers, what is complex will also depend on what the other interacting languages are. Word order will be less vulnerable if the word orders of the two languages overlap.

A way to model reduced complexity: L2

- The absence of overt morphology does not necessarily entail the absence of functional categories in the syntax.
- Rather, the learner may not have established a complete set of exponents, or the learner fails to meet the matching conditions between the exponent and the structure (Lardiere 2000).
- **Missing Surface Inflection Hypothesis** (Lardiere 2000, Prévost & White 2000) and more generally the **Feature Reassembly Hypothesis** (Lardiere 1998, 2005, 2008, 2009, Slabakova 2009, 2016, and much other work).

Separating syntax and morphology

- We need a separationist approach which distinguishes between the underlying features and the actual exponents.
- Similar to late-insertion approaches like Distributed Morphology: Syntactic structure is generated prior to insertion of Vocabulary Items (e.g., Halle & Marantz 1993, Embick & Noyer 2007, Embick 2015).
- Applied to multilingual situations (Riksem 2017, Grimstad et al. 2018, Riksem 2018, Riksem et al. 2018, Lohndal et al. 2019, Putnam, Perez-Cortes & Sánchez 2019, E. Vanden Wyngaerd 2021).

A start: Possible outcomes in heritage grammars

- Relative to a given baseline, a feature can be retained
  - in the same hierarchical position
  - and shift its hierarchical position
- Relative to a given baseline, a feature can be lost
  - in its original hierarchical position
  - and shift its hierarchical position

	Feature	**Functional sequence**	**Exponent**
Retained			
Lost			

- A feature can also be restructured, e.g., expressing fewer distinctions, or shift its position in a feature geometry.

The remainder of the talk

- Provide an empirical illustration of the possible outcomes based on a case study of grammatical gender in (some) heritage languages.

> 'In the domain of comparative grammar no subject is of greater interest than the origin of that mysterious grammatical mechanism known as noun gender' (Flom 1903: 1).

Empirical focus: Gender in heritage grammars

- Given the complexity of gender (cf. Corbett 1991), the prediction is that heritage speakers face difficulties with grammatical gender.
  - Polinsky (2008) shows that more proficient speakers of heritage Russian in the US have retained a three-gender system whereas less proficient speakers only have a two-gender system.
  - For the less proficient speakers, Polinsky (1997, 2006) already showed that these do not fully master the complex system of declension classes. Instead, they rely on a formal cue: Whether the noun in its base form ends in a /C/ or a /V/.
  - For American Norwegian, Lohndal & Westergaard (2016) show that the three-gender system is changing: General overgeneralization of the masculine, i.e., general erosion of the system (leading to eventual loss).

Change in functional structure

- Using an agreement attraction paradigm, Scontras et al. (2018) argue that the functional sequence is restructured in heritage Spanish.
- Native speakers: separate [NUM] and [GEN] features:

(2) $[_{DP}\ D\ [_{NumP}\ Num_{[NUM]}\ [_{GenP}\ Gen_{[GEN]}\ [_{nP}\ n\ ]]]]$

- Heritage speakers: one projection with two features:

(3) $[_{DP}\ D\ [_{NumP}\ Num_{[GEN,NUM]}\ [_{nP}\ n\ ]]]$

- '... feature opacity may lead to interpretive instability' and 'the feature bundle might lose feature specification altogether' (Scontras et al. 2018: 21).

This talk: A changing feature system

- The functional sequence is intact but the feature system is changing
- A change in the feature system is not just a lexical property, but rather, a structural one.
- Discuss language mixing in American Norwegian illustrating gender assignment to English nouns.

(Traditional) gender system in Norwegian (M/F/N)

	Masculine	Feminine	Neuter
Indefinite	**en** hest *a horse*	**ei** seng *a bed*	**et** hus *a house*
Definite	hest**en** *horse.DEF*	seng**a** *bed.DEF*	hus**et** *house.DEF*
Double definite (dem. or modified)	**den** hest**en** *that horse.DEF*	**den** seng**a** *that bed.DEF*	**det** hus**et** *that house.DEF*
Adjectives	en **fin** hest *a nice horse*	ei **fin** seng *a nice bed*	et **fint** hus *a nice house*
Possessives	**min** hest *my horse.DEF*	**mi** seng *my bed.DEF*	**mitt** hus *my house.DEF*

Gender on English words in American Norwegian

- Haugen (1953: 44): "All nouns become masculine unless they were associated with a homophonous fem[inine] or neut[er] morpheme or a female creature".
- Hjelde (1996): It is possible to identify morphological, semantic, and phonological rules for English nouns. M is dominant.
- We have data going back to Flom (1903).
  - A lot of variability for all three genders
  - An increase in alternating gender
- Riksem (2018) based on the Corpus of American Nordic Speech (CANS; Johannessen 2015):
  - 1265 nouns, 1034 occur in Norwegian structures
  - 66.1% M, 6.5% F, 6.2% N, 21.2% alternating.

Data from Riksem (2018)

- Indefinites

(4) a. ei **nurse** (F) (coon_valley_WI_02gm)
b. et **shed** (N) (coon_valley_WI_02gm)
c. en **chainsaw** (M) (blair_WI_07gm)

- Definites

(5) a. **field**-a (F) (coon_valley_WI_02gm)
b. **shed**-et (N) (westby_WI_06gm)
c. **chopper**-en (M) (blair_WI_01gm)

Overlapping gender? (Riksem 2018)

- Many nouns do not have the same gender in American Norwegian as in Norwegian. 61.2% of the nouns have a gender that corresponds to the Norwegian equivalent.
- Divergent gender = overgeneralization to M? Only 62.8%, the rest display different patterns.
- In many cases, it is not clear what the translational equivalents would be. Different translational equivalents also have different grammatical gender (e.g., *government* and *field* can be translated into different words that represent all three genders).

New findings: Variation in assignment

- Inter-individual and some intra-individual variation:

(6) a. en dialect (flom_MN_01gm)
b. et dialect (harmony_MN_01gk)

(7) a. ei family (harmony_MN_02gk)
b. en family (gary_MN_01gm; vancouver_WA_01gm)

(8) a. et shed (coon_valley_WI_02gm)
b. ei shed (westby_WI_05gm)

(9) a. ei store (wanamingo_MN_04gk)
b. en store (stillwater_MN_01gm; westby_WI_03gk)
c. et store (westby_WI_03gk)

(10) a. ei trip (billings_MT_01gm)
b. det trip (vancouver_WA_01gm)

Fluctuations: Changes compared to CANS

- Flom (1903)
  - *shed*: occasionally N, more often M [CANS: N,F]
  - *store*: neuter [CANS: F,M,N]
  - *trip*: masculine [CANS: F,N]
- Haugen (1953: ch. 20):
  - *bluff*: ('steep hill'): can occur in all three genders [CANS: F,M]
  - *shed*: occurs both as M and N depending on geography [CANS: N, F]
  - The other nouns are not included in Haugen's word list (he only included about 300 nouns, with each noun occurring at least 15 times or otherwise having some special feature or interest).

Interim summary

- M is the dominant and default gender for English nouns
- Assignment of F and N seems to be fairly random, depending on perceived phonetic similarity, and associations
  - "There is no reason to suppose that his subconscious should have whispered the gender of the native 'equivalent' to him when it failed to deliver the equivalent itself. But in adapting the loanword to its new context, he might easily be reminded of some native word of similar phonetic form which he could follow without qualms" (Haugen 1953: 449).
- Presumably variation in the input due to dialect variation.
- Metalinguistic awareness (of three genders in Norwegian), and 'language anxiety' (Eide & Hjelde 2015, Rødvand 2017)

Ingredients of a formal analysis

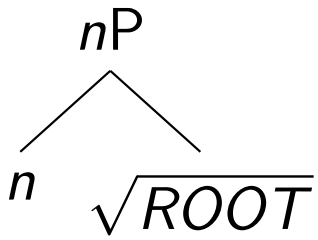
The analysis we advance here should account for the following:

- Speakers have a fixed indefinite article for the Norwegian nouns (and that the suffixed article is generally target consistent)
- English nouns can be assigned grammatical gender (through Norwegian indefinite articles) and that all three genders are represented (to some degree)
- Some speakers appear to assign different genders to the very same root

Where's gender? - The structure of 'nouns'

- Ever since Picallo (1991) and Ritter (1993), an important research question has been to determine the structural locus of grammatical gender.
- Kramer (2016) reviews the issue and argues in favor of gender being part of the noun stem, specifically on the categorizer that provides category to a root.

(11)



Different *n*-s

- Following Lohndal & Westergaard's (in press) implementation of Kramer's (2015) approach:

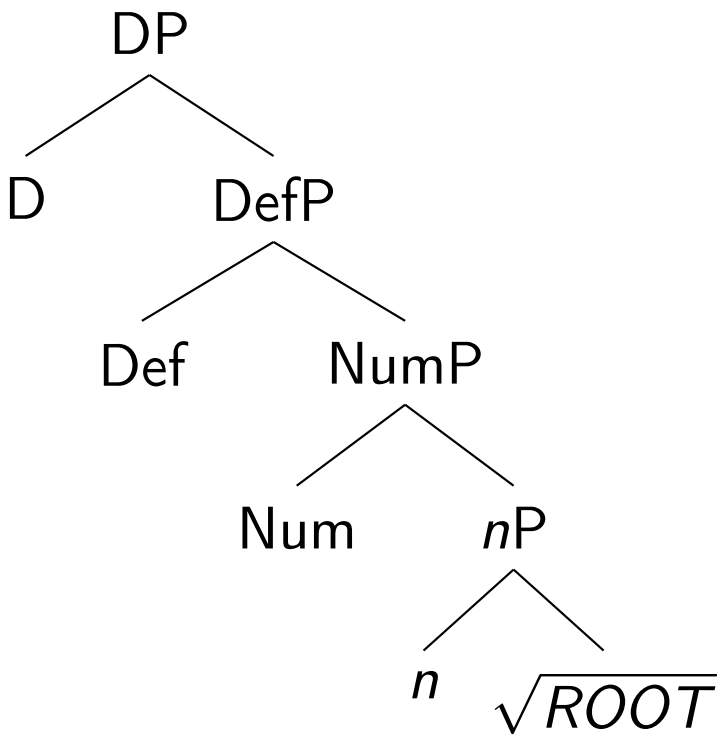
(12) *n* [_] (*neuter*)

(13) *n* [GEN: masc] (*masculine*)

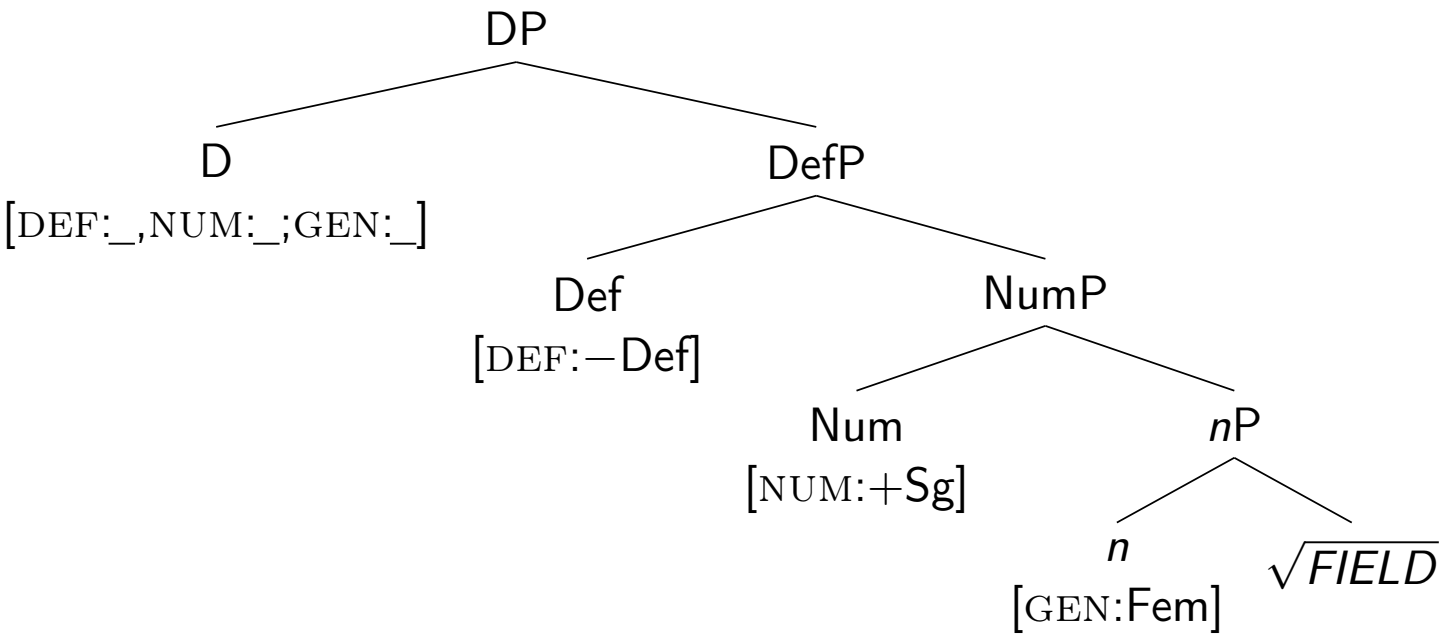
(14) *n* [GEN: fem] (*feminine*)

- Defaults: Neuter for agreement, masculine for assignment.
- The *n* agrees with the D-head and any other head along the spine, accounting for nominal concord

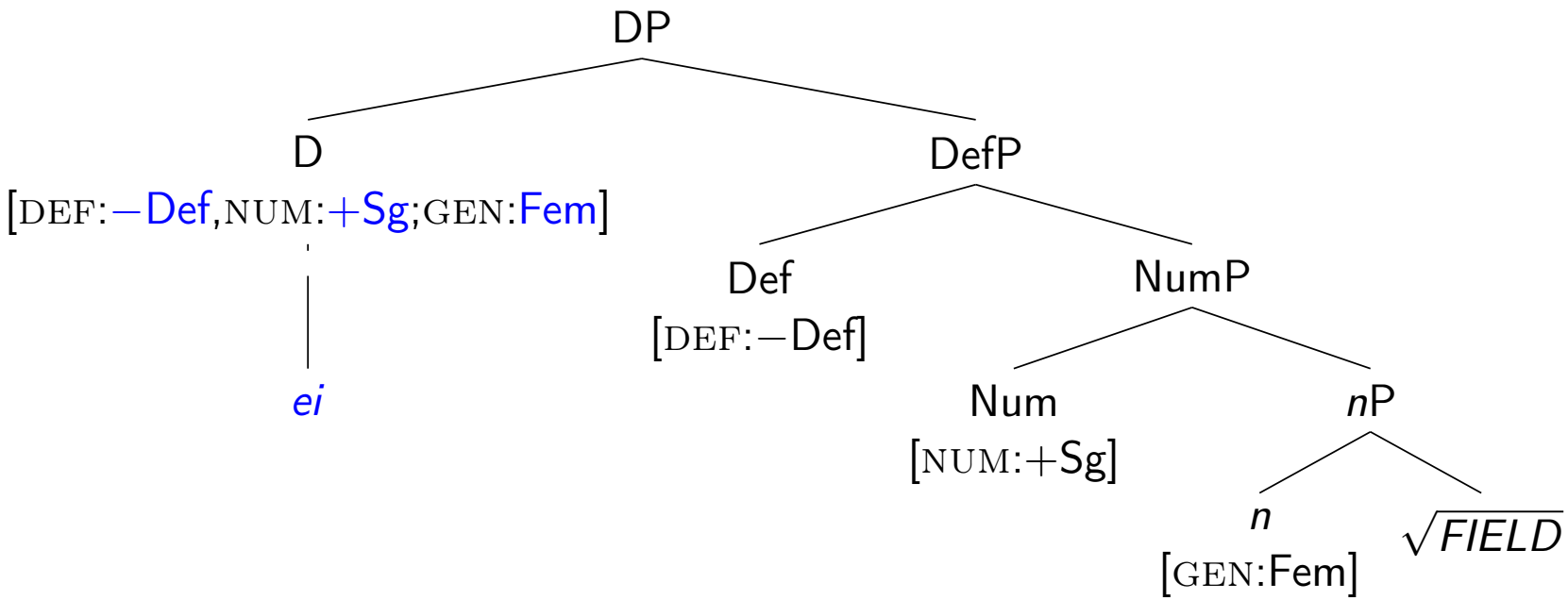
Additional functional projections in DP



DP in AmNo: *ei* ***field*** 'a field'



DP in AmNo: *ei* ***field*** 'a field'



Licensing gender and complexity

- Flavors of licensing conditions (Kramer 2015: 56)
  - Conditions on interpretation at the Encyclopedia
  - Conditions on realization at PF
- The choice of *n* with Norwegian roots is a hard condition; *n* [GEN: Masc] and *n* [GEN: Fem] have a list of possible roots that they license
- The choice of *n* with English roots is a soft condition, subject to Haugen's 'rule' → variability.
- For English roots, M really is dominant, suggesting the elimination of gender features on *n*.

Back to our typology of possible outcomes

- Scontras et al. (2018) show that the functional sequence may be restructured in heritage speakers.
- Our data show that features can be restructured without a corresponding change to the functional sequence itself.
- Complexity can both decrease and increase:
  - Decrease:
    - Fewer features to be acquired and used
    - Simpler mapping rules between gender features and gender exponents
  - Increase:
    - The relationship between gender and declension class has become less transparent.

Towards prediction

- Our talk suggests that a first step is to distinguish between underlying features and their exponents. Based on that we have the following criteria:
  - Number of synsem features
  - Number of functional projections
  - Mapping from synsem features to exponents (One-Form-One-Mapping mappings are simpler)
- Work on heritage languages provides us with important generalizations in terms of which domains of grammar that can restructure and how they may do so.

	Feature	**Functional sequence**	**Exponent**
Retained	all genders	D-Def-Num$_{[Num]}$-$n_{[Gen]}$	1:1
Lost	fewer genders	D-Def-Num$_{[Gen+Num]}$-n	x:y

Summary & remaining challenges

- Decomposing complexity into features and functional sequence provides a first pass at modeling complexity.
- The model needs to be tested on a range of other phenomena and typologically unrelated languages, including possible innovations.
- More generally, the approach here provides additional evidence for an exoskeletal architecture of grammar, whereby syntax and its exponents are fundamentally separated.
- More work needs to be done:
  - The relationship between complexity and other notions such as transparency and vulnerability
  - How this relationship should be formalized.

Complexity is a complex issue...

Thanks!

Exoskeletal grammars: Core principles & desiderata

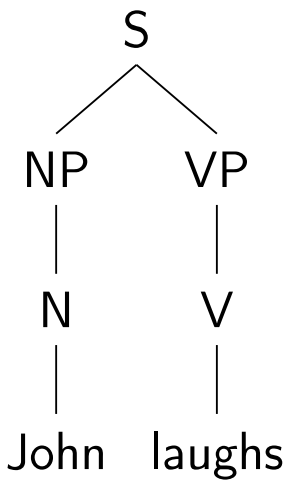
- Many formal grammars are **endoskeletal**: Grammatical structures are projected on the basis of properties of lexical items ('words'). Example: a transitive verb projects a structure with two argument positions.
- However, grammars can also be **exoskeletal**, where structures are generated independently of lexical items.
- Chomsky's early work (1955, 1957) was actually exoskeletal: Phrase structure (PS) rules generated structures and lexical items were inserted after the structure had been built.

An early exoskeletal approach

Assume a grammar with the following PS rules:

(15)
a. $S \rightarrow NP\ VP$
b. $NP \rightarrow N$
c. $VP \rightarrow V$
d. $N \rightarrow$ John
e. $V \rightarrow$ laughs

Running through the derivation provides a familiar tree:



Core evidence: Argument structure

- A lot of evidence has been mustered from the area of argument structure, in particular involving argument structure flexibility.

(16)
a. Kim whistled.
b. Kim whistled at the dog.
c. Kim whistled a tune.
d. Kim whistled a warning.
e. Kim whistled me a warning.
f. Kim whistled her appreciation.
(Rappaport Hovav & Levin 1998)

What to make of this flexibility?

- Two options:
  - Admit polysemy and accept that there are e.g., six different versions of *whistle* in the lexicon
  - Argue that this shows that verbs do not have inherent argument structure that is projected into the syntax
- This requires a handle on restrictions on argument structure (e.g., **Lisa kicked*, **Kathy gave Bill*), and various proposals exist (Borer 2005, Lohndal 2014, among many).