Noun classes in the North-East Caucasus

Maria Polinsky
polinsky@umd.edu

Language at Maryland

Outline

- Noun class as a grammatical concept
- Noun classification in the NE Caucasus
- Noun class assignment (with a touch of acquisition)
- Noun class agreement in the general model of syntactic agreement

The concept of noun class

- Many languages classify nouns by virtue of assigning them to one of several groups (Corbett 1991)
- Such groups are known as noun classes or grammatical genders
- Noun class is a broader term than grammatical gender:

  Noun class ⊃ grammatical gender

The concept of noun class

- Many languages classify nouns by virtue of assigning them to one of several groups (Corbett 1991)
- Such groups are known as noun classes or grammatical genders
- Division of nouns into groups manifested in the behavior of associated words
- Behavior of associated words:
  - Agreement and concord with a given noun (verbs, conjunctions, determiners, adverbs, auxiliaries, and modifiers)
  - Pronoun corresponding to a given noun (coreferential pronouns outside the nominal domain)

Noun class vs classifier

- Noun class (grammatical gender):
  - (relatively) small number of classes
  - nouns generally belong to only one class
  - obligatory copying of noun-class features on associated words (agreement)
- Classifier:
  - large number of classes (often 20+)
  - same noun may appear with different classifiers
  - classifier only used in certain syntactic or pragmatic contexts

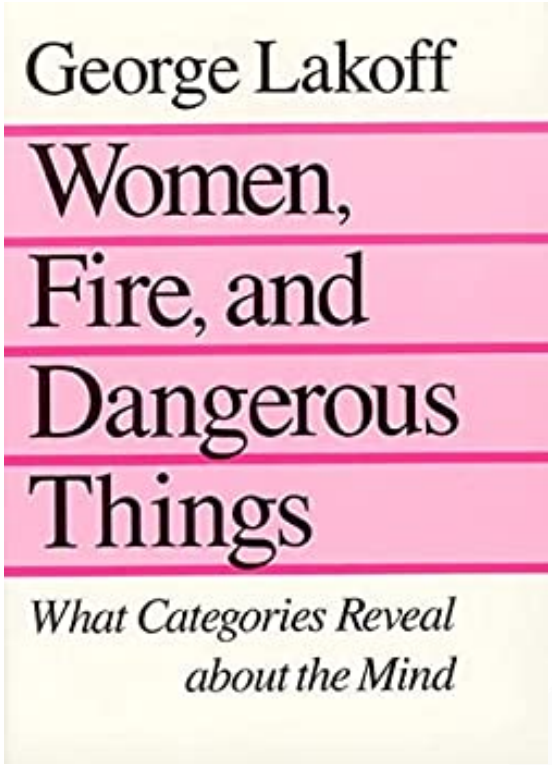
Assignment vs agreement

Two main facets of noun class as category:

- Noun class assignment: how does a language divide its nominals into classes?
- Noun class agreement: how does the noun-class/gender feature of a particular noun gets copied onto associated words?

Assignment: Myths and reality

- Noun class composition may be quite complex, hence often explained by complicated semantic rules (e.g., “women, fire, and dangerous things”)



Assignment: Myths and reality

- Noun class composition may be quite complex, hence often explained by complicated semantic rules

- Children learn the bulk of gender assignment by 36 months (individual, more complex cases linger as errors up to age 7), which makes the use of complex semantic rules highly questionable

Assignment: Reality

- Simple semantic core, recurs across the world's languages
  - Features children are sensitive to early on (natural gender, animacy)
  - Several smaller categories that do not require abstract connections or cultural knowledge young learners may not have (edible items, mobile entities)
- Highly salient formal features
  - Features of the sort accessible to young learners (initial or final segments, salient derivational suffixes)
- Children and adults may not use the exact same rules however (stay tuned)

Outline

- Noun class as a grammatical concept ✔
- Noun classification in the NE Caucasus
- Noun class assignment (with a touch of acquisition)
- Noun class agreement in the general model of syntactic agreement

NC across the Caucasus

- Nakh-Daghestanian:
  - Most languages have noun classes (genders)
  - Noun classes reconstructed for Proto-ND
  - After-effects of gender in some languages (e.g. Harris 2004 on loss of gender in Udi without loss of declension classes)
- Northwest Caucasian:
  - Only Abkhaz and Abaza: (masculine, feminine, non-human)
- Kartvelian:
  - No gender

Noun classes in languages of the NE Caucasus

- No noun classes: Udi, Lezgian, Agul, and several southern Tabasaran dialects
- Otherwise, two to eight noun classes

Q: How do we see these noun classes?

A: through the behavior of associated words

Noun class agreement in ND

- Agreement markers on verbs (and in some languages also on adjectives, demonstratives, adverbs, postpositions, particles)
  - Generally prefixal, but some infixal agreement (e.g. Lak)
  - Gender markers typically appear only on vowel-initial words
  - Agreement is with the absolutive argument (S/O)

Lak, infixal agreement on verbs

t:ul	dus-na-l	t:ul	lu	la-w-s-unni
1SG.OBL.GEN	friend-OBL-GEN	1SG.OBL.GEN	book. ABS	take-III-take -TR.PRF.3SG

‘My friend took my book.’ (Friedman 2020)

Noun class agreement in ND

- Agreement markers on verbs (and in some languages also on adjectives, demonstratives, adverbs, postpositions, particles)
  - For some aspects of the syntax of this agreement, see Polinsky (2016), Polinsky et al. (2017), Foley (2020)

Noun class agreement in ND

Archi demonstratives (Polinsky et al. 2017)

I SG	II SG	III SG	IVSG	PL	GLOSS
ju-w	ja-r	ja-b	ja-t	j-eb	this, close to the speaker
jamu	jamu-r	jamu-m	jamu-t	jem-im	this, close to the hearer
to-w	to-r	to-b	to-t	t-eb	that, further away from the speaker
gud-u	god-or	god-ob	god-ot	gid-ib	that, lower than the speaker
ʁud-u	ʁod-or	ʁod-ob	ʁod-ot	ʁid-ib	that, higher than the speaker

Noun class agreement in ND

- Agreeing verbs: 20%-30%of the verbal lexicon but about 75% of verb tokens; most common verbs (‘be’, ‘do’, ‘know’, ‘see’) show agreement
- Agreeing adjectives: about 10%-15% of the adjectival lexicon but about 70% of adjective tokens; most common adjectives (‘good’, ‘many’) show agreement
- Agreeing demonstratives: the majority of items
- The presence of agreement on very frequent items supports the maintenance of noun classes

Noun class indexing on associated pronouns

Tsez reflexives/reciprocals (compound forms, simplified, see Polinsky 2015 for details)

	Class I (males)	Classes II-IV (non-males)
reflexive	*nesä ža*	*neɬä ža*
reciprocal	*žedä žedi*	*žedä žedu*

Exotic?

- In cocktail-party discussions about noun classes in ND, people typically talk about:
  - Large number of noun classes
  - Complex assignment of nouns to classes

Exotic? Not so much

- In cocktail-party discussions about noun classes in ND, people typically talk about:
  - Large number of noun classes
  - Complex assignment of nouns to classes

- Let's see how it works

Number of noun classes: from 2 to 8

- Tabasaran (northern): 2 (human vs. non-human)

Number of noun classes: from 2 to 8

- Tabasaran, Atrik dialect (Bogomolova 2018)

	‘throw’		‘drive away’		‘throw, toss’	
	H	NH	H	NH	H	NH
SG	gat’ax-	gat’avx-	ut’uk-	u’uvk-	it-	üp-
PL		gat’ax-		ut’urk-	irč-	ürč-

The division of nouns into classes intersects with the division of denotations into singular and plural; [GENDER] and [NUMBER] operate in a shared space.

Number of noun classes: from 2 to 8

- Tabasaran (northern): 2 (human vs. non-human)
- Avar, Dargwa, and most Andic languages: 3 (male rational, female rational, non-human)
- Lak, Tsez, Hinukh, the Lezgic languages with noun classes: 4
  - Lak: male rational; (mature) female rational; animate; inanimate
  - Archi: male rational; female rational; complex division for remaining nouns
  - Tsez: male rational; female rational + inanimates; animates & inanimates; inanimates

- Chamalal, Hunzib, Khwarshi: 5
  - Hunzib: male rational; female rational; animates and inanimates spread across other three classes (Forker 2014)
- Chechen, Ingush: (traditionally) 6
- Batsbi (Tsova-Tush): (traditionally) 8

Why “traditionally”?

Batsbi (Tsova-Tush)

- Traditionally analyzed as having 8 classes

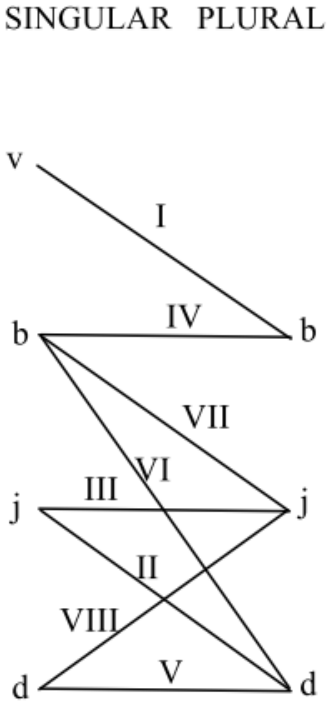


Figure 4: Agreement classes in Batsbi

(Corbett 2014)

- But, classes VI – VIII contain only ~20 words among them
- None of these three classes is independent
- "Inquorate" classes, easier to interpret them as lexical exceptions
  - new members cannot (easily) be added
  - have very few members

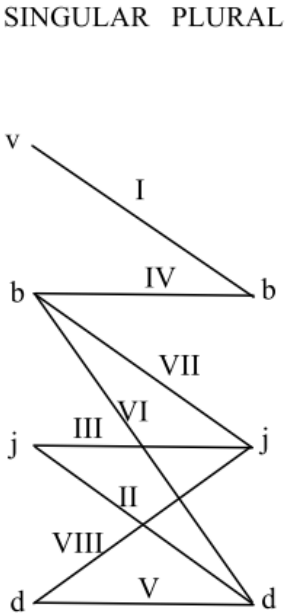


Figure 4: Agreement classes in Batsbi

Batsbi (Tsova-Tush) noun classes

	Singular	Plural	
Class I:	v-	b-	mostly denote males: *stak'* 'man' , *dad* 'father', *mar* 'husband'
Class II:	y-	d-	mostly denote females: *nan* 'mother' *pst'u* 'wife', *joħ* 'daughter'
Class III:	d-	d-	largest class: *bader* 'child' , *dok'* 'heart'
Class IV:	y-	y-	2nd largest class: *cark'* 'tooth' , *q'ar* 'rain'
Class V:	b-	d-	*pħu* 'dog', *ča* 'bear' , *matx* 'sun'
Class VI:	b-	y-	*bak* 'fist', *bʕark'* 'eye', *kok'* 'leg'
Class VII:	d-	y-	*bat'r* 'lip', *lark'* 'ear', *t'ot'* 'hand'
Class VIII:	b-	b-	*borag* 'knit slipper', *kakam* 'wool cut in fall'

(Wier 2014)

- So, under this approach Batsbi has only 5 classes

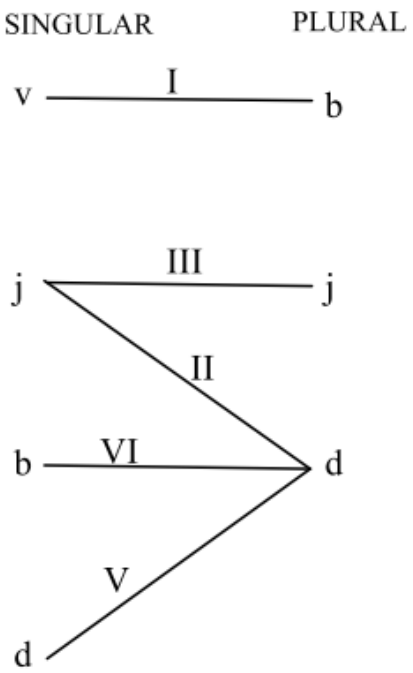


Figure 5: Gender in Batsbi (excluding inquorate genders)

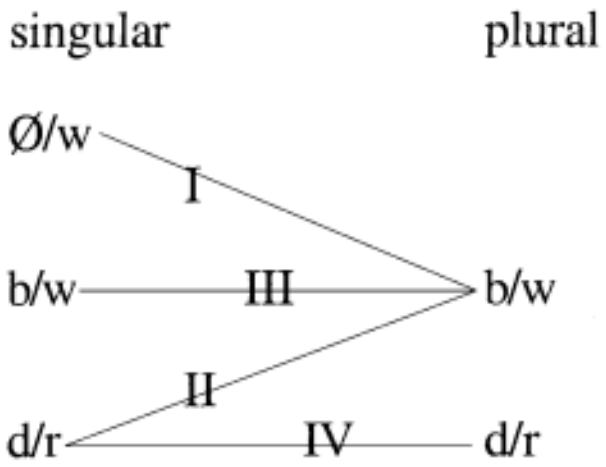
- If we do not count inquorate classes, the maximum number of noun classes found in ND drops to 5
- Always fewer class distinctions in the plural
  - Commonly, human vs. non-human
    - Godoberi (3 sg classes: male, female, non-human)
    - Hinukh (5 sg classes: male, female, animals, inanimate, inanimate)
  - Some languages have 3 classes in the plural
    - Bezhta (Tliadal dialect): 5 sg classes (male, female, animals & inanimates, inanimate, inanimate) and 3 plural classes (I & II, III & IV, V)

How many classes in Lak?

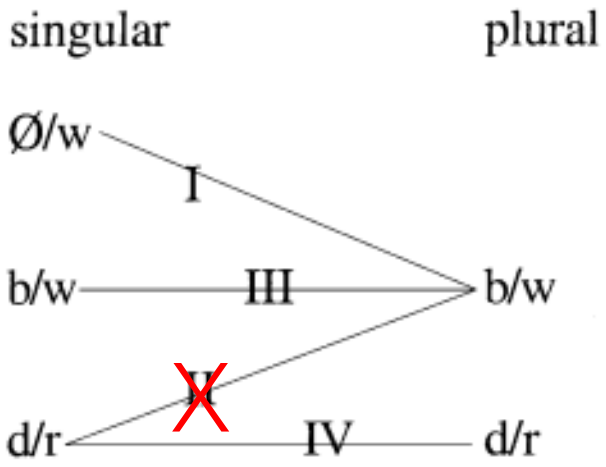
- Traditionally analyzed as having 4 classes
  - male rational; (mature) female rational; animate; inanimate
  - Historically, started with typical class II (female)
  - The vast majority of original class II (female) nouns shifted to III (animate) by attraction to *dus* 'girl' (III); impolite to use II
  - Now only ~20 words in class II, mostly terms for older family members (e.g. 'mother', 'grandmother', 'aunt', 'wife')
  - Similar shift in some southern Polish dialects and dialects transitional to Czech and Slovak: words referring to girls and unmarried women are neuter (Corbett 1991 citing Zaręba 1984-5)
    - Zuzię poszlo. 'Zuzia has gone.'
    - Jo bylo na grziby. 'I was mushrooming.'

Žirkov (1955): proposed a 5th class (inquorate)

qqatta 'house': III in sg., IV in pl.



- Friedman (1996, 2020): argues for 3 classes
  - Class II contains a small set of nouns (~20)
  - The class markers are not unique to that class
  - So, class II is inquorate



Proto-ND genders: Reconstruction

- Nakh-Daghestanian genders (Nichols):
  - v/Ø Male human
  - j/r Female human
  - b Many animates
  - d/r Inanimates (chiefly)
  - j Various non-human

Noun classes in languages of the NE Caucasus: Basic semantic core

- Animate/inanimate
  - Male/female
- Human/non-human
  - Male/female
- Grown or rational/non-rational
- Mobile/non-mobile

Aside from natural gender, all the oppositions are privative (only one member of the opposition is specified),

which seems to be the basic type of underlying representations in grammar

Outstanding questions

- Cognitive underpinnings of inquorate noun classes: what can motivate memorization?
- Main ingredients of the answer seem to include the number of lexical exceptions and their frequency
  - Too few, or too many and infrequent: loss of the inquorate category
  - Few and frequent: the inquorate survives
- We need acquisition data to know what happens in the learning of inquorates

Outline

- Noun class as a grammatical concept ✔
- Noun classification in the NE Caucasus ✔
- Noun class assignment (with a touch of acquisition)
- Noun class agreement in the general model of syntactic agreement

Main assumptions

- Noun class/Gender must be learnable by child acquirers, who do not have access to complicated semantics
- Classification should be able to be explained by appealing to simple semantic and formal features of the sort children are sensitive to
- Formal features have been shown to be relevant in a variety of languages including Russian (Corbett 1991), French (Tucker et al. 1997), Romanian (Bateman & Polinsky 2003), German (Tanenbaum 2003), Dyirbal (Plaster & Polinsky 2007, 2012), and Tsez (Gagliardi 2012, Plaster et al. 2013)

Classification in Tsez
(Plaster et al. 2013)

- Tsez is known to have a complicated assignment system, typical of the ND languages, so we set out to analyze it without appealing to complex semantics

Tsez noun classes

Tsez noun class agreement prefixes

Class	Content	Singular	Plural
I	Males (human and divine)	Ø-	b-
II	Females (human and divine) and various inanimates	y-	r-
III	Animals and various inanimates	b-	
IV	Other inanimates	r-	

Paradigm of –igu ‘good’

Class	Singular		Plural	
I	**Ø-igu** I.AGRsg-good ‘good shepherd’	aħo shepherd	**b-igu** I.AGRpl-good ‘good shepherds’	aħo-bi shepherd-PL
II	**y-igu** II.AGRsg-good ‘good wife’	baru wife	**r-igu** II-IV.AGRpl-good	baru-bi wife-PL ‘good wives’
III	**b-igu** III.AGRsg-good ‘good donkey’	ʕomoy donkey		ʕomoy-bi donkey-PL ‘good donkeys’
IV	**r-igu** IV.AGRsg-good ‘good spindle’	ʕoʎ’ spindle		ʕoʎ’-mabi spindle-PL ‘good spindles’

Class assignment

- Human and divine males > class I
  - *uži* ‘boy’
  - *allah* ‘Allah’
  - *zhek’u* ‘man’
- Human and divine females > class II
  - *baru* ‘wife’
  - *echju* ‘grandmother’
- All other animates > class III
  - *ʕomoy* ‘donkey’
  - *aw* ‘mouse’
  - *šajtan* ‘devil’
- Inanimate nouns > classes II, III, IV

- Inanimates in II: berries; paper items (letter, dictionary, newspaper); some clothing (not exclusively female); some body parts (knee, chin, shoulder blade, leg); some tools (hammer, plough, shovel); mountains, stones and rocks; some time terms (year, seasons); and various other inanimates (cage, drinking glass, salt, motor, dust, mill, science, etc.).
- Inanimates in III: some clothing; some body parts (finger, calf, arm, heel, rib); some tools (hoe, chisel, sickle, tool); some time terms (month names, minute); vehicles; many Arabic loanwords[3]; and various other inanimates (alphabet, field, call, proverb, gasoline, sun, moon, etc.).
- Inanimates in IV: derived abstract nouns in *–ƛi* or *–ni*; some clothing; some body parts (wrist, knuckles, belly, shoulder); some time terms (day names, day); and various other inanimates (wine glass, crib, university, navy beans, rye, stick, milk, etc.).

Basics of Tsez noun classes

I:	males
II:	females + [class II inanimates]
III:	other animates + [class III inanimates]
IV:	[class IV inanimates]

- But how are inanimates assigned to II, III or IV?

Possible approaches

- Approach 1: Class assignments are simply memorized
  - Large task, especially when information about class isn't robust
  - Other 'arbitrary' gender languages have been shown to be predictable (e.g. Tucker et al. 1977, Lyster 2006 for French, Harris 1991 for Spanish, Tanenbaum 2003 for German)
  - Memorization would not explain cross-speaker consistency in assignment of nonce forms

- Approach 2: Tsez is like Dyirbal (“women, fire, and dangerous things”)
  - Relies on the notion of “radial categories”: each class contains a prototype, or member that contains most of the defining characteristics of the class
  - Other nouns are included in the category based on their perceived resemblance to the prototype and language-specific principles
  - Nouns can be linked through other members
  - Under this approach, speakers learn the core members and assign other nouns accordingly

(7) Principles for assignment of noun classes in Tsez (Rajabov 1997)

(a) Material: if X is the material out of which Y is made, Y may be assigned to the same class as X (e.g. 'wood' and 'chair')

(b) Shape: flat items tend to go into class II; round, non-flat things tend to go into class III; long, thin items tend to go into class IV

(c) Internal feature: liquidness and density sometimes are relevant to class assignment ('ice' is in class III because of its association with 'rock', but it could be expected to be in class IV under the 'material' principle)

(d) Function: if Y is used for or resembles X functionally, Y may be assigned to the same class as X (e.g. 'fortress' is in class III because 'fight' is in class III)

(e) Semantic domain association: the assignment of nouns may create semantic domains (e.g. 'sock' is assigned to class IV on the basis of 'wool', and a semantic category of 'footwear' is subsequently created in class IV on the basis of the assignment of 'sock')

(f) Species to genus association: nouns referring to specific instances of more general nouns will be assigned to the class of the more general noun (e.g. the words for different fingers are assigned to the same class as 'finger')

(g) Concept association (analogy): loanwords that duplicate existing words may be assigned the class of the duplicated words

(h) Opposites: words expressing opposite concepts are placed in the same class (e.g. 'fire' and 'water' are in class IV, 'medicine' and 'poison' are in class III)[4]

- These sound more like ‘after-the-fact’ generalizations rather than operating principles
- Rajabov identifies these as ‘tendencies’ (‘may’, ‘sometimes’)
- Does not motivate the links between members and classes in an unambiguous or predictive way

- Approach 3: Tsez speakers rely on a combination of semantic and formal features to classify nouns
- Groundwork laid by:
  - Comrie & Polinsky 1999: identified connection between *i/y* and class II
  - Polinsky & Jackson 1999: identified class II as resulting from the merger of two earlier classes and performed nonce testing
    - Nonce forms beginning with i/y > 92% assigned to class II
    - Nonce forms ending in i/y > 78.5% assigned to class II
    - Initial or final bilabial or r > class III or IV, respectively, but not as robustly
- We sought to expand this approach to Tsez through computational modeling of a larger-scale, systematic analysis of Tsez nouns

Decision-tree modeling

- Decision trees: a series of connected questions, beginning with a single question (node) and resulting in decisions based on the answers given
- Induced from a data set with specified attributes
- Want to find the smallest decision tree consistent with the data, so we should ask the most determinative questions first
- (Decision trees are not the only modeling option, but we think they are a useful way to visualize the system.)

Toy example: Spanish gender

- How are Spanish genders assigned?

Spanish example

Attributes

Goal

Noun	Female?	final a?	Gender?
actriz ‘actress’	Y	N	F
chica ‘girl’	Y	Y	F
jardin ‘garden’	N	N	M
pijama ‘pajamas’	N	Y	M

If we examine the values of the two attributes, we find that the values of each attribute split the examples into the two subsets shown in (9) and (10):

(9) 'female?':

subset 1 (=Y): actriz (F), chica (F)

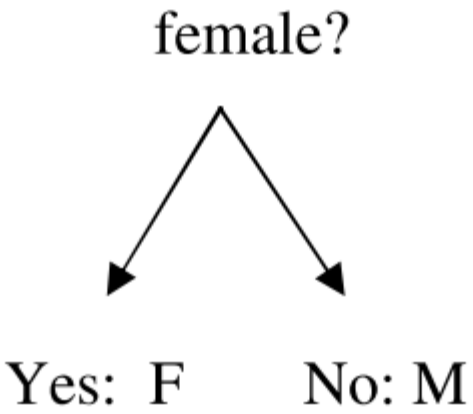
subset 2 (=N): jardin (M), pijama (M)

(10) 'final a?':

subset 1 (=Y): chica (F), pijama (M)

subset 2 (=N): actriz (F), jardin (M)

- Based on the sample, it is more informative to ask whether the noun is [female] than whether it ends in *a*
- A simple decision tree accounts for the data



Spanish example

Noun	Female?	final a?	Gender?
actriz ‘actress’	Y	N	F
chica ‘girl’	Y	Y	F
jardin ‘garden’	N	N	M
pijama ‘pajamas’	N	Y	M

- Running each noun through the tree produces the correct gender assignment

- But, Spanish is not so simple (see Harris 1991 for a full discussion)

Additional Spanish feminine nouns

Noun	Female?	Final a?	Gender?
casa 'house'	N	Y	F
leche 'milk'	N	N	F
mano 'hand'	N	N	F

- Some F nouns are not [female]; > M based on the tree
- Some F nouns end in *e* or *o* (*leche* 'milk', *mano* 'hand')

Pitfalls

- A non-comprehensive data set may not accurately represent the situation
  - Based on the forms we have seen, the final *a* is not very predictive of F, when it is, in fact.
- The choice of attributes tested is critical
  - We've tested [female] and final *a*, but maybe we also want to test [male] and final *o*? Or other features?
- A decision tree will not account for lexical exceptions
  - E.g. certain M forms end in *a* (*día* 'day', *pijama*)
  - Phonological rule: F words beginning with stressed initial *a* take M article in singular but F agreement (*el agua fresca*, *las aguas frescas*)
  - These must be memorized by speakers

Tsez data set

- Over 3,500 nouns culled from Khalilov 1999 and Rajabov (undated)
- To ensure accuracy of assignments and dialectal consistency, the classification of each noun was confirmed by native speakers of the Kidiro and Mokok dialects

Class	% of total
I	12.6%
II	12.4%
III	41.4%
IV	33.6%

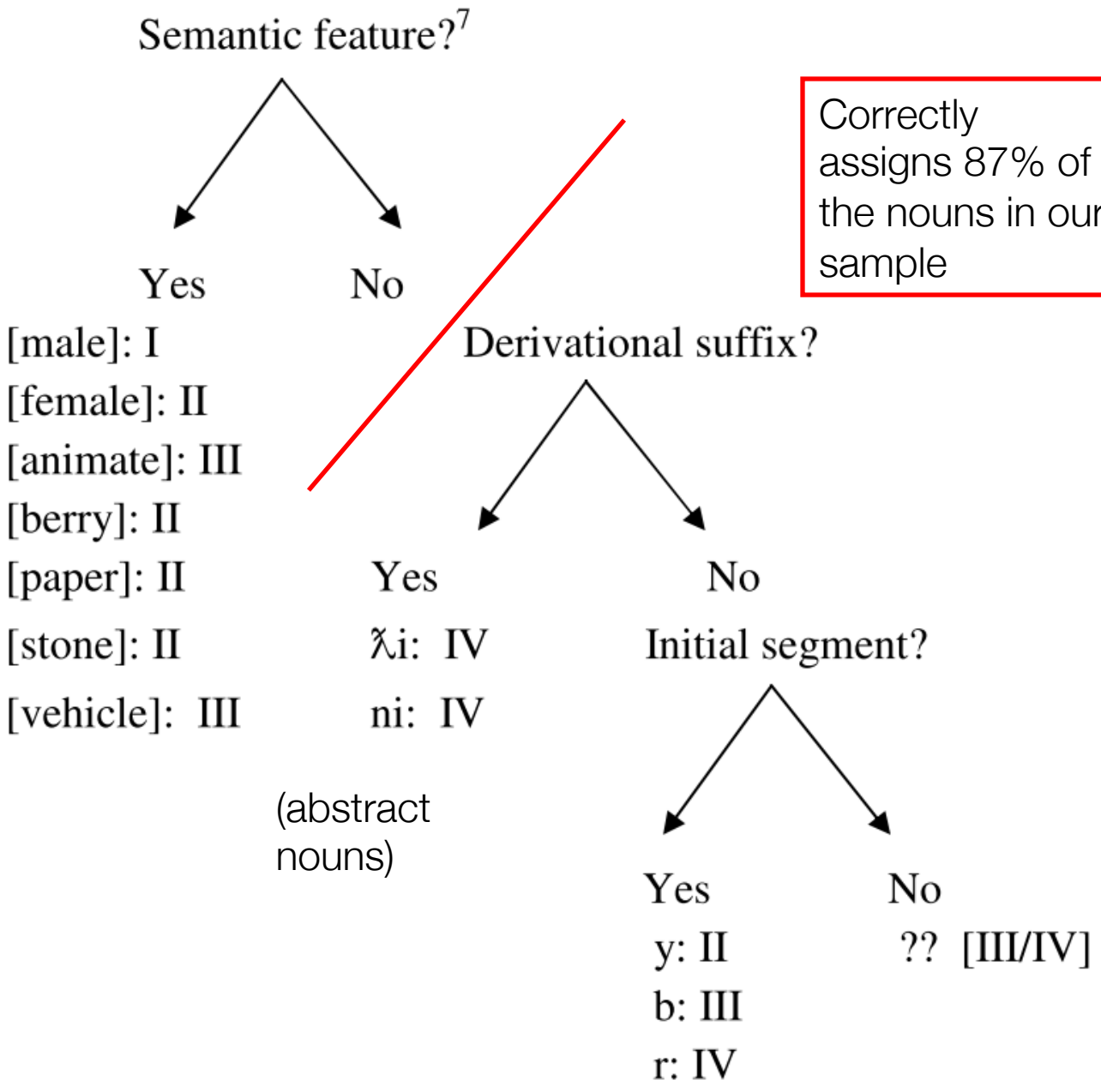
Class I slightly inflated by inclusion of M/F agent nouns

Attribute selection

- Tested a broad set of semantic and formal features of the sort children may be sensitive to:
  - Formal features: first segment, last segment, first 2 segments, last two segments, declension class, number of syllables
  - Semantic features: male, female, animate, berry, paper, edible, vehicle, container, stone
  - Some semantic features are top-down while others are likely bottom-up

Data mining

- We coded each noun for the relevant attributes
- Ran the data set through the "Orange" data mining tool (Demsar et al. 2004), based on Quinlan's C4.5 algorithm (Quinlan 1993)
  - These are simple yet powerful learning algorithms widely used in data mining and machine learning.
  - Examine the data to determine the most predictive splits.
- Using Orange we produced the following decision tree:



Correctly assigns 87% of the nouns in our sample

Semantic features

- Semantic features most predictive of assignment
- Semantic features override conflicting formal features, as is common cross-linguistically (Corbett 1991, Gentner & Namy 1999)
- All males > I, females > II, animals > III
- [vehicle] is strongly predictive of class III, so perhaps class III is [mobile] rather than [animate]; that allows us to merge [animate] and [vehicle] into [mobile]
- The smaller semantic classes are also predictive, but not exceptionless (especially strong noise in child data), so they may be abandoned

Semantic features, revised

- Semantic features most predictive of assignment

[MALE]

[FEMALE]

[MOBILE]

Derivational suffixes

- The two abstract-forming derivational suffixes were strongly associated with class IV
- Abstract nouns appear in classes II and III as well, so it's not the feature [ABSTRACT] causing assignment, but the formal shape of the suffix
  - *gaq ʻu* 'destruction', *kep* 'happiness' – class II
  - *adab* 'politeness, respect', *bax* 'luck' – class III

What about the other 13%?

- The burning question…
- Likely due to several questions and potential confounds
- 1. Could the large number of loanwords in Tsez be skewing the results?
  - Our lexicon has a large number of loanwords, some older (e.g. words of Arabic origin, such as *din* 'religion', *alim* 'scholar, teacher') and some more recent (e.g. the many Russian loans currently used, e.g. *istoli* 'table', *tilipon* 'telephone', *nowutbuk* 'laptop').
  - But these loans are part of daily life; our goal was to model the task of child learners and the competence of adult speakers, rather than to model classification in the native lexicon
  - Tsez could be in an interim stage of development as the generalizations made by speakers change as increasing numbers of loanwords enter the language; studies on acquisition of noun classes provide useful insight

- 2. Dialectal variation may be responsible for some noise
  - There are at least 5 dialects of Tsez, which have phonological and lexical differences and may also have classification differences
  - Rajabov sometimes provides different classifications from those of Khalilov 1999 and our consultants, which may reflect dialectal differences
  - Dialectal variation in noun classes is seen elsewhere in the family
    - Tabasaran: some southern dialects have lost gender
    - Khwarshi: some dialects have 5 classes while others have 4
  - If the system is in flux, dialectal variation may be even more expected

- 3. There may be other attributes involved
  - We've tried to explain as many nouns as possible without appealing to after-the-fact generalizations by restricting our attributes to those child learners have been shown to be sensitive to
  - A continued look at the possible semantic attributes may be helpful

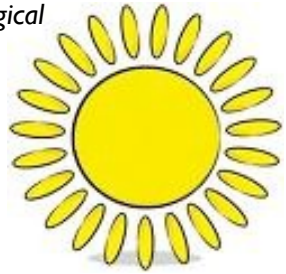
Computational models and behavioral experiments

- Gagliardi (2012): a series of behavioral experiments with children and adults, aimed at classifying real and nonce words by noun class

Classification experiment

kid (*girl)*
Class 2, Semantic Cue

buq (*sun)*
Class 3, Phonological Cue



k'uraj (*onion)*
Class 4, no Cue

zamil (*nonce)*
Class 3, Semantic Cue

Results: Real Words

Percent of words classified correctly

	Biological Semantic	Other Semantic	Phonological	No Cue	Conflicting
Young Children	79	71	84	77	**42***
Older Children	86	58	94	78	**47***
Adults	87	75	92	86	71

Classification of **real words** was compared to the words' actual class

Main takeaways of the classification experiment

- Children used less reliable phonological cues rather than more reliable semantic cues when the two conflicted

- Only certain types of semantic information are used

- Not discussed here: Tsez children used class III as default, and adults used class IV. Thus, children and adults differ in their noun class assignment, and the question remains as to how children develop the adult pattern

Tsez noun class assignment

- Noun classification in Tsez is highly predictable (87% right now)
- Simple semantic core
  - Features children are sensitive to early on (natural gender, animacy, mobility)
  - Several smaller categories that do not require abstract connections or cultural knowledge young learners may not have
- Set of highly salient formal features
  - Again, of the sort accessible to young learners (initial segments and salient derivational suffixes)

Learning about noun class assignment: Some lessons from Tsez

- Modeling approaches are a useful tool for:
  - establishing the division of nouns into classes
  - testing existing conceptions of noun classification
  - (not discussed here) reconstructing gender classifications
- Modeling should be used in conjunction with corpus data and if possible L1 acquisition data (for acquisition of Tsez genders, see Gagliardi 2012, Gagliardi & Lidz 2014)

Outline

- Noun class as a grammatical concept ✔
- Noun classification in the NE Caucasus ✔
- Noun class assignment (with a touch of acquisition) ✔
- Noun class agreement in the general model of syntactic agreement

Main issues in noun class (gender) agreement

- Agreement: covariance in features between a structural head (verb, auxiliary, tense/aspect marker, determiner) and a nominal constituent
- Main issues:
  - The structure of the feature space
  - The expression of the relevant feature on the nominal constituent
  - The mechanism of acquiring the relevant feature from the nominal constituent (where? how? finding the right match)
  - The (non-)uniformity of agreement as a morphological construct

The structure of feature spaces

- Despite the apparent dazzling cross-linguistic variation, agreement features (also known as phi-features) form a very limited set: [PERSON], [NUMBER], [GENDER]
- phi-features are privative; there is no [±feature] in syntax, there is only [PARTICIPANT], and the absence thereof

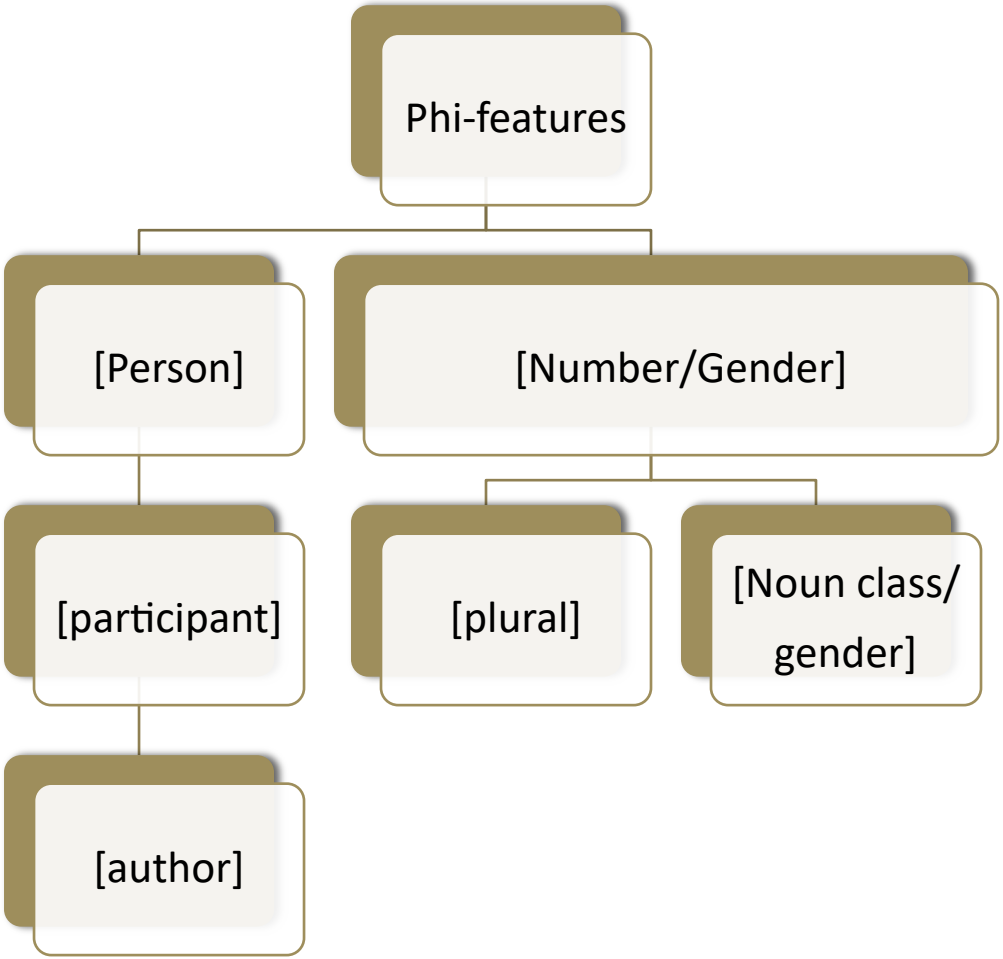
The structure of feature spaces

- Despite the apparent dazzling cross-linguistic variation, agreement features (also known as phi-features) form a very limited set: [PERSON], [NUMBER], [GENDER]
- phi-features are privative; there is no [±feature] in syntax, there is only [PARTICIPANT], and the absence thereof
- there are dependencies among different phi-features: certain privative features cannot be present unless another designated feature is present (e.g., no dual unless there is plural)

The structure of feature spaces

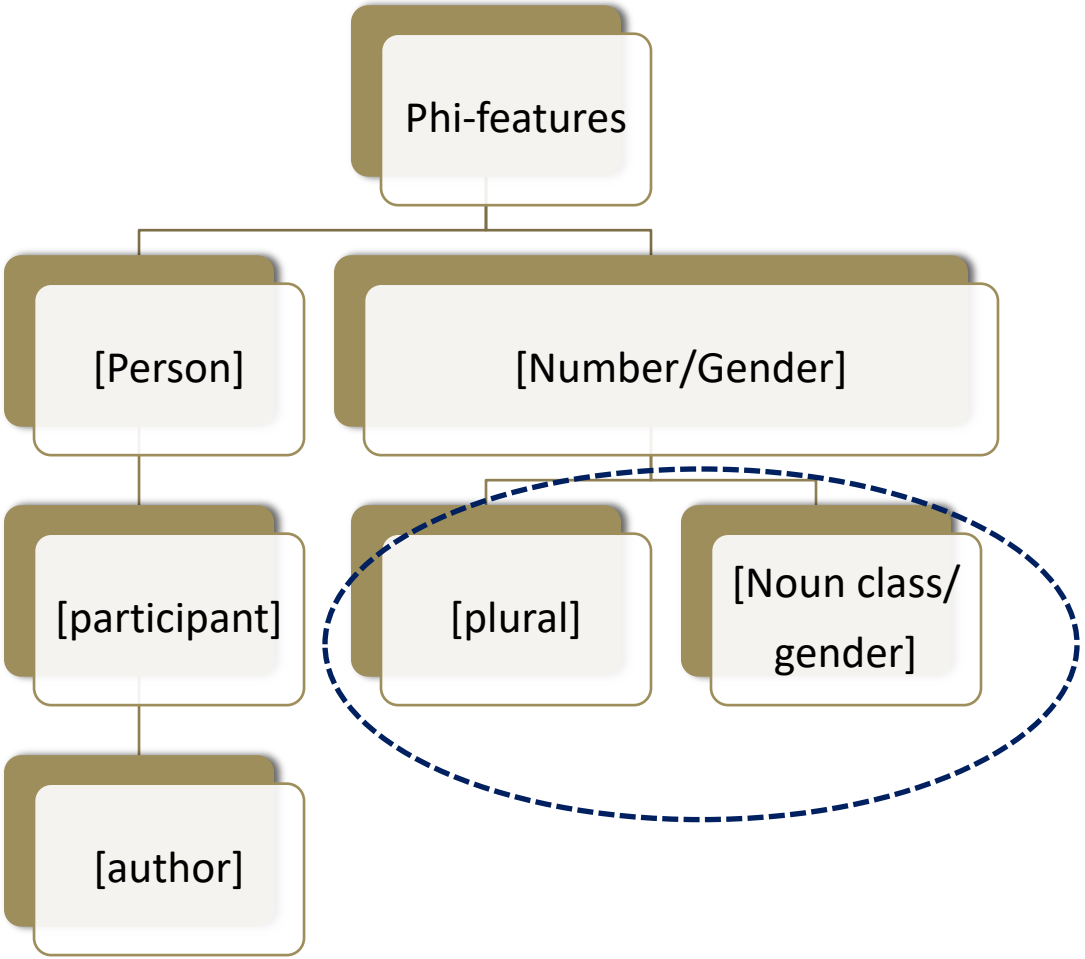
- despite the apparent dazzling cross-linguistic variation, agreement features (also known as phi-features) form a very limited set: [PERSON], [NUMBER], [GENDER]
- phi-features are privative; there is no [±feature] in syntax, there is only [PARTICIPANT], and the absence thereof
- there are dependencies among different phi-features: certain privative features cannot be present unless another designated feature is present (e.g., no dual unless there is plural)
- when agreement indexes only a subset of the phi-features of a given noun phrase, it is typically [PERSON]-agreement that goes missing (Baker 2008, 2011)
- therefore, phi-features are organized in a hierarchical structure

Hierarchies (Harley & Ritter 2002)



The connection between number and gender is accounted for under the hierarchical relationship between phi-features

Hierarchies (Harley & Ritter 2002)



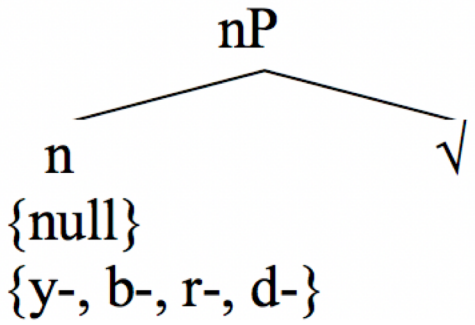
The connection between number and gender is accounted for under the hierarchical relationship between phi-features

The expression of noun class on the nominal constituent

- Noun class information is stored as a gender node at the lemma level as part of each noun entry (Carroll 1989) or as a gender node to which all nouns of this gender class are linked (e.g. Schriefers & Jescheniak 1999)
- A popular approach in terms of Distributed Morphology: acategorial roots √ combine with functional heads *n, v, a,* etc., and these heads carry the relevant grammatical information
- Default noun class: absence of features
- The hierarchical organization of features in the noun class (gender) node can follow the decision tree for noun class assignment

The expression of noun class on the nominal constituent

- Prefixes of noun class can be exponents of the categorizing *n* head that makes a noun a noun
- Exponents do not always take the same shape, but note their recurrence across the family (same in Romance)



Transfer of the relevant feature from the nominal constituent

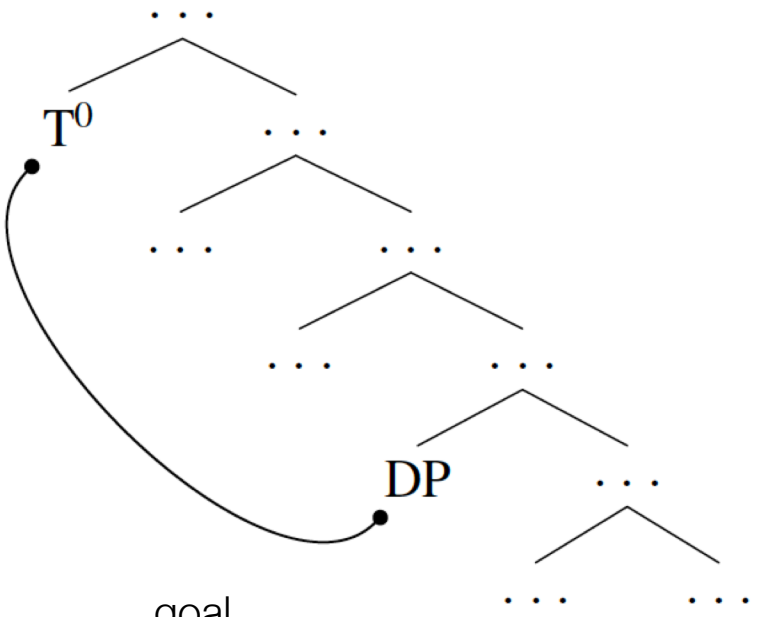
- Certain structural nodes (e.g., finite T or determiner) come into the derivation with a “need” which is met when that structural node acquires phi-feature values from a nominal
- How do we go from “I am a structural node with a need” to “here’s a nominal that can satisfy that need”?

Transfer of the relevant feature from the nominal constituent

- Certain structural nodes (e.g., finite T or determiner) come into the derivation with a “need” which is met when that structural node acquires phi-feature values from a nominal
- How do we go from “I am a structural node with a need” to “here’s a nominal that can satisfy that need”?
- The node with the need (=the probe) always c-commands the nominal(=the goal) at the stage of the derivation where the need is fulfilled

Agreement under Probe-Goal relationship, clause

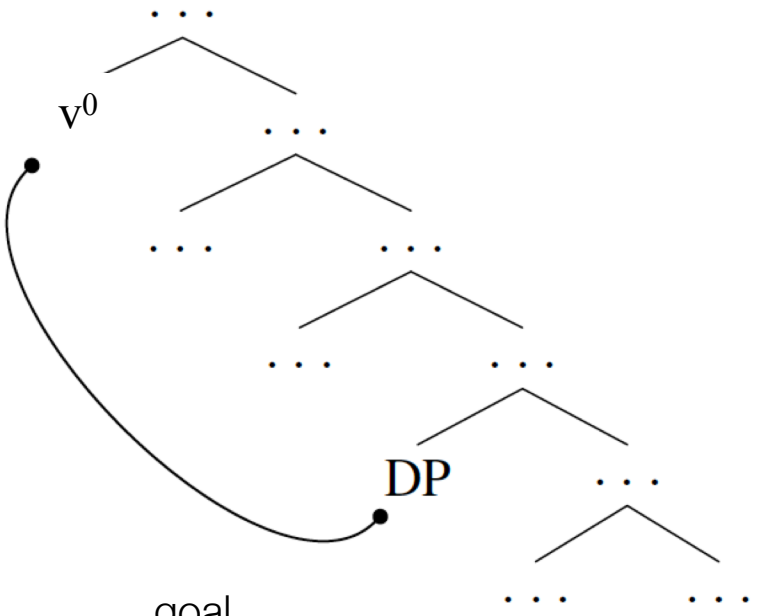
probe
(in need of
phi-feature)



goal
(has the needed
phi-feature)

Agreement under Probe-Goal relationship, verb phrase

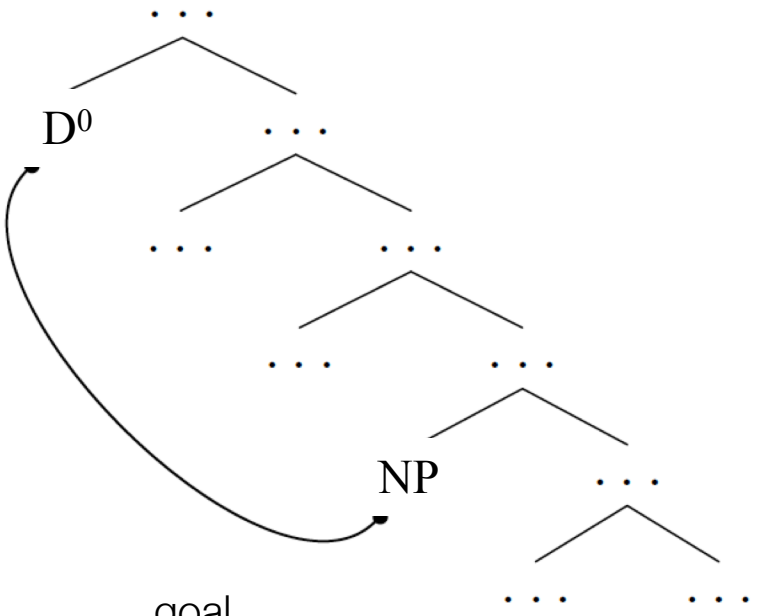
probe
(in need of
phi-feature)



goal
(has the needed
phi-feature)

Agreement under Probe-Goal relationship, noun phrase

probe
(in need of
phi-feature)



goal
(has the needed
phi-feature)

Agreement mechanism

- Approaches may vary but there seems to be a consensus that agreement is a syntactic relation subject to syntactic conditions
  - c-command
  - locality
- The output may be subject to morpho-phonological operations but their role is secondary
  - may have to do with how the relevant features are spelled-out
  - could be due performance factors

Agreement mechanism

- Syntactic structure is built from the bottom up
- when a new structural node P is introduced into the structure, all it can see is what's already in the structure
- = its sister, and everything contained within its sister
- = P's c-command domain

Agreement mechanism

- Syntactic structure is built from the bottom up
- when a new structural node P is introduced into the structure, all it can see is what's already in the structure
- = its sister, and everything contained within its sister
- = P's c-command domain

$[_{TP}$ [… DP…] T]

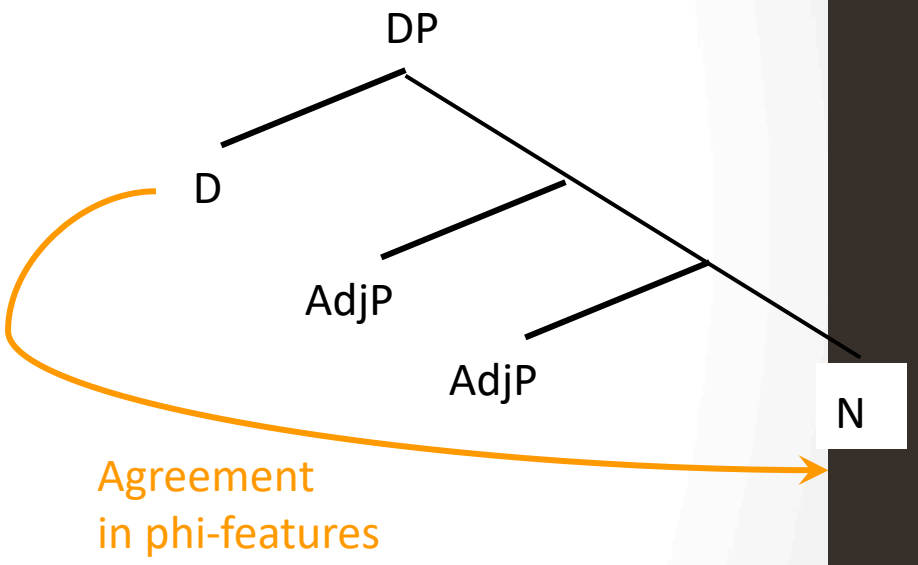
$[_{vP}$ $[_{VP}$ … DP…] v]

$[_{DP}$ $[_{NP}$ … nP…] D]

goal probe

Agreement is agreement is agreement

Probing head “looks down” and finds the necessary phi-features in its c-command domain



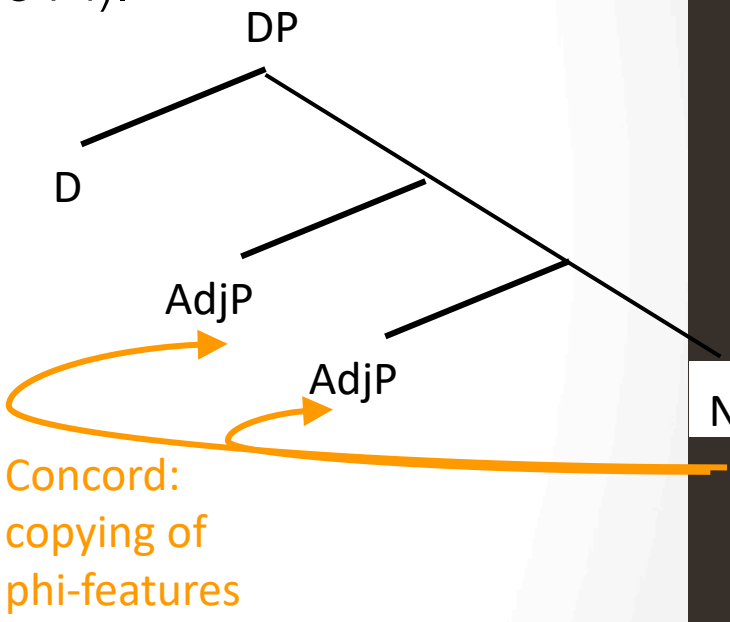
Is it all agreement all the way?

- Agreement in noun class in TPs, vPs, and DPs connects the probe (the element that ”needs” a noun class value) and the goal (the constituent that has that value); the syntactic mechanism is identical for all these domains
- Agreement on adjectival modifiers and demonstratives *may* be subject to a different mechanism: concord

Concord

It is possible that “concord” could be a different syntactic process from agreement (e.g., Norris 2014).

Concord = copying of phi-features in the phrasal domain that contains the nominal with the relevant features



Outstanding questions

Theory and description:

- More fine-grained internal organization of the phi-feature [GENDER] than usually adopted in the analyses of Spanish, Russian or German
- Relationship between [NUMBER] and [GENDER] in Nakh-Dagestanian noun class space
- Possible differences between agreement on verbal, clausal, and determiner heads on the one hand, and on modifiers, on the other: are these different syntactic operations, or one and the same operation?

Outline

- Noun class as a grammatical concept ✔
- Noun classification in the NE Caucasus ✔
- Noun class assignment (with a touch of acquisition) ✔
- Noun class agreement in the general model of syntactic agreement ✔

Taking stock

Here are two issues we may be interested in concerning noun classes in ND:

(i) How do speakers (adults and child learners) decide which noun goes into which class?

(ii) Once (i) is answered, how do heads actually come to carry noun-class agreement with their argument?

Modern linguistic theory has provided answers to both questions

Taking stock

- Modern linguistic theory has provided answers to questions concerning noun class assignment and agreement in Nakh-Dagestanian
- Detailed work on noun classifications in individual languages of the family is still needed
- We also need acquisition and experimental work on Nakh-Dagestanian noun classifications (along the lines of Gagliardi's work on Tsez)

Taking stock

- Do noun class (gender) distinctions lead to some degree of increase in distinctiveness in lexical access, pronominal reference, agreement?
- Within and beyond Nakh-Dagestanian, we still do not understand the (evolutionary) utility of noun classifications…

THANK YOU!

Selected references

Bateman, N. & M. Polinsky. 2010. Romanian as a two-gender language. In D. Gerdts et al. (eds.), *Hypothesis A/Hypothesis B*, 41–78. Cambridge, MA: The MIT Pres s.

Carroll, S. 1989. Second language acquisition and the computational paradigm. *Language Learning* 39: 535-594.

Comrie, B. & M. Polinsky. 1999. Gender in a historical perspective: Radial categories meet language change. In C. Justus & E. C. Polomé (eds.), *Language Change and Typological Variation: In Honor of Winfred P. Lehmann on the Occasion of His 83rd Birthday*. Vol . 2, Grammatical Universals and Typology, 566–589. Washington, DC: Institute for the Study of Man.

Corbett, G. 1991. *Gender*. Cambridge: Cambridge University Press.

Foley, S. 2020. Agreement in languages of the Caucasus. In M. Polinsky (ed.) *The Oxford Handbook of Languages of the Caucasus*, 845-872. Oxford: Oxford University Press.

Friedman, V. 2020. Lak. In M. Polinsky (ed.) *The Oxford Handbook of Languages of the Caucasus*, 204-242. Oxford: Oxford University Press.

Gagliardi, A. 2012. Input and intake in language acquisition. Ph.D. Diss. University of Maryland.

Gagliardi, A. & J. Lidz. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90: 58-89.

Harley, H. & E. Ritter. 2002 Person and number in pronouns: A feature-geometric analysis. *Language* 78: 482-526.

Harris, J. 1991. The form classes of Spanish substantives. *Yearbook of Morphology* 1991. 65–88.

Khalilov, M. 1999. *Tsezsko-russkij slovar'*. Moscow: Academia.

Lakoff, G. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.

Lyster, r. 2006. Predictability in French gender attribution: A corpus analysis. *French Language Studies* 16: 69–92.

Norris, M. 2014. A theory of nominal concord. Ph.D. Diss., UC Santa Cruz.

Plaster, K. & M. Polinsky. 2007. Women are not dangerous things: Gender and categorization. Harvard working papers in linguistics, vol. 12. 115–158. Cambridge, MA: Harvard University.

Plaster, K. & M. Polinsky. 2010. Features in categorization, or a new look at an old problem. In A. Kibort & G. Corbett (eds.) Features: Perspectives on a key notion in linguistics, 109-143. Oxford: Oxford University Press.

Plaster, K., M. Polinsky, & B. Harizanov. 2013. Noun classes grow on trees: Noun classification in the North-East Caucasus. In B. Bickel et al. (eds.) Language Typology and Historical Contingency. In honor of Johanna Nichols, 153-169. Amsterdam: John Benjamins.

Polinsky, M. 2015. Tsez syntax. lingbuzz/002315

Polinsky, M. 2016. Agreement in Archi from a minimalist perspective. In O. Bond et al. (eds.) Archi: Complexities of agreement in cross-theoretical perspective, 184-232. Oxford: Oxford University Press.

Polinsky, M., & D. Jackson. 1999. Noun classes: Language change and learning. In B. Fox et al.(eds.), Cognition and function in language, 29–58. Chicago, IL: University of Chicago Press.

Polinsky, M., N. Radkevich, & M. Chumakina. 2017. Agreement between arguments? Not really. In R. D'Alessandro et al. (eds.) The verbal domain, 87-120. Oxford: Oxford University Press.

Rajabov, R. 1997. The class category in Tsez: Underlying principles. In Dagestanskij lingvisticheskii sbornik. Vyp. 4. Moscow: Academia.

Rajabov, R. n.d. *Tsez dictionary* (unpublished manuscript). Los Angeles: University of Southern California.

Schriefers, H., & Jescheniak, J.D. 1999. Representation and processing of grammatical gender in language production: A review. *Journal of Psycholinguistic Research* 28: 575–600

Tanenbaum, K. E. 2003. *Modeling German gender* (unpublished manuscript). San Diego: University of California, San Diego.

Tucker, G. R., W. E. Lambert & A. Rigault. 1977. *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. The Hague: Mouton

Wier, T. 2014. What are the characteristics of the Batsbi language? Quora online https://www.quora.com/What-are-the-characteristics-of-the-Bats-Batsbi-language

The expression of noun class on the nominal constituent

- Gender on lemma or on little n
- Probing
- Agreement vs concord

- Outstanding questions (morphemes vs clitics

Reconstructing noun classifications

- Ideal case: we have information about the starting point, the end point, and intermediate changes
- Latin > medieval Romance > modern Romance languages

Reconstructing noun classifications

- Ideal case: we have information about the starting point, the end point, and intermediate changes
- Latin > medieval Romance > modern Romance languages
- Polinsky & van Everbroek (2003): used neural nets to reconstruct the change from Latin to Old French

Reconstructing noun classifications

- Polinsky & van Everbroek (2003): used neural nets to reconstruct the change from Latin to Old French
  - Used only nouns attested in the Vulgate
  - Added Celtic substrate
  - Did not model the actual phonological change
  - 79% accuracy
- Proof of concept indicating that the modeling approach works

Reconstructing noun classifications

- Ongoing project: Reconstructing Latin > Romanian (Lau, Polinsky, Stanton, in progress)
  - Used all Vulgar Latin nouns from Perseus corpus
  - Slavic adstrate
  - Actual phonological change in the model
  - Currently: 75% accuracy!
- Proof of concept indicating that the modeling approach works

Reconstructing noun classifications

- Using a modeling approach: go back, from current languages to the preceding stage. Main ingredients:
  - Current stage
  - Possible sound changes

- How well do we know the current stage?

Reconstructing noun classifications

- Using a modeling approach: go back, from current languages to the preceding stage. Main ingredients:
  - Current stage
  - Possible sound changes

- How well do we know **the current stage?**
- Evidence for the utility of modeling approach for synchronic categorization

Goals for today

- Present and analyze typical noun classification systems in North-East Caucasian languages

- Present proof-of-concept evidence for the utility of computational methods in establishing language-internal noun classifications

Outline

- General remarks on noun classification in Nakh-Dagestanian (NE Caucasian), with some ensuing questions
- Computational analysis of Tsez noun classes: good for Tsez, good for our toolkit

Head (source) gender

- Some cases of ‘head gender’ (Nichols): some nouns carry a source gender marker (compare A- and I-marking in Korean Sign Language, Byun et al.—talk yesterday)

	'brother' (1)	'sister' (2)	
Ingush	vosha	jisha	(Nakh)
Avar.chd	wac	jac	(Avar)
Akhvakh	wacci	jacci	(Andic)
Bezhta	is	is	(Tsezic)
Lak	usswu	ssu	(Lak)
Dargi.icari	ucci	rucci	(Dargi)
Archi	ush-du	dosh-dur	(Lezgian)
Rutul	shu	rishi	(Lezgian)

Head (source) gender

- Some cases of ‘head gender’ (Nichols): some nouns carry a source gender marker (compare A- and I-marking in Korean Sign Language)

	'brother' (1)	'sister' (2)	
Ingush	vosha	jisha	(Nakh)
Avar.chd	wac	jac	(Avar)
Akhvakh	wacci	jacci	(Andic)
Bezhta	is	is	(Tsezic)
Lak	usswu	ssu	(Lak)
Dargi.icari	ucci	rucci	(Dargi)
Archi	ush-du	dosh-dur	(Lezgian)
Rutul	shu	rishi	(Lezgian)

- Ingush: 10% of nouns have source gender
- Nichols (1989) proposed as a feature of PND

- 4. What is the "default" class?
  - It's unclear whether III or IV is the default
  - Class III is the largest class in our sample, but size alone can't justify making III the default
  - Gagliardi et al. (2009) and Gagliardi (2012) found an apparent difference in the default class used by kids and adults; kids seemed to use III while adults used IV.