

Mark Norris, PhD

Linguist, Language Engineer, Conversation Analyst

Typology, Morphology, Syntax, Semantics



Agreement with nouns in the world's languages

(more formally, A typological perspective on nominal concord)

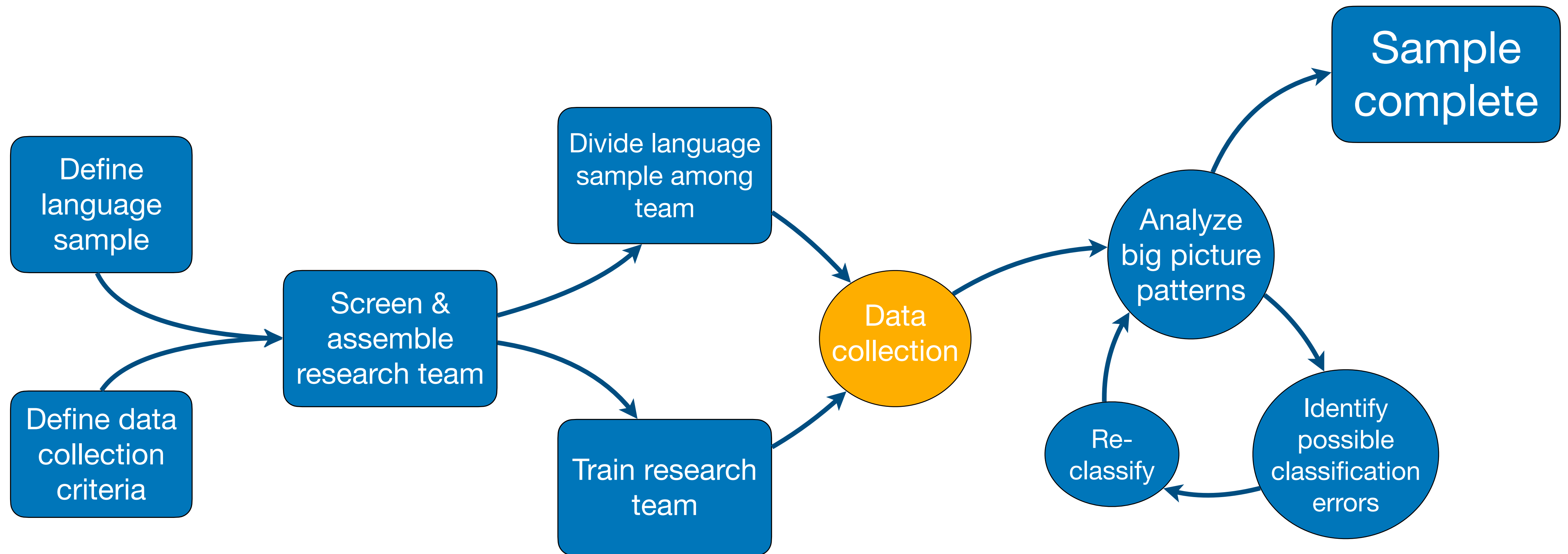


[Link to archived data](#)

- Focus: agreement with nouns akin to English *these books* vs. *this book*
- Built linguistically diverse sample of languages (174 in archive, over 200 total) including low-resourced languages
 - Trained research assistants to gather appropriate examples from secondary sources
 - Designed and implemented categorization criteria for collected examples, managed data in Excel
- Analyzed qualitative trends to determine patterns characteristic of agreement with nouns in the world's languages
- Study showed that just over half of languages had some form of agreement with nouns, with **number agreement** and **adjective + demonstrative agreement** being the most common.

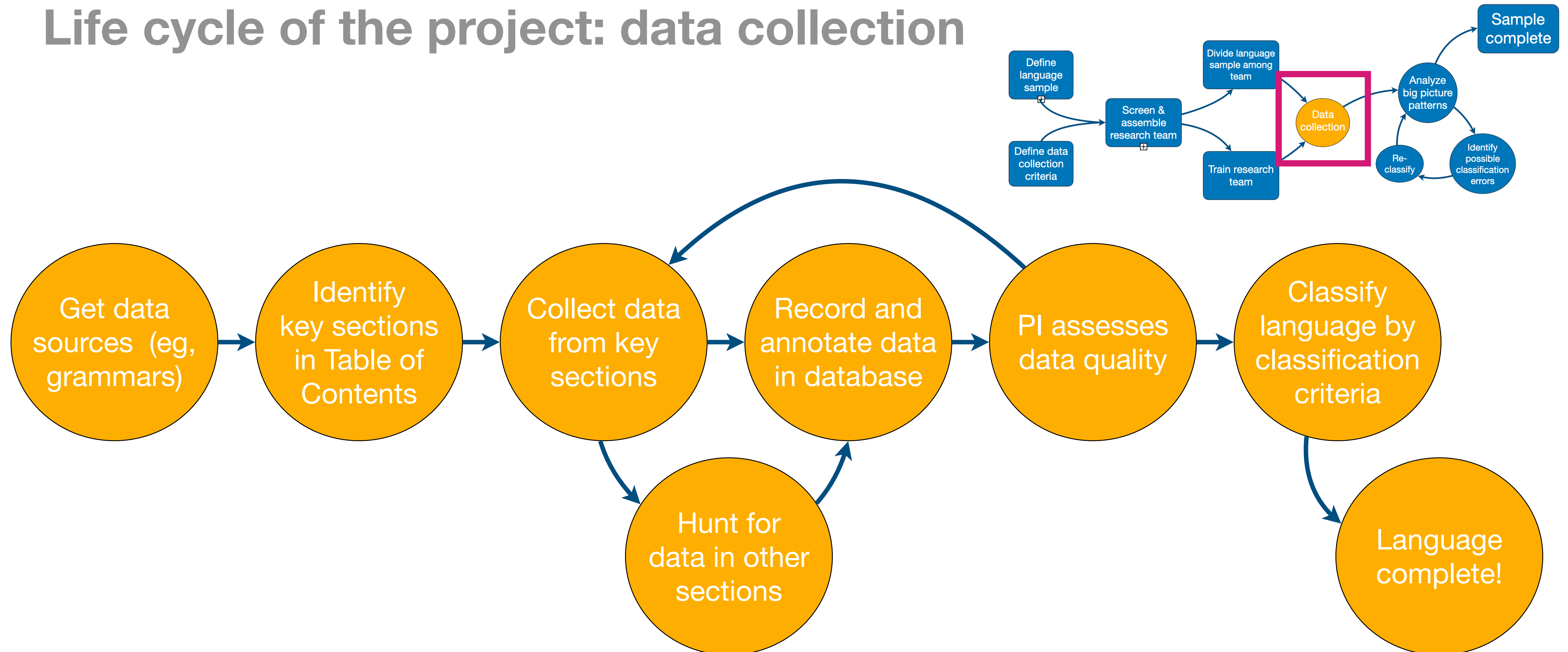
Agreement with nouns in the world's languages

Life cycle of the project: broad view



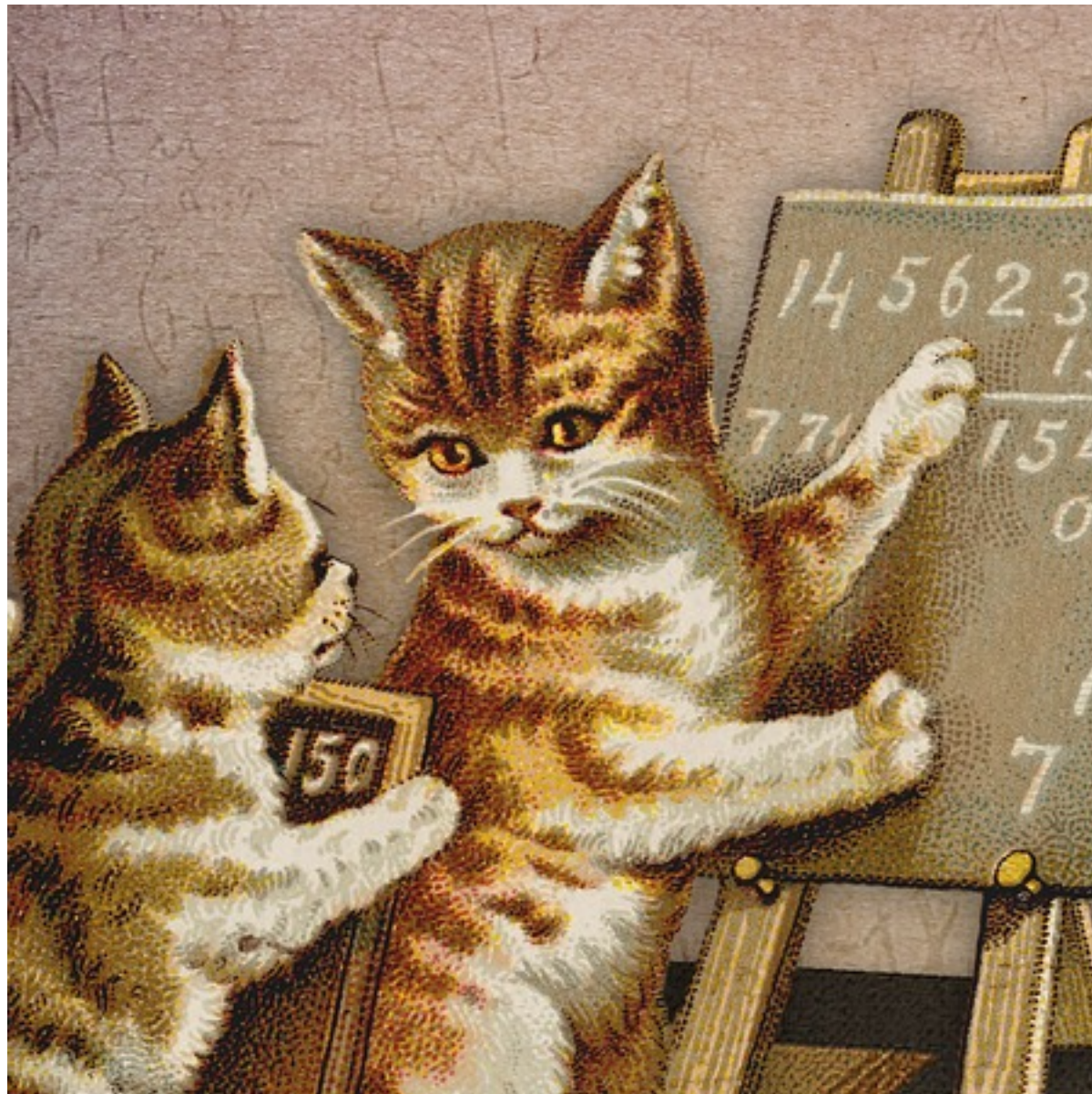
Agreement with nouns in the world's languages

Life cycle of the project: data collection



Ways of counting in Estonian

Corpus study and formal analysis



[Link to article](#)

- Focus: when nouns combine with numerals in Estonian, numeral and noun are both marked plural or both marked singular (ie, both *two mitten* and *two-s mitten-s*)
- Conducted a corpus study of plural numerals to assess range of possible meanings (the singular numerals are the default, and the plural numerals have specialized uses)
 - Collected 418 tokens of plural numerals from corpora of Estonian (available [here](#))
 - Categorized each token based on meaning; defined 4 broad types of meanings where plural numerals are used.
- Results showed that plural numerals can be used for **much broader range of words** than previously reported, but they are **used to count groups of objects** rather than individual objects.

Ways of counting in Estonian

Data categorization and management (in Excel)

Corpus	Numeral	Case	Noun (sg)	Noun (pl)	Full Token	Translation	type	example							
ettenten	three	ade	valimine	valimised	viimastel valimistel	election	pluralia plerumque	arvamus.postimees.ee : [Siiski on Rahvaliit meie hinnangul jõudnud ummikteele :] [erakonna toetus on pikka							
ettenten	two	nom	koordinaat	koordinaadid	erinevad kordinaad	coordinate	pluralia plerumque	www.geopeitus.ee : [4 november 2008 leidis Lembit [polekala] Olin varem välja arvutanud kahed erinevad k							
ettenten	two	nom	poeg	pojad	pojad	boy	contextual group	bio.edu.ee : [Aastas on neil kahed pojad ,] [esimeses kurnas on 4...6 muna , teises harilikult ühe või kahe v							
ettenten	two	nom	koordinaat	koordinaadid	koordinaadid	coordinate	pluralia plerumque	www.geopeitus.ee : [9 oktoober 2011 leidis Alex [alex] Selle aarde sõnastus suutis mind jälle mõtlema panna							
ettenten	two	all	konflikt	konfliktid	järgnevad konfliktic	conflict	contextual group	www.eetika.ee : [Õpilaste õppeedukus langeb , kahtelede järgnevad konfliktid ning suhted õpilaste ja õpetaj							
ettenten	three	ade	olümpiamäng	olümpiamängud	olümpiamängudel	Olympic games	natural group	www.eok.ee : [Kõige enam tähelepanust kõrvalejäänu :] [ungarlanna Krisztina Egerszegi ,] [kes oma võidu							
ettenten	three	nom	eestlane	eestlased	eestlased	Estonian	contextual group	debate.ee : [Seetõttu oli väga äge ,] [et eestlastel läks selles seltskonnas üle keskmise hästi :] [poolfinaali							
ettenten	two	nom	mälestus	mälestused	mälestused	memoirs	pluralia tantum	www.vikerkaar.ee : [Raamatuid [, mis seda lugu inspireerisid ,] oli aga kolm , kahed mälestused ja üks päe							
ettenten	three	nom	rahutus	rahutused	rahutused	secession/rebellion	pluralia tantum	www.ekspress.ee : [Möödunud nädalal toimusid näiteks Tais koguni kolmed rahutused samaaegselt ja üsna							
ettenten	three	nom	kangastelg	kangasteljed	kangasteljed	loom (fabric axles)	pluralia tantum	www.maaleht.ee : [Pärnumaal asuv Pootsi vaibakoda on just nii suur ,] [et sinna mahuvad kolmed kangaste							
ettenten	three	nom	poeg	pojad	pojad	boy	contextual group	www.folklore.ee : [Vanad naesed rägivad ,] [et kui taeva tähed hästi heledasti paistavad ,] [siis toovad lamr							
ettenten	two	nom	hambauk	hambaaugud	hambaaugud	cavity	contextual group	www.nupsu.ee : [Ma käisin tüdrukutega hambaarstil 3-selt,sest avastasin ,] [et neil oli mõlemal kahed hamb							
ettenten	two	nom	valimine	valimised	valimised	election	pluralia plerumque	arvamus.postimees.ee : [Läinud aastaga [, mil toimusid kahed valimised ,] jõudis avalik arvamus Eesti poli							
ettenten	two	nom	kolmik	kolmikud	kolmikud	triplet	natural group	www.epl.ee : [TÜ kliinikumi naistehaigla info kohaselt on 2008. aasta kahe kuuga sündinud 14 paari kaksikuid							
ettenten	two	ade	valimine	valimised	järjestikustel valimi	election	pluralia plerumque	www.vikerkaar.ee : [Kuid need kasutatavast süsteemist tulenevad eelised ei püsi üldiselt kauem kui ühtedel v							
ettenten	two	com	valimine	valimised	järjestikuste valimis	election	pluralia plerumque	www.keskerakond.ee : [Nüüd on ees üks aasta arupidamist , enese kogumist ja kosumist ning kohe seejärel .							
ettenten	two	nom	võistlus	võistlused	EM-võistlused	competition	pluralia plerumque	www.autoleht.ee : [Kuuekümnendate alguses võitsid Walter Schock ja Eugen Böhringer Mercedes-Benz 220 S							
ettenten	two	nom	riie	riided	trenniriided	clothes (gym)	natural group	www.nommerattaklubi.ee : [Kõigil kaasa voodipesu , käterätik , isikliku hügieeni asjad , kahed trenniriided õ							
ettenten	two	ade	valimine	valimised	valimistel	election	pluralia plerumque	arvamus.postimees.ee : [Samasugused kaotused oldi sunnitud üle elama ka sisepoliitikas :] [kahtedel valin							
ettenten	two	ade	unistaja	unistajad	unistajatel	daydreamer	contextual group	www.kalale.ee : [Mis neil kahtedel unistajatel üldisemalt laiendades vahet on ?]							
balanced	three	nom	kolmik	kolmikud	kolmikud	triplet	natural group	SLÕhtuleht_2001 : [Eestis sünnivad igal aastal keskmiselt ühed kuni kolmed kolmikud .]							
ettenten	two	nom	võitja	võitjad	võitjad	winner	contextual group	www.juhtimisaju.ee : [Oleme selgitanud kahed võitjad .]							
ettenten	two	nom	jahu	jahud	jahud	flour	kind	www.folklore.ee : [Rukkijahudega saab ka hautud :] [võetakse kahed jahud ,] [pannakse ühed panni peale							
ettenten	three	ela	mäng	mängud	mängudest	game	natural group	www.loomakaitse.ee : [Võistluste veteran Sione Belle Togiavalu [, kes on osa võtnud kolmedest mängudest							
ettenten	two	nom	kõrv	kõrvad	kenad hobusekõrva	ears (horse)	natural group	www.horsemarket.ee : [Väiksemad lapsed said seni jäätist , ma ise aga eesesisva pika retke pärast muret tu							
ettenten	three	nom	kuldmedal	kuldmedalid	kuldmedalid	gold medals	contextual group	www.fcflora.ee : [FC Flora koosseisus riputati mängija kaela kolmed kuldmedalid , üks hõbedane ja kaks pr							
ettenten	two	nom	olümpiamäng	olümpiamängud	olümpiamängud	Olympic games	natural group	www.med24.ee : [Olen töötanud Eesti suusakoondise juures alates 1997. aastast ning] [sellesse ajavahemik							
ettenten	two	nom	kont	kondid	kondid	bone	contextual group	www.folklore.ee : [Teine kord mõisarehe juures näidatud jälle ühte kohta ning pärast vanaema leidand säält ka							
balanced	two	nom	jälg	jäljed	jäljed	track	natural group	Maaleht_2002 : [Siia tulid kahed jäljed .]							
ettenten	two	nom	PT	talgud	talgud	bee/festival	pluralia tantum	www.teemeara.ee : [Tabiveres kahed talgud Tabivere alevikus viidi sellel aastal läbi kahed talgud ,] [korda							
ettenten	two	ade	olümpiamäng	olümpiamängud	olümpiamängudel	Olympic games	natural group	www.vibuklubiilves.ee : [Aga ei maksa unustada ,] [et ta on Eesti paremikus püsinud juba rohkem kui kümme							
ettenten	three	nom	läbirääkimine	läbirääkimised	läbirääkimised	negotiation	pluralia plerumque	www.tehnikamaailm.ee : [Praegu pole Eestis konkreetset passiivmaja , nullenergiama ja ega positiivenergiama							

Ways of counting in Estonian

Semantic range of plural numeral meanings

Group 1: Morphological plurals

Words that must (usually) be plural when speakers use them (like English *measles*, *pants*, or *shenanigans*)

Group 3: Contextual Groups

Words that denote objects in groups that are defined contextually (eg, *three-s Estonian-s* to mean 'three teams of Estonians' in a competition)

Group 2: Natural Groups

Words that denote objects which come in natural pairs/groups in the world (like *gloves*, *animal tracks*, or *shoes*)

Group 4: Semantic Kinds/Sorts

Words that refer to kinds of objects rather than discrete instances of that object (eg, *two-s degree-s* to mean 'two kinds of (college) degrees')

Corpus data collection for linguistic research

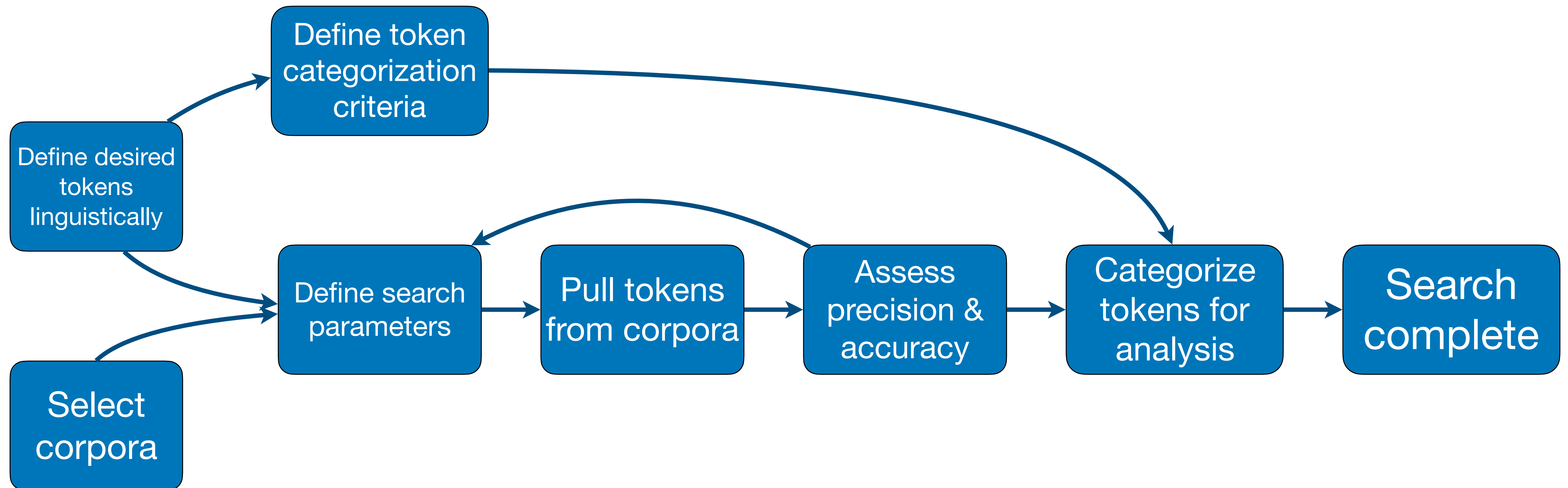
Naturalistic data for stronger empirical generalizations



- Focus: combining data collected from linguistic fieldwork with similar tokens from corpora to ensure strength of empirical generalizations and to uncover new patterns
- Collected examples for projects on case morphology, noun phrase structure, and choice of pronoun form in context
- Iterated pull requests to ensure broadest range of tokens while minimizing false positives
- Managed token collection and annotation in Excel

Corpus data collection for linguistic research

Life cycle of a corpus search



focus¹, pl. **foci**, **focuses** [ˈfoukəs, ˈfousai, ˈfoukəsiz] n. 1. Mth: *Opt: etc:* foyer *m* (de lentille, etc.); *Opt: depth of f.*, (i) profondeur *f* de foyer; (ii) profondeur de champ; in *f.*, (i) (of image) au point; (ii) (of instrument) réglé; out of *f.*, (i) (of image) pas au point; (ii) (of instrument) non réglé; déréglé; (iii) (of headlamp bulb, etc.) mal réglé; to bring sth. into *f.* mettre qch. au point; *Photo: Auto-f. camera, ap-*

THANK YOU!

MARK NORRIS, PhD
MJNORRIS@GMAIL.COM
MORPHOSYNTAX.ORG