Paola Crisma, Cristina Guardiano, Giuseppe Longobardi

Syntactic diversity and language learnability

Appendix

Part 1: Introduction to parameters and manifestations

INSTRUCTIONS

1) This Appendix provide a protocol for the replicability of data collection and their coding as parameter states. The data refer to the lists of parameters used in the following publications:

- Andrea Ceolin, Cristina Guardiano, Monica Alexandrina Irimia and Giuseppe Longobardi (2020) Formal syntax and deep history. Frontiers in Psychology 11: 488871.
- Patrícia Santos, Gloria Gonzalez-Fortes, Emiliano Trucchi, Andrea Ceolin, Guido Cordoni, Cristina Guardiano, Giuseppe Longobardi and Guido Barbujani (2020) More rule than exception: parallel evidence of ancient migrations in grammars and genomes of Finno-Ugric speakers. Genes 11, 1491.
- Andrea Ceolin, Cristina Guardiano, Giuseppe Longobardi, Monica Alexandrina Irimia, Luca Bortolussi and Andrea Sgarro (2021) At the boundaries of syntactic prehistory. Philosophical Transactions of the Royal Society B 376: 20200197.

2) The present file (**Part 1**) contains: a short description of the structure of the manifestations (section 1) and of their conditions of application (section 2), along with a glossary of some technical terms (section 3). It can also be downloaded at www.parametricomparison.unimore.it, section "Materials".

3) A very short description of each parameter along with an updated list of its empirical manifestations is contained in a separate dedicated file (**Part 2**), that can be downloaded at www.parametricomparison.unimore.it, section "Materials".

**** _____ **** _____ **** _____ **** _____ ****

1.   Formal structure of the manifestations

Each parameter is associated with one or more manifestations that set its state. Each manifestation on its own suffices to set the relevant parameter to the state '+': thus, each manifestation is translatable into a YES/NO question asking about the truth of an existential statement of the type "in language L, there is an observable grammatical object (construction/morpheme/feature/etc.) α with property X", or a conjunction of two or more statements of this type.

More precisely, such questions generally obey the following properties:

1) YES answers can always be provided just on the basis of positive evidence. For this purpose, all bare plurals in the formulation of the questions are meant to be interpreted as existential rather than generic/universal. We made sure that for every parameter there is at least one question that might be answered YES in some known language. This guarantees the minimal requirement of cognitive plausibility for that parameter, given that language learners can certainly access positive evidence.

2) one YES answer (i.e. one manifestation per parameter and per language) is sufficient to set a parameter's value unambiguously to '+'. In a language, all the manifestations of a parameter should co-vary across languages by definition. Thus, normally, one YES answer correlates with YES answers to all the questions for the same parameter. Yet, in several cases, an answer NO (more accurately the lack of an answer YES) may just be the consequence of the absence of the relevant construction in a language due to independent conditions: namely, the combinations of other parameter values or the accidental lack of the relevant morpheme as an idiosyncrasy of the functional lexicon (ultimately Saussurean arbitrariness).

3) if no question relative to any manifestation of a certain parameter in a language receives an answer YES, the value '-' will be assigned by default to that parameter in that language. Therefore, '-' is the unmarked state of each parameter, while '+' is always chosen on the basis of positive evidence (**NB**: very occasionally a manifestation/question requiring negative evidence has been

2

added to the others for the sake of easiness of parameter setting by a linguist interviewing a native speaker, but no parameter has been formulated as requiring just negative evidence to be set).

2. Conditions of application of the manifestations/questions
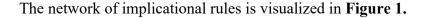
The manifestations setting each parameter are so formulated as to be relevant only if the parameter needs to be set to (either of) the two alternative parameter states '+' or '-', i.e. if it is not independently neutralized in a certain language owing to the interaction with other parameter values (the state marked as '0' as a consequence of the implicational rules): if a parameter is implicationally neutralized in a certain language, the questions for that parameter in that language must be disregarded altogether: they would be irrelevant and in some cases misleading.
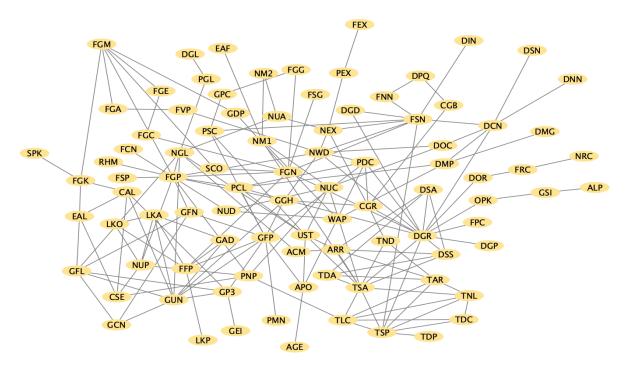
The list of the 94 parameters used in this work is contained in **Table 1** along with the implicational rules for each parameter. The table can also be downloaded at www.parametricomparison.unimore.it, section "Materials".

Each parameter in the Table is conventionally identified by a progressive number (in the first column from the left) and more rigidly designated by a combination of three capital letters (in the second column). The order of the parameters is only motivated by the ease of expression of cross-parametric dependencies, which are so organized as to proceed top-down. The conditions that must hold for each parameter to be relevant (i.e. not neutralized to '0') are indicated in the fourth column after the name of the parameter itself. They are expressed in a Boolean form, i.e.: either as simple values of other parameters, or as conjunctions (written ','), disjunctions ('*or*'), or negation ('¬') thereof. As a space-saving convention, in the implicational rules, disjunctions (which are all logically meant to be inclusive: *vel*, not *aut*) are always meant to be parsed first, conjunctions later, unless parentheses are used to explicitly signify the opposite order of embedding. Thus, as an example of how to read the notation, the implicational condition of parameter 20 (NWD) should sound as follows: p20 (NWD) can be set (to either '+' or '-') if and only if: p8 (FGP) is set to + and p9 (FSN) is not set to + or if and only if p14 (DGR) is set to + (or both disjoined conditions hold); otherwise it will be neutralized (0).

	Label	Parameter	Implication(s)
1	FGM	± grammaticalized morphology	
2	FGA	± grammaticalized agreement	+FGM
3	FGK	± grammaticalized Case	+FGM
4	SPK	± grammaticalized (ultra-)spatial Cases	+FGK
5	FGP	± grammaticalized person	+FGM
6	FSP	± semantic person	¬+FGP
7	FGN	± grammaticalized number	+FGP
8	SCO	± spread collective number	+FGM, ¬+FGN
9	GDP	± grammaticalized distributive plurality	+FGM, ¬+FGN
10	FSN	± number spread to N	+FGN
11	FNN	± number on N	+FSN
12	FGT	± grammaticalized temporality	
13	FGG	± grammaticalized gender	+FGN
14	FSG	± semantic gender	+FGN
15	CGB	± unbounded sg N	
16	FPC	± grammaticalized perception	
17	DGR	± grammaticalized Specified Quantity	-FPC, +FGN
18	DGP	± grammaticalized text anaphora	¬+DGR
19	CGR	± weak Specified Quantity	-CGB, +DGR
20	NWD	± weak person	+FGP, -FSN or +DGR
21	FVP	± variable person	+FGA, -NWD
22	DGD	± grammaticalized distality	-FSN or +DGR
23	DPQ	± free null partitive Q	+FNN, -CGB
24	DCN	± article-checking N	-FSN or +DGR
25	DNN	± null-N-licensing art	-DCN
26	DIN	± D-controlled inflection on N	+FSN
27	FGC	± grammaticalized classifier	¬+FGN
28	FGE	± general classifier	-FGM, +FGC
29	FCN	± person spread to predicate nouns	+FGP
30	HMP	± NP-heading modifier	
31	ARR	± free reduced relatives	
32	GCN	± head-marking	
33	GFN	± Person controlled marking	+FGP, +GCN
34	GFP	± agreement with all pronouns	+GFN
35	GP3	± agreement with all 3rdPers DPs	+GFP
36	GEI	± genitive inversion	+GP3
37	CSE	± full c-selection	
38	EAL	± ergative alignment	+FGK, +CSE
39	CAL	± clausal alignment	+FGK, +CSE, ¬+GP3,¬+EAL
40	LKA	± argument linker	
41	LKO	± oblique linker	-LKA
42	LKP	± predicative linker	
43	DMP	± def matching pronominal possessives	+DCN
44	DMG	± def matching genitives	+DMP
45	GUN	± uniform genitive	(-GCN or (+GFP, -GP3)), -CAL, -LKA
46	GAD	± free Gen	-LKA, ¬+GUN
47	GFL	± GenL	(-GCN or +GFN), ¬+GP3, ¬+EAL, ¬+GUN
48	PGL	± partial GenL	-DGL
49	GGH	± generalized GenH	+FGP, -CGR, +NWD, ¬+GFP, ¬+GUN
50	GSI	± grammaticalized inalienability	
51	ALP	± alienable possession	-GSI
52	GIT	± genitive-licensing iteration	
53	UST	± unstructured modifiers	+ARR
54	GPC	± gender polarity cardinals	+FGG
55	PSC	± plural spread from cardinal quantifiers	+FSN, ¬+UST, ¬+GPC
56	PCA	± plural spread through cardinal adjectives	-PSC
57	PMN	± person marking on numerals	+GFP
58	RHM	± head marking on relative clauses	+FGP
59	FRC	± finite relative clauses	
60	NRC	± participial relative clauses	+FRC
61	DOR	± def on relatives	+DGR, +FRC
62	FFP	± feature spread to particles	+FGN, (+LKA or +LKP or +LKO or +GAD), ¬+GFP
63	NUP	± NP under non-genitive arguments	+FGP, (+CSE or +LKA or +LKO)
64	PNP	± complement under P	+FGP, -CSE or -NUP
65	NUD	± NP under D	+FGP
66	NUC	± N under cardinals	¬+UST, +PNP, +NUD
67	NM1	± N under M1 As	+NUC
68	EAF	± fronted high As	-NM1
69	NM2	± N under M2 As	+NM1
70	NUA	± N under As	+NM2
71	NGL	± N under GenL	(+FGP, +UST) or +NUA, (+GUN or +GFL or +PGL)
72	ACM	± class MOD	-ARR, -NGL
73	DSN	± def spread to N	+DCN
74	DSA	± def spread to ARR	+DGR, +ARR
75	DSS	± def spread to structural categories	+DGR, (-ARR or +DSA)
76	DOC	± def on cardinals	-NWD, +DCN, +NUC
77	NEX	± Proper names in D	(-FSN or -CGR), -NWD, ¬+NUA
78	PEX	± Personal names in D	+NEX
79	FEX	± Partial personal names in D	+PEX
80	PDC	± D-checking possessives	+DGR, (¬-CGR or -NWD), ¬+GFP
81	PCL	± clitic possessives	+FGP, ¬+GFP, ¬+DMP, ¬+UST, (-PDC or +DGR)
82	APO	± adjectival possessives	¬+GFP, ¬+UST
83	WAP	± Wackernagel possessives	¬+DMP, +NUD, -PDC, -APO or (-NM1, +APO)
84	AGE	± adjectival genitive	+APO
85	OPK	± null possessive licensing article with kinship nouns	+DGR, -GSI
86	TSP	± split deictic demonstratives	-FSN or +DGR
87	TDP	± split demonstratives	+TSP
88	TDC	± D-checking demonstratives	-TSP
89	TSA	± structured demonstratives (adjectival)	¬+UST, ¬+TSP, ((+DGR, +NM1) or (-ARR, -NM1) or -NUC)
90	TAR	± unstructured demonstratives	+ARR, ¬+TSP
91	TLC	± Dem fronting to Loc	¬+TSP, ¬+TDC, (+TSA or (+PNP, +TAR))
92	TND	± long distance D-checking demonstratives	+CGR, (+TSA or +TAR)
93	TDA	± split def on adjectival demonstratives	(+DSS or +DSA), (+TSA or +TAR)
94	TNL	± DP under Loc	+TSP or +TLC or (¬+TSP, ¬+TDC, ¬+TSA, ¬+TAR)

Table 1

The network of implicational rules is visualized in **Figure 1.**

Figure 1 - The network of implications among 91 of the 94 parameters presented in Table 1 (three parameters are independent of any implicational relation: FGT, ±grammaticalized temporality, HMP, ±NP-heading modifier, and GIT, ±genitive-licensing iteration). Every line represents an implicational condition between two parameters.

3.   GLOSSARY of some technical usages

(morphological) Case

In nominal morphology it is important to have a demarcation criterion between inflectional (*lato sensu,* i.e. including agglutinative morphology) Case proper and pre-/post-positional words. Two criteria are conceivable, as a first approximation:

a.   many prepostional languages have Case *suffixes*, so that they are easily distinguished positionally from phrase-initial functional words like prepositions

b.   for postpositional languages, a demarcation criterion between Case-suffixes and postpositions is necessary, however. A distinction can be made if (and probably only if):

1)   the suffix occurs on a head noun or adjective before some other phrase-internal word (another adjective, a relative clause…)

        or

2)   the suffix occurs both on the head noun and some other phrase-internal category (adjective, quantifier, demonstrative) agrees with it through a corresponding suffix or

3)   the suffix is only on the head noun but, in the morphological structure, it is more word-internal than other noun suffixes, such as e.g. of number, possessed status etc.

Speech role

It refers to the semantic interpretation of a nominal phrase as denoting the speaker(s), the hearer(s), both, or any individual other than the above. It is encoded as the so-called person feature in many languages

Determiner

A functional morpheme (in many languages instantiated by an article, a demonstrative, a possessive, or a quantifier) normally occurring in, or connected to, a position at the highest boundary of a nominal phrase, able to shift the latter phrase into an individual-denoting expression and often also to ensure an *atomizing* function (from an unbounded to a *bounded* interpretation of the nominal). In most languages there are at least some instances of phonologically null determiner positions, possibly interpreted at a distance from other elements within the nominal phrase.

Article

It is used to refer to a determiner that does not express any meaning other than (in)*definiteness* or just φ-features, and sometimes even less interpretable content (expletive articles). In some languages, articles (normally phonologically unstressed) occur as morphosyntactically free morphemes, in others they are bound morphemes affixed to the head noun (or an adjective).

Bare noun (bare nominal argument)

A noun (or its extended maximal phrase) not introduced by any overt *determiner* in the canonical Determiner position nor with a determiner-like (*atomizing*) interpretation derived at a distance from other elements within the nominal phrase. Bare nouns in this sense may contain modifiers like arguments, adjectives or relative clauses, provided they do not contribute an atomizing interpretation.

Definite(ness)

The interpretation of the denotatum of a nominal phrase as being considered maximal in the shared domain of discourse, in many languages provided by designated *articles*, by demonstratives, *possessives,* or inherited even at a distance from certain *Genitives*. Definite arguments can be specific (i.e. assume the existence of a denotatum) or non-specific, and definiteness and specificity must by no means be confused.

Noun modifier

Any constituent within the maximal phrase of a head noun beyond the thematic arguments of the head noun and the elements occupying the position of determiners or performing an *atomizing* function at a distance from it.

Bounded vs. unbounded reading

A nominal argument will be said to have an unbounded reading when it is at the same time obligatorily indefinite, scopeless, and atelic. This reading is typically instantiated in European languages by bare plural and mass nouns and in many Turkic and Uralic languages by bare singulars. Nominal arguments introduced by an overt determiner normally have a bounded reading, which specifies the quantity of individuals denoted and *atomizes* the kind expressed by the head noun.

Phrase boundary

In principle the two (left- and right-ward) external edges of a nominal phrase. In practice the expression is used to refer to the one where the D position occurs in the language (initially in most languages, but clearly finally in such languages as Basque or Wolof). Notice that potentially floating quantifiers (such as those meaning 'all') and in certain languages even demonstratives may occur to the left of definite determiners without affecting the D-initial status of nominal phrases, i.e. the identification of a left boundary.

Possessive

A personal pronoun (or reflexive) expressing a genitive argument of a head noun if and only if it has a form different from that used to realize non-pronominal genitive arguments.

Genitive (genitive)

Genitive (with capitalized initial as a noun, but spelt with lower-case initial when used as an adjective) refers to the abstract Case considered assigned to the direct arguments of a head noun (those normally expressing the possessor, agent and theme relation). It must not be confused with just morphological Case: direct adnominal arguments realized as adpositional phrases, with inflectional marking or with no marking at all will all be considered Genitives in this sense.

Classifier

A morpheme connecting a cardinal numeral to a non-plural head noun in a nominal phrase interpreted as count.

Structured Adjectives

Adnominal adjectives occur in most languages (also) with a distribution separate from that of modifiers such as relative clauses. In this case they respect a certain fixed order when they appear before the head noun, but occur either in the same or in reverse order if they superficially appear after the noun. These adjectives are called structured. In those languages/constructions in which adjectives can or must appear with the distribution of relative clauses they appear in freer order and will be considered reduced relative clauses.

Atomizing

The process shifting the interpretation of a nominal argument from an *unbounded* reading (a free variable, indirectly bound) to a *bounded* one (a variable bound by a specific determiner).

φ-features

Any subset of person, number, gender and Case specifications in a language.