

DAS GRAMMATISCHE RARITÄTENKABINETT

a leisurely collection to entertain and instruct

Frans Plank
(Universität Konstanz)

with some items and/or languages brought to my attention by:
Michael Cysouw, Lena Filimonova, Orin Gensler, David Gil, Aditi Lahiri,
Gilbert Lazard, Irina Nikolaeva, Wolfgang Schellinger, Pieter Seuren, Horst Simon,
M. L. Souag, Ton van der Wouden

last update: 22 February 2001

CONTRIBUTIONS SOLICITED!
send to: frans.plank@uni-konstanz.de

and visit: <http://www.ling.uni-konstanz.de/pages/proj/Sprachbau/rara.html>

A *Raritätenkabinett* is a collection of things, living or more often dead, which are considered valuable, or at any rate worth collecting and perhaps exhibiting, because they are rare —not such things, though, whose commodity or exchange value would grow with rareness, but ones whose very existence, contrary to experience or indeed reason, makes the beholder marvel at the design of the universe.

The *rara*, *rarissima*, and *singularia* collected here come from the realm of grammar, wondrous enough in its everyday manifestations.

A **rarum** is a trait (of any conceivable sort: a form, a relationship between forms, a matching of form and meaning, a category, a construction, a rule, a constraint, a relationship between rules or constraints, ...) which is so uncommon across languages as not even to occur in all members of a single (middle to lower level) family or diffusion area (for short: sprachbund), although it may occur in a few languages from a few different families or sprachbünde.

Diachronically speaking, a **rarum** is a trait which has only been retained, or only been innovated, in a few members of a single family or sprachbund or of a few of them.

Thus, on this definition, a trait may well be infrequent (an **infrequentale**, so to speak; Latin purists might prefer **infrequens**) without being a **rarum**. Just think of such precious family or (former) sprachbund possessions as syntactically governed word-initial segmental alternations, generally infrequent but omnipresent at least in Celtic and Berber (and also

heard in Nias (Malayo-Polynesian, Austronesian) and in Iwaidja and Marrgu (Iwaidjan, Australian)); trilateral roots, found throughout Semitic and possibly Berber; or clicks, shared by the entire Khoesan family, though outside only by some (former) Bantu neighbours as well as, at some distance, by Dahalo (Cushitic, Afroasiatic). Presumably, however, allowances ought to be made for single-member families also being able to host rara rather than only infrequentalia.

Ararissimum is even rarer than a rarum.

A **singulare** (or, less highfalutinly, a **nonesuch**) is unique to individual languages.

A **pink** dot in front of an exhibit means a published universal is being infringed on by a rarum/rarissimum/nonesuch, which naturally throws doubt on that universal, interpreted categorically; black-dotted rara/rarissima/nonesuches are inoffensive.

If interested in universals, visit *The Universals Archive* at: <http://www.ling.uni-konstanz.de/pages/proj/sprachbau.htm>

As yet, the documentation of the Raritätenkabinett leaves much to be desired: often references are only given to secondary literature where a trait has been noted as being rare, but not to original sources. This will be remedied in future.

NB: Needless to add that, at this stage, claims of rarumhood, rarissimumhood, and nonesuchhood should be considered no less tentative than ones of universality.

Too often they rest on half-knowledge (to exaggerate vastly) and credulity.

What is now believed to be rare may prove to be less unusual upon wider and more diligent probing. What has been endorsed as an anomaly may vanish upon critical scrutiny as a figment of someone's imagination.

omnia rara praeclara

LEXICON

- b-ö-f 'ox'¹
- Úwlf² 'wolf', Údheghon³ 'earth', Úkr⁴ 'heart', etc.

... and billions of other sound-meaning matchings of lexemes and also grammemes (though not all)

... but then, aren't such matchings typically family possessions, hence infrequentalia rather than rara/rarissima/singularia, of course allowing for some more or less regular formal and/or semantic intrafamily variation?

where found

French

Proto-IE

cf. Latin *bov(-em)*, whence
Romanian *bou*,
Sardinian *boe/boi*, Italian *bove*
(regional *bo, bèu, vove, voi, goi*,
Friulian *bo*, Ladin *bouv*,
Romansch *bov*,
Occitan *buòu/buèu*, French *bœ*
Catalan *bou*, Spanish *buey*,
Portuguese *boi*, ...
that is, everywhere in the
Romance or indeed Italic family

LEXICON: BASIC TERMS

- a basic colour term for TURQUOISE²
- two basic colour terms for BLUE and LIGHT BLUE and none for PURPLE³
- two basic colour terms for RED1 and RED2⁴

Tsakhur (Nakh-Daghestanian)
Russian (Slavonic, IE)
Russian up to early 19th century
Hungarian (Ugric, Uralic)

- a sizable inventory of basic smell terms, i.e., one with more than two or three items⁵

Des(s)ana (Tucanoan, Andean-Equatorial);
Totonaca (Penutian);
Wanzi, Nzebi, Sangu, Tsogo, Fang (all North Western area, Bantu Proper, Niger-Congo);
Ser(r)er Ndut (West Atlantic, Niger-Congo);
Kapsiki (aka Higi; Biu-Mandara Chadic, Afroasiatic)

Or do suchlike finds belong with MORPHOLOGICAL rara/rarissima/nonesuches, since basicness crucially includes being morphologically simple or at any rate non-compositional?

LEXICON: SOUND – MEANING MATCHING

- place of articulation of vowels exploited for iconic expression of size (e.g., for diminutives), but with [high, front] expressing large size and [low, back] small size⁶

Korean (isolate);
Bahnar (Mon-Khmer, Austro-Asiatic)

PHONOLOGY: PHONEME INVENTORY

- labial flap /vɸ/⁷

once believed to be non-existent or very rare (and indeed lacking an IPA symbol), but now known to be rather common, attested in over 60 lgs from Africa (Chadic, Niger-Congo, Nilo-Saharan) and Asia (Austronesian)
- phonemic secondary “labio-palatalization” (simultaneous labial rounding and palatal constriction)⁸

Twɪ (Kwa, Niger-Congo)
French (Romance, IE)
- bilabial trill /ɸ/⁹
 - (i) prenasalized bilabial stop with trilled release
 - (ii) trilled fricated vowel as syllabic variant of labial fricative

Kele, Nias (West Indonesian Austronesian);
Luquan Yi (Tibeto-Burman, Sino-Tibetan)
- laminal (tongue-blade) trill /rɸ¹⁰

Czech (Slavonic, IE)
- voiceless apico-dental plosive followed by a voiceless bilabial trill /tɸ¹¹, behaving as a single segment¹¹

about a dozen words each in Wari’ and Oro Win (Chapakura; Arawakan, Andean-Equatorial)

- apical trills at three places of articulation (fronted alveolar, alveolar, retroflex, /r̥r̥ʁ̥/), both without and with palatalization¹²
 - velar /V/ as the sole fricative (if this is what it is), rather than the expected /s/¹³
 - eight or nine liquids, with three rhotics and five or six laterals¹⁴
 - the full gamut of (voiceless) palato-alveolar, alveolar, bilabial, and (dialectally) velar affricates /tʃ, tʃ, pʃ, kʃ/
 - four-way phonemic contrast of dental, alveolar, palato-alveolar, and retroflex stops¹⁵
 - lack of bilabial nasal and possibly bilabial stops¹⁶
 - phonemic contrast between diphthong and vowel followed by glide¹⁷
(e.g., [kʰ] [kʰj] ‘billiard cue’ vs. [kʰ] [kʰj] ‘cow.PLURAL’)
- Toda (Dravidian)
- Iwaidjian (non-Pama Nyungan, Australian)
- Iwaidjian (non-Pama Nyungan, Australian)
- German (Germanic, IE)
- Nunggubuyu (nonPN, Australian)
- Iroquois; Tlingit (Na-Dene)
- Maastricht Dutch (Germanic, IE)

PHONOLOGY: QUANTITY

- three-way quantity contrast of both vowels and consonants¹⁸

Estonian (?), Livonian (?),
Saami (?) (all Uralic)

PHONOLOGY: FOOT STRUCTURE

- Resolved Moraic Trochee as foot type¹⁹
(esp. older) Germanic (IE);
Cayuvava (Andean-Equatorial) (
- a ternary constraint on prosodic word size:
minimal word must be a trimoraic foot²⁰
Gilbertese (Micronesian)
- trimoraic and quadrimoraic syllables as a normal
syllable type in the native vocabulary, with moras as
prosody-bearing units (not just units of syllabic weight)²¹
Gilbertese (Micronesian)
- Foot Harmony: if at least one foot of a word constitutes
a canonical / ˘ / iamb, then all feet of the word are
made iambic; otherwise all feet are made trochaic²²
Yidiny (Pama-Nyungan,
Australian)

PHONOLOGY: WORD-STRESS

- word-stress on postinitial syllable (or mora)²³
- word-stress on post-postinitial syllable²⁴
- word-stress (in non-focussed or naturally focussed phrases) phonetically implemented by low rather than high pitch²⁵

less than 3% of languages, including Araucanian and a few Uto-Aztec languages (though reconstructed for Proto-Uto-Aztec!)

Winnebago (Siouan)

Saipan and Guam dialects of Chamorro (Austronesian; the Rota dialect has high-pitch stress; Diegueño (Yuman; as a less common alternant to high-pitch stress)

[is perhaps not really so uncommon: also found in northern dialects of British English, Schwyzertüütsch, Bengali (IE), ...]

MORPHOLOGY: WORD FORMATION

- a derivational rule for forming nouns denoting plants from nouns denoting their fruits, and also nouns denoting containers from nouns denoting their characteristic contents
(e.g., French *pomm-ier* ‘apple tree’, *frambois-ier* ‘raspberry bush’, *vinaigr-ier* ‘vinegar flask’; Dutch *kastanje-laar* ‘chestnut tree’²⁶)

French (Romance, IE),
Dutch (esp. Flemish;
Germanic, IE)

- a derivational rule for forming nouns denoting a disease, an inflammation, or an affection of a body part from nouns denoting that body part
(e.g., English *appendic-itis*, *bronch-itis*, *tonsil-itis*; also used metaphorically, as in *election-itis*.
Originally, in Greek, the suffix was used for deriving adjectives from nouns, denoting properties prototypically associated with those nouns; the medical specialization being due to the combination of such derived adjectives, in their feminine singular form, with the feminine noun *nósos* ‘illness’)

Ancient Greek (Hellenic, IE)
and European lgs borrowing
this derivational rule or indeed
the derived and basic nouns
themselves

MORPHOLOGY: NUMERAL SYSTEMS

- ‘four‘ as a regular base for a numeral system²⁷

Nyali (aka Huku, Bantu,
Niger-Congo),
Afudu (affiliation?, Africa);
Hawaiian (originally;
Polynesian, Austronesian);
Welsh (Celtic, IE) (?)

MORPHOLOGY: INFLECTION: TYPES OF EXPONENTS AND THEIR USES

- bound morphology (almost) exclusively prefixal²⁸ Athabaskan lgs
[a family possession?]
- verb-final and postpositions Athabaskan lgs
but bound morphology (almost) exclusively prefixal²⁹ [a family possession?]

- use of metathesis (CV ~ VC) for grammatical purposes, viz. to express distinctions of aspect and/or definiteness, or word-class (noun-verb)³⁰

Rotuman (Austronesian);
Clallam and other Straits lgs
of Coast Salish;
Sierra Miwok lgs and
Costanoan lgs (Penutian)

- case distinction exclusively by tone³¹

Eastern (or Para-) Nilotic
(Nilo-Saharan);
(some) Cushitic,
(some) Omotic (?) (Afroasiatic);
(some) Bantu, Bantoid, Kwa
(Niger-Congo);
Burmese (Tibeto-Burman,
Sino-Tibetan) (?)

- reduplication of verb to form finite verb of restrictive relative clauses with non-actor head nouns (not an attributive deverbal adjective or noun, for which purpose reduplication is used more commonly, not to mention its even more popular uses)³²

Wetan (Eastern Indonesian,
Austronesian)

- plural marking of a small subset of nouns (persons, kin) by lengthening of the vowel of the antepenultimate syllable, not a salient structural position otherwise (e.g., stress is on long vowels, otherwise on the penult);³³ e.g., Hawaiian *ka.NA.ka* ‘person’ – *KAA.NA.ka* ‘people’ (with two stresses in plural, main on penult and secondary on lengthened antepenult)

Hawaiian, Samoan
(Polynesian, Austronesian)

MORPHOLOGY: INFLECTION: ZERO

- negation expressed negatively, by omission of material present in affirmative clause (marked realis vs. unmarked irrealis mood inflection)³⁴
- verb inflection with non-zero exponent for 3rd person, but zero for all other persons
- zero exponence for plural, non-zero for singular (not singulative!) as well as dual³⁵
- inverse number marking: same exponent denoting singular, dual, or plural in different declensions, depending on which number is least expected from the basic meaning of nouns in that declension class (its prototypical members, at any rate)³⁶

Tamil, Kannada, Toda, and elsewhere in South Dravidian; Malakmalak (Pama-Nyungan, Australian) (?); Achumawi (Hokan) (?)

Standard English (Germanic, IE)

Imonda (Trans-New Guinea, “Papuan”)

Kiowa-Tanoan family
[a family possession, hence an infrequentale?]

MORPHOLOGY: INFLECTION: MARKEDNESS: CASE

- absolutive singular (marking intransitive subject) as the morphologically most complex noun form, expressed by a variety of suffixes or also by reduplication of stem³⁷
- absolutive (marking intransitive subject and transitive direct object) as the morphologically marked case vis-à-vis (zero-marked) ergative (marking transitive subject), on NPs which are not focused or topicalized and then are not case-marked at all (Nias), on NPs which are focused (Yukaghir, Komi)³⁸

Chukchi (Chukchi-Kamchatkar

Nias (Malayo-Polynesian, Austronesian);
Yukaghir (isolate);
dialects of Komi (Permic, Finno-Ugric, Uralic);
Old Wardaman (Australian) (?)
Kabardian (Northwest Caucasian) (?)

- **nominative (marking intransitive and transitive subject) as the morphologically marked case vis-à-vis (zero-marked) accusative (marking direct object), for all nouns or for some noun classes³⁹**

**some older Germanic lgs (IE);
Old French (Italic, IE);
Oromo, Dasenech, Kambata,
and perhaps a few further
Cushitic lgs (Afroasiatic);
Zayse and perhaps a few further
Omotic lgs (Afroasiatic);
some Berber lgs (Afroasiatic);
Old Egyptian (pronouns)
(Afroasiatic);
Proto-Afroasiatic (!);
Eastern Nilotic lgs with tonal ca
(Nilo-Saharan);
some Surmic lgs (Nilo-Saharan)
Mojave and most other Yuman
lgs (Hokan);
varieties of Maidu (Penutian);
Wappo (Yukian);
Shokleng (Jê, Tupi, Andean-
Equatorial);
Malakmalak (nonPN,
Australian);
Houailou (Oceanic,
Austronesian)**

MORPHOLOGY: INFLECTION: SUPPLETION

- stem suppletion of an adjectival word (viz., the cardinal numeral 'one') determined by gender⁴⁰ Classical Greek (Hellenic, IE)
- stem suppletion of a noun ('man, person') essentially determined by number, with one stem used for all singular cases and another stem for all plural cases, but with both stems used for the dual, distributed by cases (most taking the plural stem, but a few the singular stem)⁴¹ Slovene (Slavonic, IE)
- number suppletion of the indefinite pronoun, with the singular form identical to the numeral 'one' and (here comes the nonesuch) the plural (and also mass) form supplied by the plural of an interrogative pronoun Standard German (Germanic, IE)
(*einer* 'someone' – *welche* 'some')⁴²

MORPHOLOGY: INFLECTION: CUMULATION

- definite article co-expressing tense (past vs. non-past)⁴³
- internally unanalysable independent personal pronoun expressing the meaning '1st person singular non-topic acting on 2nd person singular topic'⁴⁴

Chamicuro (Maipuran Arawakan)

Tagalog (Philippine Austronesian)

MORPHOLOGY: INFLECTION: LOCUS

- tense not only marked on verbs, but simultaneously on other words too, including nouns (tense agreement?)⁴⁵
- nouns requiring marking for person
e.g., Elamite
sunki-k; sunki-t; sunki-r
'I, the king; you, the king; (he) the king'⁴⁶
- use of an oblique case on a non-finite verb form to express evidentiality (hearsay, inferential)⁴⁷

Kayardild, Lardil (Australian)

Elamite (extinct isolate, Ancient Near East); Nama (Khoisan)

Estonian and South Estonian dialects (Uralic);
Latvian (Baltic, IE);
Kayardild ((non?)PN, Australian)

MORPHOLOGY: INFLECTION: COMPLEXITY (PARADIGMATIC, SYNTAGMATIC)

- words taking as many as four cases simultaneously⁴⁸ Kayardild ((non?)PN, Australian)
- an extra-large case inventory, of above 20 cases⁴⁹ Hungarian (Ugric, Uralic) (?)
(variously analysed as having 16, 18, 21, 23, or 27 cases)
- an extra-large inventory of pronominal words, of some 200⁵⁰ Rennell (Futunic, Polynesian, Malayo-Polynesian, Austronesian)
- lack of even a minimal binary contrast (proximal/distal) in basic local adverbial deictics Bavarian (Germanic, IE)

MORPHOLOGY: INFLECTION: CATEGORIES AND THEIR DOMAINS: GENDER, RESPEC

- a divine gender (for nouns for gods, saints, comets, and other supernatural referents), contrasting with masculine, feminine, animal, and inanimate genders⁵¹

Jicaltepec Mixtec (Oto-Manguen)
- gender contrast of nouns, but not of pronouns (except interrogative ones, in Teop)⁵²

Turkana (Nilotic, Nilo-Saharan)
Teop (Bougainville, Austronesian)
- gender contrast (MASC/FEM) in personal pronouns for 2nd, 3rd, and sometimes also 1st plural and for 2nd singular but not for 1st and 3rd singular, in some pronominal paradigms⁵³

Tāhāggart, Tuareg, Kabyle,
Proto-Berber (?!)
(Berber, Afroasiatic)
- gender contrast in personal pronouns exclusively in 1st person singular⁵⁴

Tocharian A (IE);
Thai (Daic, Austroasiatic)
- gender contrast (MASC/FEM) in personal pronouns exclusively in 1st person plural, both inclusive and exclusive⁵⁵

Shirumba (Heiban, Kordofanian)

- a mixed (aka “common”, “resolution”, “indefinite”) gender in 3rd person (probably also 2nd and 1st person) non-singular pronouns for reference to mixed sets of antecedents, which naturally is without a singular counterpart (illustrated schematically, based on French: ‘il et il ... ils; elle et elle ... elles; il et elle ... ulles’; real French uses a more common resolution strategy, viz. to give one gender priority over the other(s) in cases of conflict, with the most common strategy of course being not to distinguish genders in non-singulars in the first place)⁵⁶
- contrast in pronouns for 2nd and 3rd person for sex of the referent and also of the speaker⁵⁷
- contrast in pronouns for 3rd person for number (SG /PL), gender (both of the referent and of the speaker), kin relationship between speaker and referent (and if affine, then also whether in ascending or descending generation), and ethnic identity (Yuchi vs. other)⁵⁸
- contrast in personal pronouns for moyety, kin class, and generation level membership⁵⁹

Nama (Khoesan);
 Vanimó (Sko, “Papuan”);
 Bukiyip (Torricelli, “Papuan”)

Diuxi Mixtec (Oto-Manguean)

Yuchi (isolate, east central
 Oklahoma, USA)

Adnyamadhanha and other
 Australian lgs

- choice of 3rd person personal pronoun determined by gender (MASC/FEM/NEUT, contrasting in SG only and not corresponding to animacy distinctions in any consistent manner), except in adpositional phrases (with adpositions governing the dative or genitive), where the relevant contrast rather is one of animacy or related notions, with masculine and feminine pronouns now only referring to animate (or concrete, figure, salient) referents and with the neuter pronoun (both in singular and plural) replaced by local-adverbial forms (*da*, *wo*), preceding rather than following their adpositions, for inanimate (abstract, ground, not-so-salient) referents

e.g., *Er kauft ihn/sie/es*
he buys him/her/it (MASC/FEM/NEUT.ACC)
Er träumt von ihm/ihr
he dreams of him/her (MASC/FEM.DAT)
Er träumt da-von
he dreams there-of ('of it')

- inflection of the great majority of words—nominal, pronominal, verbal—for sex of the speaker (and, in the case of male forms, also of the addressee, with males using female forms when addressing females), with female forms formed from male forms by final syllable reduction or male forms formed from female forms by adding a syllable (whose form depends on the class of word so inflected)⁶⁰
- pronoun of most respectful address recruited from pronoun of identity (*(allerhöchst)dieselben*), rather than from indefinite or least definite pronoun⁶¹

Yana (Hokan)

18th century German
(Germanic, IE)

MORPHOLOGY: INFLECTION: CATEGORIES AND THEIR DOMAINS: POSSESSION

- a three-way distinction of possession for possessive pronouns, which also distinguish inclusive and exclusive:
alienable, inalienable, family possession⁶²

Khinalug (Lezgetic, NE Caucasian)

MORPHOLOGY: INFLECTION: CATEGORIES AND THEIR DOMAINS: PERSON, NUMBE

- inclusive/exclusive contrast for 2nd person ('you several, non-SAPs included/excluded')⁶³

SE Ambrym (Austronesian),
Port Sandwich (Austronesian) (?
Yokuts (Yokutsan);
Ojibwe, Cree, Fox, Potawatomi,
Naskapi (Algonquian) (?);
Otomi (Otomanguean);
Abkhaz (East Caucasian) (?);
Ghomala' (Grassfield Bantu);
Nama (Khoesan);
Ordoss (Eastern Mongolian,
Altaic);
Bavarian (IE) (?)

- dual with 1st person exclusive but not with 1st person inclusive⁶⁴

Yagua (Ge-Pano-Carib);
Tübatulabal (Uto-Aztecan) (?);
Savosavo (East Papuan) (?)
Ngankikurrungkurr (nonPN,
Australian) (?)

- a three-way distinction among personal pronouns between ‘singular speaker’, ‘singular non-speaker’ (i.e. addressee or other), ‘plural’ (any speech-act role or combination of roles)⁶⁵

Balante (West Atlantic, Niger-Congo)

- an ambal (aka “paral”, only referring to natural pairs) in addition to a dual (referring to any two referents)⁶⁶

Tocharian A and B (IE)
(especially B, since in A the dual itself is obsolescent)
- a quadral number with personal pronouns⁶⁷

Sursurunga, Tangga (both Patpatar-Tolai, Austronesian) (?)
- independent personal pronouns for 1st and 3rd (animate) person inflecting for both number and case, but that for 2nd person inflecting for neither category (defectiveness of 2nd person pronouns in number alone being more common)

Modern Standard English
(Germanic, IE)
- no number inflection at all on any personal pronouns⁶⁸

Mura Pirahã (isolate, Macro-Chibchan);
Kawi (Old Javanese, Malayo-Polynesian, Austronesian)
- number marking (for plural and dual) on personal pronouns (suppletive) as well as on expressives (or ideophones; by suffix, which is unusual in Mon-Khmer, a family otherwise preferring prefixes and infixes), but not on any other word class (including nouns and verbs)⁶⁹

Bahnar (or Bahnaric in general)
(Mon-Khmer, Austro-Asiatic)

- duals up and down the referential hierarchy (ignoring INCL/EXCL)⁷⁰

PRONOUNS			NOUNS
1ST	2ND	3RD	

+DU----->

+DU-----> -DU

+DU-----> -DU----->

+DU -DU----->

+DU----->

-DU +DU----->

-DU-----> +DU----->

-DU-----> +DU

+DU-----> -DU +DU

+DU -DU +DU----->

-DU-----> +DU -DU

-DU +DU -DU----->

+DU -DU-----> +DU

-DU +DU-----> -DU

+DU -DU +DU -DU

-DU +DU -DU +DU

in abundance

65+ languages

3+ languages

9+ languages

over half of all (existing) lgs

Cl. Arabic (Semitic, Afroasiatic) -

Ugaritic (Semitic, Afroasiatic) +

10+ languages

Kewa (Papuan) (?)

northern dialects of Ostyak (Ura

4+ languages

in no (attested) language (?)

in no (attested) language (?)

Rawa (Trans-NG, "Papuan") (?)

Sedang (Austroasiatic)

in no (attested) language (?)

SYNTAX: WORD INTEGRITY

- endoclitics occurring not only inside morphologically complex words, but even within monomorphemic words,

with the positioning of the clitic

- (a) regulated by one general phonological rule
(e.g., *Chi=fuckin=CAgo*, *fan=bloomin=TAStic*;
but are *fuckin'*, *bloomin'*, etc. CLITICS?)

English (Germanic, IE) (?)

For one thing, they are categorially rather selective,
combining only with nouns)

- (b) not regulated by a general phonological rule,
although when endoclititic to simple verbs,
the clitic always comes before the last segment
of the verb stem⁷¹

Udi (NE Caucasian)

(subject person-number markers are cliticized
to constituents in focus, being always enclitic with
arguments in focus, but en- or endo-clitic with
verbs in focus)

SYNTAX: WORD CLASSES

- kinship verbs, i.e., pure kinship relations normally expressed by canonical transitive verbs rather than by nouns (which may also exist, though primarily for purposes of address)⁷²

Iwaidjian lgs (Australian);
Central Guerrero Nahuatl
(Aztec, Uto-Aztecan);
(some lgs of) some North
American Indian families
[possibly an infrequentale]

- interrogative pro verbs⁷³
(schematically, ‘The dog WHATed the boy?’ – ‘It bit him’,
‘The dog HOWed? – ‘It howled like this’)

Southern Paiute (Uto-Aztecan);
Chukchi (Chukchi-Kamchatkan);
Mandarin (Chinese, Sino-Tibetan);
Yankunytjatjara (a.k.a. Western
Desert (PN), Kayardild ((non?)
PN), and many other Australian
lgs;
Hua (‘Papuan’);
Tagalog (Philippine),
colloquial varieties of Malay/
Indonesian,
Tahitian (Polynesian),
probably widespread throughout
Austronesian;

- adverb as a distinct word class but not adjective⁷⁴

Kinyarwanda (Bantu, Niger-Congo)

**Kambera (Malayo-Polynesian);
Kwaza (unclassified, S. Rondônia;
Brazil)**

SYNTAX: WORD ORDER

- object before subject in basic word order⁷⁵
- expression of numerical approximation ('about 5') by inversion of word order from Numeral + Noun to N + Num



Russian (and other Slavonic?)
(Slavonic, IE)

SYNTAX: WORD ORDER: V2

- (finite) verb-second word order in main declarative clauses (i.e., specifically VERB-second rather than more generally clitic-second or X-second)⁷⁶
- (finite) verb-second word order in main declarative clauses generally optional, but obligatory if the first constituent is the unmarked narrative consecutive adverb ‘then’ (*þa*, homonymous with the temporal conjunction ‘when’ introducing subordinate clauses, whose finite verb generally is final and never second)
- (finite) verb-second word order in main declarative clauses only if the first constituent is an adverbial with strong negative force (such as *ever before*, *hardly ever*)

many Germanic languages (IE), excluding English (ignoring a few marginal traces of V2); Sorbian (Slavonic, IE, influenced by German); Old French and some other early Romance lgs (IE) (???), Rhaeto-Romance (Romance, IE) Breton, Middle Welsh (Celtic, IE) Kashmiri (Dardic, Indic or Iranian, IE);

Old English (Germanic, IE)

Modern English (Germanic, IE)

SYNTAX: PRONOMINALIZATION: ZERO

- parasitic gaps (rather than an overt resumptive pronouns), licensed in main clauses if and only if a coreferential NP in a topicalized, fronted subordinate clause has been moved ahead of the clause-initial complementizer;

e.g. Bavarian:

an Sepp wenn i Ø dawisch daschlog i Ø
the Sepp_i if I get Ø_i I slay Ø_i
*wenn i an Sepp dawisch daschlog i eam/*Ø*
if I get the Sepp_i I slay him_i/*Ø_i
'if I get Sepp I'll slay him'⁷⁷

Bavarian, West Flemish
(Germanic, IE),
Polish (Slavonic, IE) (?),
Latin (Italic, IE) (?)

SYNTAX: AGREEMENT

- some noun-phrase internal agreement but no agreement at clause level⁷⁸

(some) Pama-Nyungan (Australian);
modern Continental Scandinavian Germanic (IE)
- relative pronoun as the only target for agreement in animacy (human: *who*; non-human: *which*)

Modern Standard English (Germanic, IE)
- attributive adjectives agreeing (in gender, number, case) when they precede their noun but not when they follow it (using the invariable stem here)⁷⁹

German (Germanic, IE)
- determiners and modifiers agreeing in all relevant agreement categories (gender, number, case) with only one of two or more conjoined nouns sharing these determiners and modifiers (rather than with both, relying on some resolution strategy in case of conflicting agreement categories), the agreed-with noun being the last (rather than the first) as well as the most distant conjunct (with nouns coming after determiners and modifiers)⁸⁰

Low German (Germanic, IE)

- Suffixaufnahme in more or less prototypical form, i.e., with a noun in an attributive case (genitive and/or other) agreeing in case (and perhaps further inflectional categories) with its head noun⁸¹

some Australian lgs
 (indeed about half of them);
 Awngi (Central Cushitic), Burji,
 Darasa (Highland East Cushitic)
 (Cushitic, Afroasiatic);
 Maasai (and perhaps other
 Paraniotic lgs, Nilo-Saharan or
 Afroasiatic);
 Kikongo group (Bantu, Niger-
 Congo);
 many or indeed most modern
 Indo-Aryan lgs (except all of
 Magadhan and Singhalese-
 Maldivian)(IE);
 Malto (alone in Dravidian);
 Old Georgian and perhaps
 further old Kartvelian lgs;
 Tsezic group (NE Caucasian);
 Chukchi and perhaps other
 Chukchi-Kamchatkan lgs;
 Evenki, Even (Tungus, Altaic);
 Hurrian and Urartean [i.e., all of
 the long extinct Hurro-Urartean
 family, hence an infrequentale?]

SYNTAX: NUMERALS AND NUMBER

- plural marking of numeral (including ‘one’) to express that the reference of the NP is to groups of entities denoted by the accompanying noun (e.g., one-SG shoe-SG ‘one shoe’ – one-PL shoe-PL ‘one group (pair) of shoes’; five-SG shoe-SG ‘five (individual) shoes’ – five-PL shoe-PL ‘five groups (pairs) of shoes’)⁸²

Finnish, Estonian (Uralic);
Icelandic (Germanic, IE) (?)

SYNTAX: GRAMMATICAL RELATIONS: ALIGNMENT

- ergative (or neutral) case marking higher and accusative case marking lower on the referential hierarchy:⁸³

PRONOUNS			NOUNS	
1ST	2ND	3RD		
ERG----->			ACC	Parachi and Yazguljami, past tense (Iranian, IE)
ERG----->	ACC	NEUT		Iha, SG, past tense (West Bomberai, "Papuan")
ERG----->	NEUT	ACC		Tirahi, past tense (Dardic, IE)
NEUT----->			ACC	Nganasan (Samoyedic, Uralic)
ERG----->	ACC	ERG		Karitiana (Tupi-Guarani)

- accusative alignment for proper names and kin terms (as used for address), tripartite alignment for singular pronouns, neutral alignment for non-singular pronouns and plural common nouns, ergative alignment for non-plural common nouns⁸⁴
- double-oblique alignment, with one case (or other coding device) for both subject and object in transitive clauses and another case (or other coding device) for intransitive subject⁸⁵
- 3rd person independent pronouns as the only NPs to inflectionally distinguish subject and object

Kalaw Lagaw La (Saibai dialect)
(Pama-Nyungan, Australian)

Rushan and a few other Pamir lg
(Iranian, IE) as well as a few
neighbouring Dardic lgs (Indic
or Iranian, IE)

Washo (Hokan)

SYNTAX: GRAMMATICAL RELATIONS: OBJECTS

- two object arguments in one construction showing the same syntactic and morphological behaviour as the single object of a primary transitive verb (lgs known as “symmetrical” or “true double accusative”);
- a benefactive applied object in a multiple objects construction behaving like the single object of a primary transitive verb, esp. with regard to extraction⁸⁶

(not all) Bantu lgs (Niger-Congo)
Yagua (isolate, Peru);
Martuthunira (PN, Australian)

Bajau (Malayo-Polynesian,
Austronesian);
LuHaya (Bantu, Niger-Congo)

SYNTAX: GRAMMATICAL RELATIONS: VOICE

- passive constructions, with subject demoted, corresponding to active ones containing a reflexive pronoun (or reflexive marker) as a direct or indirect object, with the reflexive thus not controlled by a subject in the passive
- grammaticalization of these verbs, denoting activities with subjects in an agentive role, as auxiliaries of passives where non-agents are promoted to subject:
 - (i) 'give'
(should actually be more common since 'give' is common as an existential verb and existential verbs are common as passive auxiliaries)⁸⁷
 - (ii) 'eat'
(but then, the agentive subject of 'eat' is also a receiving one, cf. 'to take in' or 'to stomach')⁸⁸

German (Germanic, IE),
Lithuanian (Balto-Slavic, IE),
Latin (Italic, IE);
Turkish (Turkic, Altaic)

Luxemburgian German
(Germanic, IE)

(some) Munda (Austroasiatic);
Sinhala, Bengali (?) (Indo-
Aryan, IE)

SEMANTICS

- verbs of cognition (especially ‘to think, know, remember’ —more precisely, non-communicative cognition, as opposed to cognition based on communication, such as ‘to understand’) polysemous or heterosemous with, and historically deriving from, verbs of hearing rather than of seeing⁸⁹
- the two meanings ‘sun’ and ‘future tense’ expressed by the same form (because the future is grammaticalized from the adverb *sina* ‘tomorrow’, which in turn derives from *si-na* ‘sun-come’)⁹⁰




widespread among Australian lg
[an infrequentale?—No:
Kurna and Arrernte have
‘see’ > ‘know, think’];
Hausa (Chadic, Afroasiatic) (?);
Ommura (“Papuan”) (?);
Suya (Ge, Ge-Pano-Carib) (?)

Mandinka (Mande, Niger-
Congo)

... and lots of combinatory rara/rarissima/nonesuches, i.e., unique combinations of traits which individually couldn't be more ordinary (e.g., having all of these properties: syllable-final devoicing, four cases cumulated with two numbers, non-concatenative tense marking on verbs, definite and indefinite article, relative pronouns identical to either demonstrative or interrogative pronouns, and something arguably rare, namely strict verb-second in declarative main clauses)

—and, naturally, the more numerous the traits-in-combination you take into account the nonesucher.

1 Saussure 1985: I, 1, §1.
2 Davies, Sosenskaja, & Corbett 1999.
3 Corbett & Morgan 1988.
4 Wierzbicka?
5 Plank & Plank 1995, Plank 1997, with further references.
6 Kim 1977, Diffloth 1994.
7 Hajek, Demolin, & Olson 2000.
8 De Jong & Obeng 2000.
9 Ladefoged & Maddieson 1996, Dryer 1997.
10 Ladefoged & Maddieson 1996.
11 Ladefoged & Everett 1996.
12 Ladefoged & Maddieson 1996.
13 Evans 2000a.
14 Evans 2000a.
15 Dryer 1997.
16 ? (my LT paper)
17 Gussenhoven 1999.
18 Lehiste 1966.
19 Drescher & Lahiri 1991, Key 1961, Hayes 1995.
20 Blevins & Harrison 1999.
21 Blevins & Harrison 1999.
22 Dixon 1977, Hayes 1995.
23 Hyman 1977, Munro 1977.
24 Miner 1979, 1981, Lockwood 1982.
25 Topping, Chung; Langdon 1977
26 p.c. Ton van der Wouden.
27 Gvozdanović 1999, Justus 1999, *The Linguist List* 6.894, Elbert & Pukui 1979: 161-162.
28 A. A. Kibrik 2000.
29 A. A. Kibrik 2000.
30 Thompson & Thompson 1969, Okrand 1979.
31 Bennett 1974, Plank 1995, Okell (for Burmese, p.c. David Gil).
32 Carpenter 1996.
33 Elbert & Pukui 1979: 106.

34 Master 1946, Pilot-Raichoor 1997, 1998a, 1998b, Forest 1993.
35 Dryer 1997.
36 Wonderly, Gibson, & Kirk 1954, Merrifield 1959, Mithun 1999.
37 Skorik 1961/1977.
38 Donohue & Brown 1999; Irina Nikolaeva (personal communication) for Yukaghir and Komi.
39 Plank 1985a, with further references, Dixon 1994, Gensler 2000.
40 Plank, plenary paper on suppletion, Morphologietagung Wien 1999. 
41 Plank 1994c.
42 Glaser 1992, Plank 1994b.
43 Parker 1999.
44 David Gil, LINGTYP List, 12 Sep 2000. Actually, anomalies in combinations of 1st and 2nd person would seem more common;
see Heath in *Paradigms* and more recently in *Language*.
45 Nick Evans, *Science*, vol. 2888, 19 May 2000, p 1156.
46 mention grammars
47 Bernhard Wälchli, LINGTYP List, 16 Feb 2000.
48 Evans 1995a.
49 Plank 1986.
50 Elbert 1965.
51 Bradley 1970.
52 Dryer 1997, Mosel & Spriggs (undated conference handout).
53 Plank & Schellinger 1997.
54 Van Windekens 1979, Krause & Thomas 1960, Schulze & Siegling 1931; for Thai?
55 Schadeberg 1981: 82-83, who suspects that he got something wrong here, since such misbehaviour would be unique in this family,
with no other Heiban language showing pronominal gender anywhere.
56 Plank & Schellinger 1997: 79-81.
57 Simon 1999.
58 Wagner 1938.
59 Yallop 1977. Adnyamadhanha has ten dual pronouns for distinctions of moyety, kin, and generation level.
60 Sapir 1929.
61 Simon 1999.
62 Kibrik, Kodzasov, & Olovyannikova 1972.
63 Simon 1999, others.
64 Plank 1996 and DuDa (Konstanz Dual Database, maintained by Frans Plank and Wolfgang Schellinger).
65 Wiesemann 1986.

66 Van Windekens 1979, Krause & Thomas 1960, Schulze & Siegling 1931.
67 Hutchisson 1986.
68 Wieseemann 1986 (information from Daniel Everett). For Kawi Humboldt?
69 Diffloth 1994.
70 Plank 1996 and DuDa (Konstanz Dual Database).
71 Harris 2000.
72 Amith & Smith-Stark 1994a, b, Evans 2000b.
73 Ultan 1969: 53; Sadock & Zwicky 1985: 184; Bickel 1991: 96f., 124; Lazard & Peltzer 2000; Nick Evans, p.c.; David Gil, p.c.
74 Klamer 1998, van der Voort 2000.
75 Derbyshire & Pullum 1981.
76 Haiman 1974, Haider & 🍏🍏🍏 1986, Adams 1987, Vance 1997, Willis 1998, Schafer 1995, Borsley & Kathol 2000, 🍏🍏🍏.
77 Bayer 1983, Haegeman
78 Plank 1994a.
79 Plank 1977.
80 Rohdenburg 19🍏🍏, Plank 1991, Corbett 2001.
81 Plank (ed.) 1995; Schellinger 1998 (adding Malto).
82 Hurford 2001.
83 Filimonova 2000.
84 Comrie 1981.
85 Payne 1980.
86 Donohue 1996.
87 Nübling 1999, Plank 1999.
88 Von der Gabelentz 1861, Plank 1999.
89 Sweetser 1990, Evans & Wilkins 2000.
90 Claudi 1994.