# Progress in Typology: The first 369 years, and those yet to look forward to 

Frans Plank<br>Universität Konstanz

## 1. Doubting Fra Tommaso: Which words to inflect, and for which categories?

Giovanni Domenico (Fra Tommaso) Campanella (1568-1639)
education and profession: Dominican monk from early youth, theologian, philosopher, astrologer, utopian social theorist (La città del sole, 1602); under arrest or imprisoned for about 30 years (for religious and political reasons, not for being an avantgarde typologist).

Relevant text: first part of his Philosophia rationalis (1638), devoted to grammar.
Special focus on the parts of speech and their universal definition: Campanella finds more crosslinguistic diversity than envisaged by BLT, calling for an ALT:
nouns are not (to be defined as) words inflecting for case and number (witness Romance vernaculars, Hebrew, Arabic [!], with nouns only inflecting for number), nor as words either inflecting for case or accompanied by (p)articuli (witness Ancient Greek, which has both);
verbs are not (to be defined as) words inflecting for tense, person-number, and possibly further categories (witness Chinese and Vietnamese).

Crosslinguistic awareness (i.e., UG no longer entirely speculative and Greek/Latin-inspired): Hebrew (and Chaldean), Arabic, Turkish, Ancient Greek, Latin, the contemporary Romance vernaculars (Italian, French, Spanish), Chinese, and Vietnamese (lingua Coconchinesium or Concincinorum).

Discovery 1: The inflectional behaviours of nouns and verbs are not independent of one another.
Case inflection and verbal inflection, according to Campanella:

| NOMINAL CASE <br> INFLECTION | VERBAL <br> INFLECTION | combination attested in |
| :---: | :---: | :--- |
| + | + | Latin, Greek, Turkish |
| + | - | --- |
| - | + | Italian, French, Spanish, Hebrew, |
| - | - | Arabic [!] |
|  | Chinese, Vietnamese |  |

- If nouns inflect for case, then verbs inflect too, for whatever category (tense, person-number, ...).
[In 1965 contradicted by Arthur Capell, typology of "concept domination".]
Why? ???

Discovery 2: Two inflectional categories, case and number, of one part of speech, noun, are not independent of one another.

Case and number inflection, according to Campanella:

| CASE <br> INFLECTION | NUMBER <br> INFLECTION | combination attested in |
| :--- | :---: | :--- |
| + | + | Latin, Greek, Turkish |
| + | - | -- |
| - | + | Italian, French, Spanish, Hebrew, Arabic [!] |
| - | - | Chinese, Vietnamese |

- If nouns inflect for case, then they also inflect for number.

Why? ???

## 2. Mesgnien-Meninski compares Turkish and finds word orders harmonic

François (de) Mesgnien/Meninski (1620/23-98)
education and profession: language teacher (author of grammaticae civiles of French, Italian, Polish), traveller, diplomat in Polish and Austrian service, long based in Constantinople.

Relevant text: Thesaurus linguarum orientalium, accompanied by a contrastive grammar, Linguarum orientalium turcica, arabica, persica institutiones seu Grammatica turcica (Vienna 1680); especially part 6 of the Grammatica, De syntaxi, section 2, De ordine constructionis (1680: 146-148). [First edition largely destroyed by fire at siege of Vienna in 1683; re-edited later, 1756, 1780-82, too late for next generation of word order typologists.]

Crosslinguistic awareness:
Turkish, Arabic, and Persian, also Greek, Latin, French, Italian, German, Polish, Hungarian, and some others mentioned in passing.

Mesgnien finds a lot more diversity than had been envisaged by BLT: definite articles (not found in Turkish and Persian), prepositions (postposed in Turkish), personal and possessive pronouns (the former only optional in some languages, the latter suffixed to nouns in Turkish), genders (no such contrast in Turkish and Persian), most cases (ranging from really only a single one in Persian, contrasting with a multi-purpose form in $r a$, to three in Arabic and six in Turkish, with the nominative here coinciding with a special accusativus indeterminativus), some numbers (with a dual only in Arabic, and with the nominal plural identical to the verbal plural in Turkish), and negation (part of verbal inflection in Turkish); inflection agglutinative in Turkish (separation rather than cumulation).

Highlight: linear order of constituents in a whole range of constructions, in particular these:
(i) subject, object, and VERB
(ii) nominal attribute and HEAD NOUN
(iii) adjectival attribute and HEAD NOUN
(iv) noun phrase and ADPOSITION
(v) clause and CLAUSAL PARTICLES (such as conjunctions and interrogative words)
(vi) subject and verb in interrogative clauses
(vii) indirect object, direct object, and subject
(viii) core arguments and circumstantial specifications

Discovery: Linear order does not vary independently across a range of constructions, especially ones instantiating the relation of dependency/government.

Generalisation(s):

1. Regens debet semper postponi suo recto, seu casui quem regit, ideoque Verbum, quòd omnia regere videatur, ultimum orationis locum obtinet [in one group of languages, including Turkish, also Hungarian, German]. ... Sed horum ferè omnium contrarium evenit [in another group of languages, including Arabic, and less strictly Persian].
2. Substantivum Adjectivo suo postponitur, ut \& alteri Substantivo quod regit in genitivo ... Sed horum ferè omnium contrarium evenit in lingua Arabica ... (1680: 147)
[1. The governor must always be placed after its governee, or the case which it governs, and likewise the verb, which seems to govern everything, comes last in its clause. 2. The substantive is placed after its adjective, and also after another substantive in the genitive governed by it ... But the opposite of nearly all this obtains in Arabic ...]

Why such cross-categorial harmony? Simplicity: Easy to learn, process, produce (?)
Why one way (Head-Dependent) or the other (Dep-Head)? ??? [a basic choice, not
dependent on anything else?]

## 3. Beauzée of the Encyclopédie: Progress?

Nicolas Beauzée (1717-89)
education and profession: educated at Jesuit college, then unemployed or school teacher, sometime professor of Grammar at École Royale Militaire, demoted to language teacher, then received into the Académie Française, editor, writer, and translater (from Latin).

Relevant texts: articles in Diderot \& d'Alembert's Encyclopédie (vol. 9, 1765), especially 'Langue', and, with little difference, Grammaire générale, ou Exposition raisonnée des éléments nécessaires du langage (Paris 1767)

Crosslinguistic awareness:
such staple fare as French, Italian, Spanish, English, German, Latin, Ancient Greek, and Hebrew, but also a medley of more unusual languages, such as Portuguese, Swedish, Breton, Irish, Polish, Basque, Saami, Arabic, Aramaic (in different varieties), Chinese, and (Peruvian) Quechua

Conception of Universal Grammar:
envisaged as a fund from which particular grammars may choose, rather than as the largest common denominator of all particular grammars.

Building on Gabriel Girard's Les vrais principes de la langue françoise (1747, propagating three génies: analogue, transpositif, mixte/amphilogique), Beauzée's focus is on the expression of grammatical relations, especially through the linear order of constituents though not only SVO, but in a whole range of constructions, in particular these, all involving the relationship of détermination:
\(\left.$$
\begin{array}{ll}\text { DETERMINED } & \begin{array}{l}\text { DETERMINING CONSTITUENTS } \\
\text { subject }\end{array}
$$ <br>

predicate\end{array}\right]\)| complement (objects, prepositional phrases) |
| :--- |
| preposition | | complement (noun phrase) |
| :--- |
| head noun |
| head noun |
| head noun | | relative clause |
| :--- |
| complement (prepositional phrase) |

- In analogous (=rigid-order) languages the rigid linear order in all these constructions is uniform, with the determined constituent always preceding the determining one.

Why not also determining before determined?
Scholastic and later doctrine about the ordo naturalis: the analytic order of ideas is governed by the principles that the cause comes perforce before the effect, the action perforce before whatever or whoever is acted upon, and that prius est esse quam sic esse - and these analytic orders are only mirrored by linear constituent orders with determined constituents (subject, verb, preposition, head noun) preceding rather than following determining ones (predicate, complements, attributes).

- And attributive adjectives in English and Swedish?
- And Celtic, Semitic (VSO)?
- And too bad Turkish was not in Beauzée's sample!

Beauzée's taxonomy of génies:

|  | ANALOGOUS |  | TRANSPOSITIVE |  |
| :--- | :--- | :--- | :--- | :--- |
|  | UNIFORM | FREE | UNIFORM | FREE |
|  | rigid | rigid, but <br> inversions | free within <br> limits | free |
| ORDER | absent | absent | present | present |

- If constituent order is free without limits, then nominal words will be inflected, but not vice versa (because of the uniform transpositive génie);
- if constituent order is rigid without inversions, then there will be no nominal inflection, but again not vice versa (because of the free analogous génie).

Why? Simplicity/economy, iconicity.
Relationship to diachrony: Filiation is contingent on ressemblance du génie, but no longer (as with Girard, who denied on typological grounds that French was derived from Latin) absolute immutability of génie (just like fixity of species had begun to be questioned): possible (gradual) change from (simpler, more iconic) analogous to transpositive (again ruling out a derivation of French from Latin).

## 4. Smith and Anonymous of the Encyclopedia Britannica: The economy of inflection

Adam Smith (1723-90)
education and profession: tutor, then professor of Logic and Rhetoric, then of Moral Philosophy. Relevant text: 'Considerations concerning the first formation of languages, and the different genius of original and compounded languages' (1761/1767, often reprinted and translated).

Anonymous $=$ William Smellie? $(1740-95$, a hack writer and first editor of the $E B)$ Relevant text: article 'Language' and others in the Encyclopedia Britannica (1771).

Crosslinguistic awareness: very modest, anecdotal ("conjectural history"); en passant references to Ancient and Modern Greek, Etruscan, Latin, Gothic, the older Germanic tongues of the Lombards, Franks, and Saxons, French, Italian, (Old) Armenian (10 cases!, or so thought Smith and his source), Hebrew, the languages of some savage nations Smith had read of (North American Indian in particular), and above all of English.

Focus on the development of inflectional systems: From lexicon to grammar, by means of abstraction and analysis (excrescence rather than coalescence), from original to compounded languages (=mixed, involving contact, requiring adult L2 learning).

Gender, case, and number in phase of inflectional expansion (analysis of holistic expressions), according to Adam Smith:

| GENDER | CASE | NUMBER | combination conjectured to be |
| :---: | :---: | :---: | :---: |
| - | - | - | possible |
| + | - | - | possible |
| + | + | - | possible |
| + | + | + | possible |
| - | + | + | impossible |
| - | - | + | impossible |
| - | + | - | impossible |
| + | - | + | impossible |

- Number implies case, which in turn implies gender.

Gender, case, and number in phase of inflectional reduction (occasioned by L2 learning in language contact situations), according to Adam Smith:

| GENDER | CASE | NUMBER | combination conjectured to be |
| :---: | :---: | :---: | :---: |
| + | + | + | possible |
| - | + | + | possible |
| - | - | + | possible |
| - | - | - | possible |
| + | - | + | impossible |
| + | + | - | impossible |
| - | + | - | impossible |
| + | - | - | impossible |

- Gender implies case, which in turn implies number.
- Numerals/quantifiers imply prepositions, which in turn imply adjectives - which in turn imply the presence of nouns substantive, and these only exist by virtue of being in contrast to the other class of principal words, (personal) verbs.

Four parameters of accidence expressions, according to Smith and Anonymous:

| TIGHT BOND | CUMULATION | SYNONYMY | HOMONYMY | believed attested |
| :---: | :---: | :---: | :---: | :---: |
| + | + | + | + | yes (=flexion) |
| + | + | + | - | no |
| + | + | - | - | no |
| + | - | - | - | no |
| - | - | - | + | no |
| - | - | + | + | no |
| - | + | + | + | no |
| + | - | - | + | no |
| + | - | - | + | no |
| + | + | - | + | no |
| + | + | - | + | no |
| - | - | - | - | no |
| - | - | - | no |  |
| - | - | - | yes $(=a g g l)$ |  |
| - | - |  | - |  |

Co-variation (=co-evolution) not a "physical necessity" but only a "moral certainty": empirical contingencies, motivated by the need to keep grammars within the bounds of what can be handled with facility (=economy) by human speakers and hearers.

## 5. The nineteenth century

[Skipped for reasons of time. Was the century of Lautgesetze, "laws" of change: concerning laws of language, properly so called, it was eminently forgettable, with the two exceptions to be noted presently.]

## 6. The Humboldt Moratorium

Wilhelm von Humboldt (1767-1835) education and profession: trained in cameralistics; a diplomat and civil servant (secretary of state for education); an amateur philosopher and, late in life, amateur linguist.

Advice: First things first!

- namely an "encyclopedia of languages" (grammars and lexicons, on uniform principles), and an "encyclopedia of categories", re-arranging the information provided in the encyclopedia of languages. ("Categories" are to include constructions, cf. Hans Conon von der Gabelentz's Über das Passivum (1861), demonstrating that it is always forms also used for something else that are being re-used for "passive" constructions.)
'Über den Dualis' (1830), this being the only category covered by WvH, fragmentarily:
- languages can independently have or lack a dual in pronominal and in nominal inflection, with the latter in particular not dependent on the former;
- having a dual for any nouns implies having one for nouns denoting natural pairs;
- having a dual for pronouns of 3rd person implies having one for 2nd person, which in turn implies having one for 1st person (the speaker-hearer dyad being a prototypical natural pair);
- using a dual for arbitrary sets of cardinality 2 implies using a dual for confirmed pairs;
- having a duo-paucal (for any sets of cardinality 2) implies also distinguishing a paucal (for three or a few more) from a plural (for multitudes).


## 7. Gabelentz's two conjunctures

Georg von der Gabelentz (1840-93)
education and profession: trained in law, then a professor of oriental languages (esp. Chinese) and linguistics.

Relevant texts: 'Hypologie der Sprachen, eine neue Aufgabe der Linguistik' (1894), as well as Die Sprachwissenschaft (Leipzig 1891/1901).

Crosslinguistic awareness:
as wide-ranging as the Gabelentz family library, perhaps the most comprehensive at the time.
"Jede Sprache ist ein System, dessen sämmtliche Theile organisch zusammenhängen und zusammenwirken." (1901: 481, essentially a quote from Georges Cuvier, reproducing 'le principe de la corrélation des formes dans les êtres organisés' and 'le principe de la subordination des caractères', applied so successfully in reconstructing fossils; same exaggeration in lingistics regardless of imagery: language a mechanism, an organism, a system) - well, perhaps these two:

Relational alignment of case marking and relationship between genitive and adjective ordering (information other than that supplied by Gabelentz is in square brackets):

| ERGATIVE <br> ALIGNMENT | DIVERGENT GENITIVE <br> AND ADJECTIVE ORDER | combination attested in |
| :--- | :---: | :--- |
| + | + | Basque, Tibetan, Eskimo |
|  |  | [Hurrian, Urartean, Australian, Papuan] |
| + | - | Australian [?] |
| - | + | $[?]$ |
| - | - | [German, Ancient Greek, ...] |

- If a language has ergative-absolutive alignment, it will with more than chance PROBABILITY also have divergent genitive and adjective ordering; if it has divergent genitive and adjective ordering, it will, wITH CONSIDERABLY HIGHER PROBABILITY (IF NOT CERTAINTY), also have ergative-absolutive alignment.

Why? Answer: Ethnopsychology??? Co-evolution??? ("denn das Gleichzeitige muss ja organisch zusammenhängen", 1891: 180/1901: 168; whatever coexists at any one time coexists by necessity!?; better turned round: what organically coheres, or appears to, must have developed in sync)

Cf. this one: If a language has infixes, then it will also have adfixes.
Why? Because infixes can only originate from adfixes by means of metathesis, to optimise syllable structures, even at the cost of positional unfaithfulness. (That is, morphologically, infixes are "pre-/sub-infixes", or rather "in-prefixes" or "in-suffixes") (1891: 330-331/1901: 348-349)
And not? Because discontinuous constituents are harder to store, process, and produce.
The new task for linguistics:
(i) Die Arbeit verlangt eine Kommission, und die Kommission verlangt ein bis ins Einzelnste ausgearbeitetes Programm, und dies Programm verlangt mehr selbstentsagenden Gehorsam, als man von der Mehrzahl der Gelehrten erwarten darf. Doch solche Schwierigkeiten sind zu überwinden.
Unter dem Programme aber denke ich mir eine Art Fragebogen, der kategorienweise alle grammatischen Möglichkeiten erschöpft, so dass jede Frage mit einem Ja oder Nein beantwortet ist. Eine solche Fragestellung ist schwierig für den Fragesteller selbst, manchmal auch für den Beantworter; aber Unmögliches wird keinem der Beteiligten zugemutet. (1894: 6)
[The work needs a committee, and the committee needs a programme attending to every single particular, and this programme needs more self-denying obedience than can be expected from most scholars. But such difficulties have to be overcome. I conceive of the programme as a kind of questionnaire which covers all grammatical possibilities category by category, with each question answerable by a mere yes or no. Framing such questions is difficult for those asking them and probably also those answering them; but nothing impossible will be required of either.]
(ii) Einem nothwendigen, die Gewähr der Richtigkeit in sich tragenden Gedanken darf man aber nicht darum entsagen, weil der erste Versuch, ihn zu verwirklichen, fehlschlug. Es gilt, ihn in eine kontrollierbare Form zu kleiden, und besser kontrollierbar ist keine als die statistische. Hier wünschte ich die Arbeit anfangen zu sehen. (1894: 4)
... jene grammatische Statistik ..., die ich vorhin als Typologie bezeichnete (1901: 484)
[A necessary idea, guaranteed to be right, must not be abandoned because the first attempt to realize it failed. It needs to be framed in such a form as to be controllable, and no form is better controllable than the statistical. Here I would wish to see work to commence.
... that grammatical statistics ... which I above referred to as typology.]

- For the first time, then: WITH MORE THAN CHANCE FREQUENCY ...


## 8. 1901-2006

[Another century to be skipped here. Is well known, anyhow: Greenberg, Principles and Parameters, us.]

## 9. The next 369 years: Finishing unfinished business?

- Acquiring (and making generally accessible) knowledge about more languages ("reference"/descriptive grammars, questionnaires).

NB: A pity so many (how many? mass decimation, or only constant diversification?) have been discontinued undescribed/undocumented.

- Acquiring (and making generally accessible) knowledge about particular languages, especially ones as yet poorly described, that is
- more comprehensive
- more detailed/in-depth ("thick" rather than "thin" description)
- more reliable
than what is known currently.
- Identifying more and more individual variables, and ascertaining whether or not they co-vary.
(In effect, feeding all that is known about all human languages into your computer and let it check all possible correlations.)

NB 1: The more is known about different languages, the more difficult it will be to perceive unity in diversity, =invariants. It's easier to persuade yourself that something is invariant if you don't know much about variation.

NB 2: Why should anything be EXPECTED to depend on/co-vary with anything else (unless there is a logical connection)?

- common heritage? (But why jointly retained?)
- common contact possession? (But why jointly borrowed?)
- sub-system held together by abstract connections, structured by abstract general organising principles?
- real-time dimension in developmental scenarios?
e.g., Why should noun inflection imply/be dependent on verb inflection?
(Campanella)
perhaps because verb inflection (i) is created more quickly, (ii) is more pertinacious, (iii) is gotten rid of more slowly than noun inflection.
- Improving methods of inductive generalisation, of drawing valid inferences about universals from limited crosslinguistic evidence:
constructing "representative" samples, representative of crosslinguistic diversity: not necessarily bigger samples, but samples whose members are maximally independent, to guard against mistaking (i) common heritage from protolanguages or (ii) borrowings among (macro-)areal neighbours for (statistical or absolute) universals.

NB: Sceptics doubt that, once (pertinacious) inheritance and borrowings have been identified, anything will remain that could claim universality. If they are right (which is an empirical question), the typological research programme can be discontinued in favour of historical research, in particular areal linguistics and population history.)

- Furthering specialisation: Quest for unity in/constraints on diversity an end in itself, not the by-product of other concerns (e.g., diachronic ones: quest for laws of change - but then, co-variation may turn out to be due to co-evolution, see above).


## 10. The next 369 years: Constraints on what - analysis or re-analysis?

Linguistic typology is about the diversity and unity of languages - or, to avoid the problematic concept of "a language", about the diversity and unity of grammars \& lexicons. By "a grammar \& lexicon" is meant the linguistic know-how of an individual, represented in that individual's mind (well, brain). The know-how to express and communicate thought can be shared among individuals, or else it distinguishes one group of individuals from another: speech communities, with no differences in linguistic know-how among their members (an idealisation); families/genera/phyla and Sprachbünde/diffusion zones/macro-areas, with the members of each sharing what is "inherited" or "borrowed"; the whole subspecies of homo sapiens sapiens, with at least that in common which earns it the attribute loquens.

Now, grammars \& lexicons are acquired: the linguistic histories of individuals or populations are the histories of the acquisition of linguistic know-how and its diffusion across populations. When grammars \& lexicons are, in some respect, the same, this may be due to (i) the relevant linguistic experiences of learners having been the same, or (ii) the linguistic know-how of acquirers being innately the same (leading them to arrive at the same grammars \& lexicons irrespective of experience). This innate know-how can guide/constrain structuring or restructuring.
(i) Timeless laws

Universals, seen as timeless laws, impose (absolute or conditional) limits on variation across grammars \& lexicons at any and all times;
they thereby constrain change insofar as a grammar \& lexicon must not change (not be restructured) so as to violate such a universal, or at any rate not without subsequent changes swiftly redressing the balance.
(Possibly: There are no laws of change itself; anything can be reanalysed as anything else independently of anything else.)
(ii) Laws of change

Particular targets (forms, categories, constructions, rules, constraints) can only result by particular mechanisms of change (reanalysis) operating on particular sources (forms, etc.);
such laws of change thereby impose limits on how grammars \& lexicons can differ: they can only be what they could become.
(Possibly: There are no timeless universals. Or, a view I share:
Co-variation is due to co-evolution, with concomitance or consecutiveness of reanalyses being superintended by timeless laws.)

The view that has diachrony in charge of constraining (ii) is the more traditional one; for a while superseded by position (i), it has lately been regaining popularity, much inspired by Greenberg's programme to "dynamicise" typology. In the case of implicational constraints, the most straightforward way of dynamicisation has been to read "implies" as "derives from". (Three examples: (a) Prep NP $\supset /<$ N Genitive, NP Postp $\supset /<$ Genitive N; that is, when head nouns in attributive constructions are grammaticalised as adpositions, if nothing else happens, they will be remain in the same position relative to the NPs they continue to be in some sort of similar construction with. (b) Infixes $\supset /<$ adfixes, the (almost) only source of infixes being adfixes, metathesised inside stems, around initial or final consonants, to improve syllable structures. (c) Nasal vowel $\supset /<$ corresponding oral vowel, the (almost) only source of nasal vowels being oral vowels nasalised through a following nasal.) Though straightforward, this way of dynamicisation is too facile: the full diachronic stories behind implications are more complex.

## 11. The next 369 years: Constraints on what - abstract and/or concrete representations?

## Ingredients of concrete representations (BLT)

See the Lingua Descriptive Series for typical concrete descriptions (following the questionnaire of Comrie \& Smith 1977), and see The Universals Archive (at http://typo.unikonstanz.de/archive) for typical typological findings cast in terms of concrete grammar.

- forms-in-constructions, at complexity levels of syllables, feet, stems, (syn/phon) words, (syn/phon) phrases, (syn/phon) clauses, sentences, paragraphs, texts;
- (sub-)classes of forms, (sub-)classes of constructions:
form classes, distribution/position classes, meaning classes;
- relation of (immediate) precedence (i.e., linear order);
- relation of (immediate) constituency (i.e., part-whole);
- syntagmatic relationships between manifest constituents of constructions, not necessarily specific to particular classes of constructions:
- subjunction, or dependency (head - dependent),
with subtypes: modification, determination, complementation, attribution, predication, apposition, ...
predicate-argument, circumstance quantification, classification
- conjunction,
- adjunction,
- ...;
- several kinds of influence (preferably local, possibly also at a distance) among manifest constituents of constructions, not specific to particular classes of constructions [and how to express such influence in your descriptive framework is your own business]:
- exclusion of co-presence,
- requirement of co-presence,
- license of co-presence,
- government, command,
- referential binding,
- scope-taking,
- ...;
- paradigmatic relationships;
- kinds of constructional marking:
- relationship-identifying,
- relatedness-indicating,
- linking, separating,
- speech-event and context anchoring;


## typically concrete MEANINGS/FUNCTIONS

- reference, reference-modification, predication;
- topic, comment, focus;
- assertion, question, command, exclamation, ...;
typically concrete EXPLANATION
- expressiveness [allowing/facilitating the expression of thought];
- simplicity (formal economy, markedness) [facilitating storage, production, processing];
- clarity (ambiguity avoidance) [facilitating processing];
- iconicity (function motivating form) [facilitating what? giving pleasure?];
- inertia (Don't reanalyse! or, Only reanalyse within reason/limits!);
- extravagance (Be different!);


## How can representations of forms and constructions be abstract? (ALT)

- first, by virtue of excluding all sorts of things present in the speech signal, but considered irrelevant for grammar;
- second, by virtue of including some sorts of things not present in the speech signal, but considered relevant for grammar.

Abstractness can involve:

- structurally relevant PRECEDENCE of parts of wholes differing from manifest precedence (movement/displacedness);
- structurally relevant constituency differing from manifest CONSTITUENCY (restructuring);
- structurally SOMETHING represented by manifestly NOTHING, with subtypes of manifest nothingness:
phonological (pro, PRO, trace), syntactic (ellipsis);
- structurally NOTHING represented by manifestly SOMETHING (including underspecification).

Three cases where typological insight is gained (only) by assuming abstract representations, as being those subject to constraints, and by accounting for manifest diversity through modularisation

## (i) The coronal syndrome: Underspecification of segments for place of articulation

"First, coronals are the most frequent articulator choice on a number of counts. In the UG phonetic alphabet, the Coronal articulator supports a larger number of dependent (consonantal) features than Labial, Dorsal, and Pharyngeal. In the phonemic systems of individual languages, coronals typically outnumber the other Place categories as well. Finally, they have been documented as among the most frequent consonants in speech corpora of English and Spanish. Second, coronal is the normal outcome of rules and constraints that neutralize Place contrasts [...]. Third [...] phonologists have the impression that coronal is the most commonly chosen epenthetic or othewise dummy oral consonant [...]. Fourth, coronals more freely combine with each other as well as with other consonants, eluding phonotactic restrictions that are enforced on labials and velars. [...] Fifth, coronals are more susceptible to Place assimilation than noncoronals [...], [If noncoronals assimilate in Place, then coronals assimilate in Place too - Mohanan 1993.] Finally, coronals are
more likely to be transparent to transconsonantal vowel-echo rules than labials or velars." (Kenstowicz 1994: 516-517, see further Paradis \& Prunet (eds.) 1991, and much work since, done in Konstanz and elsewhere.)

Explanation of family of relevant universals: Underspecification of lexical representations, with PoA unspecified for coronal segments (or only specified when feature [Coronal] is phonologically active).

## (ii) Concrete $\neq$ abstract order in morphology: No infixes in morphology

- Some languages have "infixes", always in addition to adfixes (suffixes and/or prefixes).
- Nonetheless, morphological variation concerning the positioning of affixes relative to stems can plausibly/insightfully be constrained as follows:

There are only adfixes and no infixes in any language, morphologically speaking.

- In morphological representations of relevant words in relevant languages, what are manifestly "infixes" (e.g., Latin present stems of the relevant conjugation class, such as $f u-\underline{N}-d-$ 'shed', vi- $\underline{N}-c$-'conquer', ru- $\underline{N}-p$ - 'break'; Tagalog perfective verbs such as $k$-um-ain 'ate', $p$-um-asok 'entered'; Tiene causative verb stems such as $l$-as-ab'cause to walk'; Ulwa construct state forms such as suu-ka-lu '(his) dog', siwa-ka-nak '(his) root', karas-ka-mak '(his) knee', with the "infixes" underlined) are (abstractly) adfixes. This enables the universal to be upheld, though valid only for morphological representations, where the order of word parts may be different from the manifest ordering of these parts.
- The adfixes of morphological representations can be subject to metathesis or other reorderings, which are phonological means with the aim of optimising prosodic structures, namely those (i) of syllables (as syllabified at the stem or perhaps also the word level, aiming at CV patterns or at sonority sequencing; as in Tagalog and Latin) or (ii) of syllable groupings, i.e., feet (aiming at the foot type preferred in the language; as in Ulwa, or also in English expletive and -ma- "infixation": abso-bloomin-lutely, sophisti-ma-cated), or also (iii) of stem templates (as in Tiene, where derived verb stems are of the shape $\mathrm{C}_{1} \mathrm{VC}_{2} \mathrm{VC}_{3}$ and need to form a "prosodic trough" (Hyman 2006) with $\mathrm{C}_{2}$ coronal and $\mathrm{C}_{3}$ non-coronal).
- Diachronically speaking, phonological reordering is the only productive mechanism to create "infixes". Very occasionally, "infixes" have been created through "entrapment", with an outer affix reanalysed as part of the stem. Whether they have been reordered or entrapped, "infixes" have always been adfixes in their previous lives.
- Synchronically speaking, "infixes" are always edge-bound: they are never found further inside stems than after/before the initial/final constituents of the relevant prosodic unit (before plosive syllable-coda in Latin, after syllable-onset in Tagalog, after first iambic foot in Ulwa, before the final non-coronal consonant of the template in Tiene). And they are, in the relevant languages, always also realised as adfixes with stems that do not require prosodic optimisation through phonological reordering (cf. Latin stems without a stem-final plosive such as si- $\underline{N}$ - 'leave', $k e r-\underline{N}$ - 'separate', (con-)tem- $\underline{N}-$ 'despise’, pell- (< pel- $\underline{N}-$ ) 'expell'; Tagalog vowel-initial stems such as umawit PERF of 'sing'; Tiene verb stems with a final coronal, forming a "prosodic trough" when suffixes with a non-coronal consonant remain external, such as mat-is-
(-> maas-) 'cause to go away'; Ulwa iambic stems of two morae, contributed by one heavy syllable or two light ones, or of three morae, contributed by a light syllable followed by a heavy one, such as kii-ka '(his) stone', sana-ka '(his) bee', sapaa-ka '(his) forehead'). Edge-boundedness and external occurrence in the case of some edges support the analysis of "infixes" as created by phonological reordering from morphological adfixes.
- Crucially reasserting their morphological adfix status, "infixes" can subsequently be re-externalised, even from the very stems they used to surface inside of (as in Toba Batak: Crowhurst 1998). (This won't happen once they have become lexicalised, as probably in Latin; but then, being lexicalised, they aren't part of productive morphological representations, either.)
- The remaining question is what decides the competition between morphology and phonology at any given historical stage: morphological faithfulness has adfixes remain adfixes in concrete word constructions; phonological (in particular, prosodic) optimality has affixes in a manifest order which sounds better than the morphologically faithful ordering. Before and after the internalisation of adfixes morphology wins (why?); at internalisation phonology wins (why?).
- As to the accessibility of abstract morphological representations: given that there are always instances of adfixes in both a stem-external and a stem-internal position, depending on the prosodic shape of stems, a learner can infer adfixal positioning for abstract representations.
(iii) Concrete $\neq$ abstract order in syntax: Attributive adjectives are always ordered iconically, before nouns or after

English et al.
a beautiful big red ball
VALUE SIZE COLOUR N

Bahasa Indonesia et al.
bola merah besar tjantik
N colour size value
Maltese et al. (incl. Celtic)
ballun sabiћ kbir aћmar
N value size colour
Italian et al.
una bella grande palla rossa
value size N colour
unattested (or so it seems)
redd bigg beautifull ball
colour size value N
In establishing such unmarked orderings all kinds of considerations have to be borne in mind:

- make sure the language permits the stacking of property-concept words to begin with (some/many don't, or only do so very reluctantly);
- the several property-concept words shouldn't be (asyndetically) coordinated (the only option in non-stacking languages);
- property-concept words should be modifiying the noun rather than one the other (a beautifully red ball);
- there should be no intonation breaks between the property-concept words suggesting non-cohesiveness of the modifier construction;
- sometimes inherently emphatic property-concept words (especially 'big') push forward;
- there may be further factors relevant for ordering, such as the length or other phonological properties of property-concept words;
- further semantic classes of property concepts should be taken into consideration, such as weight, age, material, provenance, and purpose, increasingly closer to the noun ( $a$ beautiful big heavy new red woollen Swiss medicine ball).

Descriptive generalisation over ( $1 \mathrm{a} / \mathrm{b}$ ), the crosslinguistically most common orderings: The relative distance (position class) of semantic classes of property-concept words from the noun is the same, whether the noun comes after or before propety-concept words. (That is, the two most common orderings are mirror images of each other.)

This generalisation is richly supported in the relevant literature, most substantially by Hetzron 1978 and Sproat \& Shih 1991.

Higher-level generalisation, perhaps explanation of ( $1 \mathrm{a} / \mathrm{b}$ ):
The nounier a modifier, the closer to the noun.
The nouniness ranking COLOUR $>$ SIZE $>$ VALUE is independently motivated, on languageparticular and general grounds. Relevant evidence includes: the nounier modifier words actually are themselves nouns, or are derived from nouns; they can enter a morphological relation with head nouns (compounding); their range of applicability to nouns of different semantic classes is narrower. The less nouny modifier words are verbal forms (participles) or are derived from verbs; they don't compound; their range of application is wider. In the smallish adjective inventories of languages which prefer to express property concepts through nouns and/or verb, specifically adjectival lexicalisation of property concepts also follows that scale.

Even higher-level generalisation/explanation, in terms of ICONICITY:
Linear closeness mirrors conceptual distance.
(And it remains to be seen whether the many other mirror image orderings identified by Cinque 2006 can be subsumed under this same generalisation.)

It follows from this account that (2b) should not occur (and it apparently doesn't).
But neither should (2a), which does, even if not so frequently!
Considering the existence of (2a), the obvious question is:
WHY is there no mirror image of (2a), i.e., (2b)?
The answer:
Maltese et al. are really, at a relevant level of syntactic representation where conceptual proximity matters (which is abstract), like English et al., i.e., their NPs are N -final, and N is fronted in manifest form (half-way fronted in Romance).
(See Cinque 1993, Longobardi 1994)
That is, iconicity rules ok everywhere!
But for this to be possible, syntactic representations need to be assumed which are abstract, in the sense of manifest order $\neq$ structural order.

Can such abstract syntactic representations be indirectly inferred by a learner from any concrete forms and relations between forms?

In Maltese (possibly like in other relevant languages with abstract precedence differing from concrete precedence), a few adjectives do precede their nouns, but only in highly marked formal and stereotyped constructions:

```
il-kbir Alla
DEF-big god
`God Almighty'
l-g\hbararef ћu-k
DEF-wise brother-2SG
'your clever brother'(only ironic)
```

Synthetic comparatives, themselves postnominal, move ahead of their noun to express the superlative grade (with one definiteness marker suppressed):

```
il-belt il-qawwi-ja
DEF-city DEF-beautiful-FEM
'the beautiful city'
belt aqwa
city beautiful.COMP
'a more beautiful city'
l-aqwa belt
DEF-beautiful.COMP city
'the most beautiful city'
```

Analytic comparatives/superlatives as such stay behind; only the comparative marker on its own may move ahead (again, with one definiteness marker dropped):

```
il-ktieb l-aktar sabi\hbar
DEF-book DEF-more beautiful
'the more/most beautiful book'
l-aktar ktieb sabih
DEF-more book beautiful
'the most beautiful book'
```

And, being in some ways similar to adjectival modifiers, numerals other than 'one' also precede nouns (which are in the singular with numerals above 'ten'):

```
g\hbaroxrin suldat
twenty soldier (SG)
'twenty soldiers'
```

Whatever the right "thick description", it seems to me doubtful that the manifest ordering of nouns in very marginal modifier constructions could suffice to inspire learners to set up abstract representations at odds with manifest ordering in ordinary constructions encountered much more frequently.
[To mention an analogous case, abstract uniformly verb-final representations, thus also for V2 main clauses, in Germanic are perhaps easier to infer for learners, on the assumption that verbfinal patterns are unmarked, lacking all sorts of things that make a clause "main".]

But what other inspiration could there be?

Answer: Universal Grammar prohibiting representations of modifier constructions at odds with iconicity, thereby forcing abstract representations on learners of languages where manifest ordering is at odds with iconicity.

The remaining questions are (i) why iconic representations can be interfered with through N fronting, and (ii) why there are no abstract representations in line with iconicity ending up with counter-iconic concrete precedence relations through N -backing - deriving manifest (2b) from abstract (1b). Ruled out by UG?

A lead that remains to be explored is that languages which have N -fronting (including Semitic and Celtic, but not in fact Maltese) seem to be ones where V is initial too, at least in abstract representations. If there really is a connection between the construction types of noun phrases and of clauses, how can this be made sense of? Would seem to require abstract representations where noun phrases and clauses are seen to share constructional principles that are perhaps impossible to express in concrete representations. (Inflectional differences in the marking for gender and number have also been claimed to be implicated; but this seems more dubious.)

