

Better Typology

1. What is typology? Slightly rephrased: What is it about?

Answer: diversity (across languages and in time) and unity — i.e.,

- How do languages (rather, grammars & lexicons) differ?

How have they come to differ?

- How do languages (grammars & lexicons) not differ?

How have they come not to differ?

Pretty profound questions, it would seem, which are central to linguistics as an intellectual field and an academic discipline, not to be relegated to a marginal subspecialisation, taught in optional classes for students not keen on “theory”.

Doing typology is to pursue these questions in accordance with accepted professional standards of scientific research.

(Which are a bit tricky, since linguistics is both humanities and science.)

2. What do/should typologists expect from “formal” “theoreticians”?

Unanswerable, owing to presupposition failure:

- Typology is theory; it’s evidently not applied linguistics.

[In the case of grammar writing it is perhaps less evident that this is theory rather than application. But then, typology isn’t grammar writing: it only depends on it, and vice versa — see below.]

- What’s “formal”, anyhow? (algebraic linguistics, yes, but procrastination? greed? ...)

3. Question rephrased: What do/should typologists expect from themselves?

Answer (though it’s obvious): doing typology well (or optimally, under the circumstances) — generally speaking,

- asking important questions about diversity and unity;
- giving plausible answers (so far as circumstances permit);
- which raise even more important questions;
- whose answers will be even more intellectually satisfying.

4. Question of what it needs to do typology well answered somewhat more specifically:
- being good at linguistic description and analysis;
 - being knowledgeable about crosslinguistic diversity and its history:
 - being curious: seeking to obtain all knowledge available and then extend it;
 - showing respect: paying equal attention to all languages;
 - having a sense of priorities: paying more immediate attention to languages not well described and especially to ones endangered;
 - being good at inductive methodology: at generalising from evidence that will remain limited (the domain being all human languages, including those extinct without a trace) no matter how conscientiously current knowledge is being accessed and extended;
 - serendipity (=the natural ability of making fortunate and unexpected valuable discoveries by chance).

5. Though (one hopes) typology has been done well for a while, and is being done well in these respects, there is potential, and need, for improvement — most urgently, in my view, along three lines.
6. First, the relationship between typology and diachrony needs to be clarified:
Who is in charge of constraining?
[See elsewhere for what I think is the answer.]

TYPOLOGY, IN RELATION TO DIACHRONY

Typology is about variation across languages (as mentally represented and socially shared knowledge), diachrony is about variation of languages in time. Neither typological nor diachronic variation appears to be unlimited; against the backdrop of the possibilities for diversity, yet to be mapped out comprehensively, the limitations need to be reliably identified and to be plausibly accounted for.

The question is how constraints on variation across languages and in time, to the extent that they are not due to accidents of population histories (and it is not always easy to disentangle structural necessity and historical chance), bear upon each other. Two alternative answers have been given, attributing (possibly exclusive) responsibility to either typology (a) or diachrony (b):

- (a) Universals, seen as timeless laws, impose (absolute or conditional) limits on variation across languages at any and all times; they thereby constrain change insofar as a language must not change 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.)

- (b) 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 languages 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 changes being superintended by timeless laws.)

The view that has diachrony in charge (b) is the more traditional one; for a while superseded by position (a), it has lately been regaining popularity, inspired by Greenberg's programme to "dynamicise" typology. In the case of implicational constraints, the most straightforward (arguably too facile) way of dynamicisation has been to read "implies" as "derives from". (Examples: (i) 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 remain in the same position relative to the NPs they continue to be in some sort of similar construction with. (ii) Infixes \supset /< adfixes, the (almost) only source of infixes being adfixes, metathesised inside stems, around initial or final consonants, to improve syllable structures. (iii) Nasal vowel \supset /< corresponding oral vowel, the (almost) only source of nasal vowels being oral vowels nasalised through a following nasal.)

7. Second, (i) the existing crosslinguistic knowledge needs to be made better use of and (ii) the acquisition of such knowledge needs to be better guided.
- crosslinguistic knowledge is growing at an unprecedented rate: book-length grammars for ca. **300** languages have been published in the last decade only, many previously un- or ill-described.

The *ALT* Grammar Watch

[2005, at <http://www.lancs.ac.uk/fss/organisations/alt/>]

Although the *ALT* Grammar Watch [drawing attention to recently published grammars, as a service for those with a professional interest in linguistic diversity and unity; regularly updated in the *ALT News*] only covers about a decade [1995–2005], and cannot claim completeness, the list is already longer than those which entire previous centuries could take credit for. With grammar-writing a growth industry, linguistics, long remarkably uncurious and given to hasty generalization, is, for the first time, seriously closing in on its subject matter, human language in **all** its manifestations. As each grammar is a token of our profession's respect for a culture, so is the attention we pay to each grammar a measure of our self-respect as theoretical linguists.

- however, existing crosslinguistic knowledge is not utilised optimally by typologists (notwithstanding (i) a growing general awareness of the extents of diversity and (ii) much improved sampling methodology);
- even when utilised extensively, lots of information is typically found missing in descriptive grammars (or is easy to overlook, owing to grammars being organised less-than-transparently):
 - decent grammar of sounds (including intonation/prosody);
 - various concrete features, such as proper naming or word formation;
 - abstract grammar [on which below];
 - features of oral, informal speech (as opposed to written, formal speech);
 - in general the full range of variability across a speech community;
 - a corresponding lexicon;
 - texts of various genres, to show grammar and lexicon at work;
 - distinction between what is language-particular and what reflects universals;
 - a reliability guarantee (“the truth, the full truth, and nothing but the truth”).

8. Third, in order to avoid such shortcomings, grammar writers want/need better guidance from experts in crosslinguistic diversity and unity (=typologists).

Grammar writers need to bear in mind, and perhaps occasionally to be reminded,

- that ANALYSIS (or “thick description”) is needed in addition to (i) observation, (ii) recording, and (iii) (“thin”) description, and analytic decisions need to be justified;

[“Thick description” is from Clifford Geertz, *The Interpretation of Cultures*; his favourite example, borrowed from Gilbert Ryle, is the wink of an eye, thinly described as the rapid contraction of a person’s right eyelid, thickly described as the practicing of a burlesque of a friend faking a wink to deceive an innocent into thinking conspiracy is in motion.]

- that grammar needs to be recognised to involve ABSTRACT representations.

9. Why abstract representations?

- Abstract description naturally leads on to explanation, insofar as the particular is more easily seen as a special instantiation of something more general also instantiated otherwise (going beyond type-instantiations through exemplars).
 - Abstract representations, perhaps more centrally than concrete representations, are involved in
 - speech production and processing;
 - brain activity/neural events, in terms of local and temporal patterns discernible by current methods of neuro-imaging.
 - Abstract representations, specially benefiting typology, may in fact reinforce explanatory principles commonly appealed to in typology, while concrete representations are at odds with them.
- [Case studies: (i) adjective ordering, governed by iconicity; (ii) no infixes, morphologically speaking; (iii) segments underspecified or specified differently.]

INGREDIENTS OF CONCRETE REPRESENTATIONS

[concrete vs. abstract better characterises current alternative ideologies than functional/descriptive vs. formal/theoretical]

See the Lingua Descriptive Series for typical concrete descriptions (answering the questionnaire drawn up by Comrie & Smith 1977), and see The Universals Archive (at <http://ling.uni-konstanz.de/pages/proj/sprachbau.htm>) for typical typological findings cast in terms of concrete grammar.

FORM

- forms-in-constructions,
at complexity levels of 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:

- 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 MEANING/FUNCTION

“functions” of forms-in-constructions:

reference, reference-modification, predication;

topic, comment, focus;

assertion, question, command, exclamation, ...;

typically concrete EXPLANATION

explanatory concepts (communicative maxims), behind highest-level descriptive generalisations:

- 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!);
- ...

In the most general terms, then: How can representations be abstract?

- 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 in the (more interesting) “inclusive” sense can involve:

- structurally relevant PRECEDENCE of parts of wholes differing from manifest precedence (movement);
- 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). [Or is this “exclusive”-type abstractness?]

structurally something represented by manifestly nothing: Bavarian *Bua(b)* ‘boy’

manifest form of SINGULAR: [buə] — no coda consonant

structurally: /buəb/ — with coda consonant /b/

However, abstract /b/ can be inferred from other manifest forms of the same stem in construction with inflections (or also from derivationally related forms such as diminutive [biəvəl], where /b/ is intervocalic and realised as a labial fricative) and from the manifest form of the relevant inflectional exponents in construction with other stems:

manifest form of PLURAL: [buəm]

*/buə-m/ because elsewhere (after non-labials, with nouns of the same declension), PLURAL is /-n/

*/buə-n/ because this cannot be mapped on manifest form

/buəb-n/ can be mapped on manifest form, through

- labial assimilation, hence [buəbm]
- homorganic cluster simplification, hence [buəm]

Hence, coda /b/ is not really abstract in the strictest sense: as a segment specified for labiality, it is manifested indirectly, through an adjacent segment participating in a regular non-labial/labial alternation.

Only in cases of absolute neutralisation would there be no (indirect) way of inferring an abstract contrast.

structurally something(s) represented by manifestly nothing: trace, PRO, and cliticisation of *to* in English

*Who_i do you_j want PRO_j to succeed t_i?
(You want to succeed who?)*

... wanna ...

*Who_i do you_j want t_i to succeed PRO_{arb}?
(You want who to succeed?)*

**... wanna ...*

Again, indirect manifestation of structural something(s), through possibility or impossibility of cliticisation and attendant reductions.

structurally relevant PRECEDENCE of parts of wholes differing from manifest precedence: relative proximity of attributive adjectives (or rather, property-concept words) of different meaning to their noun

English et al.

<i>a beautiful big red ball</i>	VALUE SIZE COLOUR N	(1a)
---------------------------------	---------------------	------

Bahasa Indonesia et al.

<i>bola merah besar tjantik</i>	N COLOUR SIZE VALUE	(1b)
---------------------------------	---------------------	------

Maltese et al. (incl. Celtic)

<i>ballun sabiħ kbir aħmar</i>	N VALUE SIZE COLOUR	(2a)
--------------------------------	---------------------	------

Italian et al.

<i>una bella grande palla rossa</i>	VALUE SIZE N COLOUR	(2a')
-------------------------------------	---------------------	-------

unattested (or so it seems)

<i>redd bigg beautifull ball</i>	COLOUR SIZE VALUE N	(2b)
----------------------------------	---------------------	------

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 modifying 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 (1a/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 first or last.

(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 (uncharitably or sloppily failing to acknowledge or even mention Hetzron).

Higher-level generalisation, perhaps explanation of (1a/b):

The nounier a modifier, the closer to the noun.

The nouniness ranking COLOUR > SIZE > VALUE is independently motivated, on language-particular 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.

It follows from this account that (2b) should not occur (and it 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).

That is, iconicity rules ok everywhere!

But for this to be possible, syntactic representations need to be assumed which are abstract.

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-gharef hu-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 sabiħ

DEF-book DEF-more beautiful

‘the more/most beautiful book’

l-aktar ktieb sabiħ

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’):

ghoxrin 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 main clauses, in Germanic are perhaps easier to infer for learners, on the assumption that verb-final 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 isn’t at odds with iconicity.

A remaining question is 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) 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.)

The infix story, in stark outline

- Some languages have “infixes”, always in addition to adfixes (suffixes and/or prefixes).
- Nonetheless, there are grounds to assume that morphological variation concerning the positioning of affixes relative to stems is strictly 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 *fu-N-d-* ‘shed’, *vi-N-c-* ‘conquer’, *ru-N-p-* ‘break’ etc.; Tagalog perfective *k-um-ain* ‘ate’, *p-um-asok* ‘entered’ etc.; Ulwa construct state *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, viz. those of syllables (as syllabified at the stem or perhaps also the word level; as in Latin and Tagalog) or of feet (as in Ulwa, or also in English expletive and *-ma-* “infixation”: *abso-bloomin-lutely*, *sophistica-ma-cated*).
- 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 at the relevant prosodic level (before plosive syllable-coda in Latin, after syllable-onset in Tagalog, after first iambic foot in Ulwa). 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-N-*

‘leave’, *ker-N-* ‘separate’, (*con-*)*tem-N-* ‘despise’, *pell-* (< *pel-N-*) ‘expell’; Tagalog vowel-initial stems such as *um-awit* PERF of ‘sing’; 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’). This supports the analysis of “infixes” as created by phonological reordering from morphological adfixes.

- Diachronically speaking, “infixes” can be re-externalised, even relative to the same stems they used to surface inside of.
- 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 affixes in both a stem-external and a stem-internal position, depending on the prosodic shape of stems, a learner could infer affixal positioning for abstract representations.

The underspecification story: demnächst in diesem Theater.