

W. D. C. de Melo
Week 1, Hilary term 2007

Linguistic Typology: Introduction

Before we begin

Timetable: Wednesday, 10–11 o'clock, weeks 1–8 in the Centre for Linguistics and Philology, Walton Street, Room 206

Lecturer: Wolfgang de Melo

Contact: wolfgang.demelo@all-souls.ox.ac.uk

If you need to contact me, please write 'Typology' in the subject heading of your e-mail.

Unity and diversity

There are between 4,000 and 6,000 languages spoken today. They are all different in many ways. But they also share many features.

(1) English: *David gave the book to me.*

(2) Biblical Hebrew: *nāthan dāwīdh eth-has-sēfer lī.*
give.3RD.PERF David OBJ-ART-book PREP.me
'David gave the book to me.'

(3) Cantonese: *David béi sīu ngóh.*
David give book I
'David gave the book to me.'

Note some similarities:

Phonology: all three languages have vowels and consonants.

Morphology, syntax and semantics: all three languages distinguish between nouns for people and things and verbs for actions. All have means to show unambiguously who is giving what to whom.

Note some differences:

Phonology: English uses pitch contours to distinguish different sentence types ('intonational language'). With a falling pitch contour, Ex. 1 is interpreted as a statement, with a rising contour it is a surprised echo-question. Cantonese pitch is marked lexically. A number of words differ only in their pitch patterns ('tone language'). Cf.: *sī* (high-level) 'poetry', *sī* (mid-level) 'to try', *sīh* (low-level) 'a matter', *sīh* (low-falling) 'time', *sí* (high-rising) 'to cause', *sīh* (low-rising) 'city'.

Morphology, syntax and semantics: In English and Hebrew the verbs show that the action is past; the nouns show that only one book was transferred. In

Cantonese, neither of this is clear without context; time words and numerals are needed if you want to be explicit.

In English and Cantonese the subject (agent) is in first position, followed by the verb, followed by the direct object (patient), followed by the indirect object (recipient); but in Hebrew the verb is in first position, followed by the other constituents.

In Cantonese, the semantic roles are marked by word order only; in English, subject and direct object are marked by word order, and the indirect object is marked by word order and the preposition *to*. In Hebrew, the subject is marked by word order only, and direct and indirect objects are marked by both word order and prepositions.

What does typology do?

The aim of typologists is to uncover the unity behind the diversity of language structures. They want to find out about constraints on what constitutes a possible language.

- What structures occur in all languages, and why? For example, why do all languages have nouns and verbs?
- What structures are impossible, and why?
- Why are some structures more frequent than others across languages?
- Why do certain structural elements always or never co-occur with certain others, e.g. why do languages with nasal vowels always have oral vowels, but why do languages with oral vowels not always have nasal ones?

Constraints on what constitutes a possible language are also called universals.

Universals

Universals are cross-linguistic generalizations.

Universals may be absolute. Absolute universals / substantive universals have no exceptions in any language. There are not many absolute universals. Examples are:

- All languages have vowels and consonants.
- All languages have nouns and verbs.

But many things that you might expect to hold for every language are not absolute universals; for instance, not every language has adjectives. Chinook (North America) uses verbs instead, and Igbo (Kwa family, part of the Niger-Congo family) has only eight adjectives.

Other universals are statistical. Statistical universals hold in the majority of languages. For example, in more than 90% of all languages, the subject precedes the object (other constituents may of course intervene). Languages in which the object precedes the subject are exceedingly rare. One of the few known languages for which this is the case is Hixkaryana (Carib family), which has object-verb-subject order.

The last category is implicational universals. A universal is implicational if two features correlate; for example, object-verb order often goes hand in hand with postpositions.

Implicational universals can be absolute or statistical. The universal that object-verb order goes hand in hand with postpositions is statistical. Hindi follows it, but classical Latin (as written by Caesar) does not. Caesar has a fairly strict object-verb order, but prepositions.

Hierarchies

Most universals are implicational. If there is more than one implication, we can speak of a hierarchy. Let us look again at the example sentences above. Cantonese marks core semantic roles (subject, direct object, indirect object) by word order. English uses a preposition for the indirect object. Hebrew has prepositions for direct and indirect object. Adpositions in general can be regarded as analytic case markers. Latin goes one step further and marks subject, direct object, and indirect object by cases.

These differences are not random. You are unlikely to find a language which marks subjects, but not direct and indirect objects, or a language which marks subjects and direct objects, but not indirect ones. Thus, subject marking implies object marking, and direct object marking implies indirect object marking, but not subject marking. We can represent the implications in a hierarchy:

subject marking $>$ direct object marking $>$ indirect object marking,

where $x > y$ means that if x is marked, then everything to its right will be marked as well.

Reasons for universals

Languages can share features for a variety of reasons:

Common origin: all Romance languages (Italian, French, Spanish, Portuguese, Rumanian etc.) share a number of phonological, morphological and syntactic features, for instance a two-gender system (sometimes with vestiges of a third gender). Compare: French *le loup* masc. ‘the wolf’, *la lune* fem. ‘the moon’; Italian *il lupo* masc., *la luna* fem.; Spanish *el lobo* masc., *la luna* fem.

Language contact: Balkan area, India etc. In the Balkan region, we find Indo-European languages from separate branches (Greek, Romanian (Romance), Bulgarian (Slavonic), Albanian), which have developed a number of shared features:

absence of infinitives, futures formed with the verb ‘to want’, post-posed articles (Romanian *lup-ul* ‘the wolf’; not in Greek). In India, we find mainly Indo-European and Dravidian languages, which have influenced each other (retroflex consonants, subject-object-verb order, postpositions, preposed genitives).

Chance: see constituent order in English and Cantonese above

But features which can be explained like this are always absolute or statistical universals. And even if you exclude all these factors, there remains a substantial amount of similarities that has to be explained. But what is most important is that implicational universals and hierarchies cannot be explained by a common ancestral language, language contact or chance.

How can universals be explained?

Chomsky’s explanation of universals

Human beings are genetically predisposed to learning languages. The similarities between languages are the result of similar or identical brain structures. Human brains have a ‘language acquisition device’ (*LAD*), which accounts for the rapid acquisition of a first language despite the supposedly impoverished input.

Why are there different languages despite similar or identical brain structures? Every generation and even every culture within one generation has different technologies and strategies for coping with the environment. You need new words for new things, so languages have to be flexible. But languages cannot differ randomly, otherwise it would be impossible to master them. Some elements of language have to be fixed and immutable, others have to be flexible.

‘Principles and parameters’ theory: the fixed and immutable elements of language are fixed and immutable because of our brain structures; these are the so-called principles, which roughly correspond to the absolute universals above. The flexible elements can be divided into those which are completely unpredictable and those which can at least to some extent be predicted. The completely unpredictable elements comprise e.g. lexical items; a French child simply has to learn that a dog is a *chien*, just as an English child has to learn the word *dog*. Other elements can be predicted. If your language has postpositions, it is also likely to be verb-final (implicational universals). These are the so-called parameters: a child understands a postpositional phrase, realizes that heads follow their dependents, and then generalizes from here to other categories like objects and verbs.

A functional-typological approach

It is likely that Chomsky is at least partly correct. However, the principles and parameters approach is not very satisfying. In the absence of convincing proof it is not a good idea to postulate innate structures and a language acquisition device without exhausting alternative explanations first.

Typologists often look for functional explanations. The hierarchy above for example can be explained by the functional principle of economy: if a verb takes one constituent only, it will be a subject; if it takes two, they will normally be a subject and a direct object; if it takes three, they will usually be subject, direct object, and indirect object. It is more important to mark rarer constituents, as they are the ones which are most likely to be confused. So if a language marks only one type of constituent morphologically, it will be the indirect object; if it marks two, it will be both object types; if it marks three, it will be the subject and both objects.

Historical background

- Friedrich von Schlegel (1772–1829) first established a dichotomy between inflecting languages and non-inflecting ones.
- Wilhelm von Humboldt (1767–1835) first distinguished between isolation, agglutination, inflexion, and incorporation. Note that he regarded these as general strategies used in different languages, not as language types; each language is considered a type in its own right.

isolation: see Cantonese above; every word consists of just one morpheme

agglutination: words can consist of several morphemes, but each morpheme has just one function; cf. Turkish *çocuk* ‘child’, *çocuk-lar* ‘children’ (child-PL), or *ev* ‘house’, *ev-ler* ‘houses’, *ev-ler-im* ‘my houses’ (house-PL-POSS)

inflection: words can consist of several morphemes, with each morpheme fulfilling several functions at once; cf. Latin *laud-ant* ‘they praise’ (praise-3PL.PRES.ACT.IND)

incorporation: words consist of several morphemes, some of which have grammatical functions, others of which stand for concepts; in extreme cases, word and sentence coincide (e.g. Eskimo)

- August Schleicher (1821–68) postulated a diachronic development between isolating, agglutinating, and inflecting languages: isolating → agglutinating → inflecting → isolating
- Franz Nikolaus Finck (1867–1910) refined previous typologies and distinguished between eight types, such as root inflecting (Hebrew, Arabic) vs. stem inflecting (Greek, Latin).

These early approaches all focus on morphology. Greenberg’s seminal study on word order (1963) opened up other fields of enquiry.