

PARAMETERS OF LINGUISTIC VARIATION

Roberta D'Alessandro
LUCL Leiden
r.dalessandro@hum.leidenuniv.nl

1. THE LIMITS OF SYNTACTIC VARIATION

(1) 'Languages can differ from each other without limit and in unpredictable ways' [Joos 1957:96]

Is this statement true?

Looking at the differences

Looking at the similarities

(2) '[t]he fact that some features are [...] widespread is worthy of notice and calls for explanation; when we have adequate data about many languages we shall have to return to the problem of general grammar and to explain these similarities and divergences'

(3) 'looking forward to a larger synthesis, a General Grammar, which will register similarities between languages' [Bloomfield 1934:46]

What are the limits of syntactic variation? What are the main approaches to observe and study them?

1.1. MORPHOLOGICAL TYPOLOGY

Schlegel (1817), Schleicher (1862) and Sapir (1921)

A classification of languages on the basis of word-formation patterns

i. **ANALYTIC** (also called isolating) [one word-one morpheme]

ii. **AGGLUTINATING** [one morph-one morpheme]

iii. **INFLECTING** (also called synthetic or fusional) [one morph-several morphemes]

iv. **INCORPORATING** (also called polysynthetic) [verb with all its arguments in one word]

v. **INFIXING** [infixation of vowels in a root which contains only consonants]

[Katamba 1993: 56ff]

Typological classification of languages. Greenberg (1955):

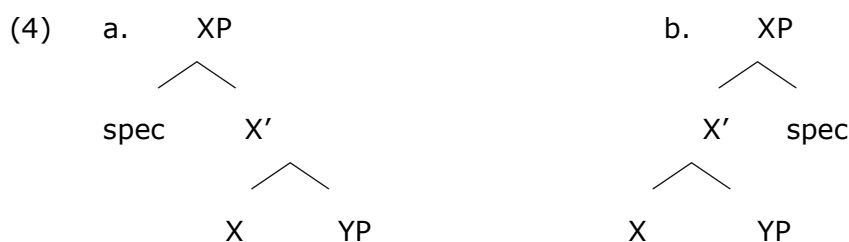
- If a language has between 1.00 and 1.99 morphemes per word it is **ISOLATING (ANALYTIC)**
- If a language has between 2.00 and 2.99 morphemes per word it is **INFLECTING (SYNTHETIC)** if the realisation of the different morphemes tends to be simultaneous
- If a language has between 2.00 and 2.99 morphemes per word it is **AGGLUTINATING** if each morpheme tends to be realised by a separate morph
- If a language averages 3.00 morphemes per word or more it is **INCORPORATING**

EXAMPLES OF UNIVERSALS

[TYPOLOGICAL IMPLICATIONAL UNIVERSAL]: if a language has distinct reflexive pronouns (i.e. distinct from non-reflexive pronouns) in the 1st or 2nd person, then it has distinct reflexive pronouns in the 3rd person.

We need a number of languages to check whether this is true.

[GENERATIVE UNIVERSAL from the GB era]: given an X' structure for syntactic phrases of the kind in (6), for all phrase types in a given language either the specifier precedes (a) or it follows (b) the head



Comrie's consideration : this universal was proposed on the basis of English.

BUT: in Malay determiners follow their noun while auxiliaries precede their verb.

Hence: this is a tendency, not a universal.

We will come back to this objection later.

TWO MAJOR APPROACHES TO LINGUISTIC VARIATION

1. the typological approach
2. the generative approach

Comrie (1981): two main differences between the two approaches

- the data base for research on language universals
- the degree of abstractness of analysis required to state language universals
- the kinds of explanations advanced for the existence of language universals

In this class, we wish to concentrate on the **similarities** between the two approaches, on what they have in common.

ACCOUNTING FOR UNIVERSALS

Many possible accounts

- no need to account for anything: this is just accidental and the universal we see are only apparent
- historical approach: universals are there because all languages have a common origin
- areal approach: universals are there because of language contact
- functional approach (usually adopted by typologists): these universals are due to communicative and/or processing needs
- formal approach (adopted by generativists): these universals derive from innate grammar

NB: formal here means 'related to the form of rules', but also 'stated in mathematically precise terms as in formalized'

2. APPROACHES TO LINGUISTIC VARIATION: THE TYPOLOGICAL APPROACH

2.1. THE DATA BASE

(5) 'surely in order to establish that something is universal in language one would need to look at a wide range of languages'. [Comrie 1981:2]

(6) 'Language universals would be those innate linguistic principles which facilitate the child language-learning task' [Comrie 1981:4]

analyzing the chemical properties of iron vs analyzing human behaviour under stress

How do we select a language sample?

Bell (1978): one should ensure that each group of languages should be given equal representation, where a group is defined as a genetically related set of languages separated from their common ancestor by a time-depth of 3500 years.

2.2. CLASSIFICATION OF LANGUAGE UNIVERSALS

SUBSTANTIVE UNIVERSALS: categories that are present in all languages (e.g. verb, noun, noun phrase, subject, ...)

FORMAL UNIVERSALS: statements on the form of grammar rules (e.g. no language can form questions by simply inverting the word order in a sentence)

IMPLICATIONAL UNIVERSALS: a property of a given language is present if another property is also present

NON-IMPLICATIONAL UNIVERSALS: properties that are found (or not found) in natural language without reference to any other property

ABSOLUTE UNIVERSALS: exceptionless universals

UNIVERSAL TENDENCIES: universals that exist as a tendency but still have exceptions

These classes of universals can be combined:

- absolute + non implicational universal: all languages have vowels
- absolute implicational universal: if a language has 1st/2nd person reflexives, then it has 3rd person reflexives
- non-implicational tendencies: nearly all languages have nasal consonants (but some Salishan languages don't, for instance)
- implicational tendencies: **if a language has SOV basic word order, it will probably have postpositions** (but Persian, SOV, has prepositions)

Possible explanations for language universals

- i. common genetic origin
- ii. external explanations (human genetic predisposition; correlation with other aspects of human cognitive psychology; functional explanation)

2.2.1. LANGUAGE TYPES

Language types based on word order [Greenberg (1963), based on work by Lepsius (1880) and Schmidt (1926)]

- (7) VO: Aux-V, Prepositions, Noun-Adjective, Noun-Genitive, Noun-Relative
OV: V-Aux, Postpositions, Adjective-Noun, Genitive-Noun, Relative Noun

HARMONIC AND DISHARMONIC WORD ORDERS

WALS (World Atlas of Language Structures) [Haspelmath, Dryer, Gil & Comrie 2005]

- (8) Correlation between OV/VO and adpositional structure in the 862 languages surveyed:

OV & Postpositions	427 (47,87%)
VO & Prepositions	417 (46,75%)
OV & Prepositions	10 (1,12 %)
VO & Postpositions	38 (4,26%)

94% of the languages surveyed are "harmonic"

2.3. FUNCTIONAL APPROACHES TO LANGUAGE VARIATION

Ease of processing

Prominence scales (see also OT accounts --> last class)

Scales of processing difficulty

- (9) "More frequent items tend to be shorter than rarer items, and if one member of an opposition is zero, processing ease dictates that it should be the more frequent one" [Haspelmath 1008:98]

Ex: objects tend to be inanimate and indefinite, so languages tend to restrict overt case marking to the rarer cases, definite and human objects.

But: Jaeggli's generalization?

(10) *Condition on Arbitrary Pronominals (CAP)*

Overt pronouns may not be arbitrary in reference iff the overt/empty alternation obtains. [Jaeggli 1986:66]

FREQUENCY EFFECTS

Another example: the more frequent person-role combinations (especially Recipient $\frac{1}{2}$ + Theme 3) tend to occur as bound forms, whereas the rarer combinations often do not (Haspelmath 2004)

Rarity of use may lead to the need for a special form to help the hearer in the interpretation (it is rare that Subj and Obj are coincident. When this happens we mark "reflexivity")

2.3.1. THE PERFORMANCE-GRAMMAR CORRESPONDENCE HYPOTHESIS (PGCH)

(11) "Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by distributional patterns of selection in corpora and by ease of processing in psycholinguistic experiments" [Hawkins 2004:3]

(12) Grammatical predictions of the PGCH:

- a. If a structure A is preferred over an A' of the same structural type in performance, then A will be more productively grammaticalized, in proportion to its degree of preference; if A and A' are more equally preferred, then A and A' will both be productive in grammars.
- b. If there is a preference ranking $A > B > C > D$ among structures of a common type in performance, then there will be a corresponding hierarchy of grammatical conventions (with cut-off points and declining frequencies of languages).
- c. If two preferences P and P' are in (partial) opposition, then there will be variation in performance and grammars, with both P and P' being realized,

each in proportion to its degree of motivation in a given language structure. [Hawkins 2004:6]

Ex.

Tendency of heads to precede their complements:

V-NP, P-NP, A of NP, N of NP

Light-before-heavy tendency involves more than a head-complement relation. The canonical order of VP constituents is lighter-to-heavier:

(13) V-NP-PP-CP (*convince my students of the fact that all grammars leak*)
[Newmeyer 2004:217]

Single adjectives and participles can appear in pre-head position:

(14) a. a silly proposal

But: if these adjectives are "heavy" (i.e. if they have complements), there is heavy XP shift

(15) a. * a sillier than any I've ever seen proposal
b. a proposal sillier than any I've ever seen

[Newmeyer 2004:217]

All other things being equal, the greater the length differential between the two Aps, the more likely the speakers will be to put the shorter one first.

The three EFFICIENCY PRINCIPLES

(16) MINIMIZE DOMAINS (MID)

The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combinations and/or dependencies are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain [Hawkins 2004:32]

Do generativists necessarily exclude parsing considerations?

3. THE UNIVERSAL GRAMMAR APPROACH

Chomskyan linguistics = a species of formal linguistics → **GENERATIVE** (vs structuralist, functional, etc.); other types of generative syntax: Generalized Phrase Structure Grammar (GPSG); Head-Driven Phrase Structure Grammar (HPSG); Lexical Functional Grammar (LFG); Optimality Syntax (OT syntax) ...

What does it mean for syntax/grammar to be generative?

A generative grammar must be able to generate all possible sentences in a given language and only those. It appeals to linguistic **COMPETENCE** and almost completely disregards **PERFORMANCE**.

3.1. GENERAL FEATURES OF GENERATIVE GRAMMAR

- A (generative) theory of grammar must be **FULLY EXPLICIT**, i.e. give enough information to enable us to account for the generation of all and only the structures that are grammatical in a given language (but see Haspelmath observation about the explicitness of parameters)

Moreover, a generative theory of grammar must meet three **LEVELS OF ADEQUACY** (from Biberauer 2006):

- **OBSERVATIONAL** adequacy: a theory that accounts for the data in a given corpus, (i.e. for available positive evidence) but nothing beyond. No predictive power
- **DESCRIPTIVE** adequacy: a theory that accounts for both corpus data (i.e. positive evidence) and native-speaker intuitions about grammaticality (i.e. positive and negative evidence)
- **EXPLANATORY** adequacy: a theory that is descriptively adequate and, additionally, also accounts for how children could have acquired this grammar.

Chomskyan syntax strives towards explanatorily adequate grammars (and, more recently, beyond – cf. the title of Chomsky's (2004) "Beyond Explanatory Adequacy", in which he suggests syntacticians should "seek a level of

explanation deeper than explanatory adequacy, asking not **WHAT** the properties of language are, but **WHY** they are that way" [Chomsky 2004: 105])

3.1.2. THE METHODOLOGY OF CHOMSKYAN GENERATIVE SYNTAX

characteristics common to all instantiations of generative syntax (from Biberauer 2006):

- commitment to the **SCIENTIFIC METHOD**: hypothesis construction → confrontation with data → hypothesis revision → confrontation with data, etc ... with hypotheses about linguistic structure having to be precise/explicit enough to be testable and, hence, falsifiable
- object of study: the unconscious knowledge that underlies ordinary language use, i.e. our **MENTAL GRAMMAR**

Chomskyan terms for this knowledge = **COMPETENCE/I-LANGUAGE** (I=internal); contrast **PERFORMANCE/E-LANGUAGE** (E=external)

- a distinguishing feature of Chomskyan generative syntax is the commitment to the **INNATENESS HYPOTHESIS**, i.e. to the idea that there is a biological basis – Universal Grammar (UG) – that underlies our linguistic abilities and that makes us (uniquely and, under normal circumstances, without exception) able to acquire language in the first place.

The similarities we observe between languages are attributable to the fact that all human beings share a Universal Grammar.

From a performance/processing view (functionalist) to an "internal" model of the competence. Is this mismatch of point of views necessarily bad, as many claim?

PHRASE STRUCTURE RULES (TRANSFORMATIONAL GRAMMAR)

Aimed at descriptive adequacy : the focus was on the differences between languages, all of which were meticulously described and "derived" via essentially language-specific rules (phrase structure rules and transformations).

(Chomsky 1973)

- Verb-movement parameter

...

Remember: UG=Principles and parameters

Remember: variation is encoded in the UG

As opposed to: variation emerges in languages because of processing/performance reasons

3.1.3. WHY UG?

(20) "Despite the apparent complexity of human language(s), children acquire them easily under less than optimal conditions and regardless of general intelligence. By contrast, we (mostly!) have to work hard at learning artificial languages, mathematical systems, etc. and some people NEVER succeed in learning these." [Biberauer 2006: 2]

Chomsky's proposed explanation: humans are born with a special language faculty (UG/LAD) which specifies that part of linguistic structure that all languages have in common, i.e. all languages share some of their structure and babies don't need to learn this because it's already "hard-wired" into their brains at birth. Specifically, the innate knowledge that babies are born with is supposed to be largely syntactic.

(21) "By studying the properties of natural languages ... we may hope to gain some understanding of the specific characteristics of human intelligence."
Chomsky (1975: 4-5)

Questions (and objections that functionalists pose to UG):

- What kind of content do we have to postulate for UG in order to concretise the Innateness Hypothesis?
- What is the form of parameters?
- How many parameters are there?

But notice something important:

Native speakers of a language unconsciously know many things about their language. They know, for instance:

1. which combinations are **grammatical** and which not (*):

- (22) a. *In the future everyone will be famous for fifteen minutes*¹
b. * *In the future will everyone for 15 minutes famous be*
c. * *In the future will be everyone famous for 15 minutes*
d. * *Future, everyone minutes 15 for famous be will*

2. which words "go together" and which don't:

- (23) a. He **looked up** the meaning of syntax
b. He **looked** the meaning of syntax **up**

- (24) a. He **looked up** the chimney
b. *He **looked** the chimney **up**

I-language: the body of unconscious language knowledge that speakers have. And it is this that Chomskyan syntacticians aim to characterise and explain.

3.1.4. THE DATA BASE

Grammaticality judgments

Real production corpora

Let's go back to typology:

(25) **Distribution of S(ubject), V(erb) and O(bject) in the world's languages**

SOV (*Laura flowers bought*): 44.78%

SVO (*Laura bought flowers*): 41.79%

VSO (*Bought Laura flowers*): 9.2%

¹ Attributed to Andy Warhol. This specific set of examples is borrowed from Roberts (1997).

- OVS** (*Flowers bought Laura*): 1.24%
OSV (*Flowers Laura bought*): (practically) 0%
VOS (*Bought flowers Laura*): 2.99%

[Tomalin 1986]

Recall also:

(26)	OV & Postpositions	427 (47,87%)
	VO & Prepositions	417 (46,75%)
	OV & Prepositions	10 (1,12 %)
	VO & Postpositions	38 (4,26%)

- (27) VO → prepositions
→ auxiliaries preceding V
→ possessor (i.e. genitive) following N
→ adjectives follow N

- b. OV → postpositions
→ auxiliaries following V
→ possessor (genitive) preceding N
→ adjectives precede N

- (28) The Head Parameter
Heads uniformly precede their complements or uniformly follow them
[Smith 1989, Freidin 1992, Chomsky & Lasnik 1993 and many others]

or:

- (29) Complements are to the left or to the right of the head
[Newmeyer 2004:183]

Head left (head-complement): English, Swahili, ...

Head right (comp-head): Japanese, Lakhota, ...

But: this parameter runs into empirical problems. > TOMORROW

Question: what is the right level of abstraction to capture the factors that determine linguistic universals?

REFERENCES & READING SUGGESTIONS

- Bell, A. (1978). Language samples. In: Greenber & al. (eds), *Universals of human language*. Stanford: Stanford University press. I: 123-156.
- Biberauer, T. (2006). *Syntax – Lecture notes*. University of Cambridge.
- Bloomfield, L. (1934). Review of *Handbuch der erklärenden Syntax* by Wilhelm Havers. *Language* 10: 32-39.
- Chomsky, N. (1973). Conditions on transformations. In Anderson & Kiparsky (eds), *A Festschrift for Morris Halle*. New York: Holt, Rinehart & Winston, 232-286.
- Chomsky, N. (1975). *Reflections on language*. New York: Pantheon.
- Chomsky, N. (1981). *Lectures on Government and Binding Theory*. Dordrecht: Foris.
- Chomsky, N. (2004). Beyond Explanatory Adequacy. In Belletti, A. (ed.) *Structures and Beyond. The Cartography of Syntactic Structures* vol 4. Oxford: OUP, 104-131.
- Chomsky, N. & H. Lasnik (1993). The Theory of Principles and Parameters. In: Jacobs & al (eds), *Syntax: An International Handbook of Contemporary Research*. Berlin: Mouton de Gruyter, 104-131.
- Comrie, B. (1981). *Language Universals and Linguistic Typology*. Chicago: University of Chicago Press.
- Freidin, R. (1992). *Foundations of generative syntax*. Cambridge, Mass: MIT Press.
- Greenberg, J. (1955). *Studies in African Linguistic Classification*. New Haven: Compass Publishing Company.
- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (ed.). *Universals of Language*. Cambridge, MA: MIT Press, 58-90.
- Haspelmath, M. (2004). Does linguistic explanation presuppose linguistic description? *Studies in language* 28: 554-579.
- Haspelmath, M. (2008). Parametric versus functional explanations of syntactic universals. In: Biberauer, T. (ed.), *The Limits of Syntactic Variation*. Amsterdam: John Benjamins, 75-108.
- Haspelmath, M. Dryer, M. Gil, D & B. Comrie (2005). *Word Atlas of Language Structures (WALS)*. Oxford: OUP.
- Hawkins, J. (2004). *Efficiency and Complexity in Grammars*. Oxford: OUP.
- Jaeggli, O. (1986). Arbitrary plural pronominals. *Natural language and linguistic theory* 4, 43-76.

- Joos, M. (1957). *Readings in Linguistics: The Development of Descriptive Linguistics in America since 1925*. Washington DC: American Council of Learned Societies.
- Katamba, F. (1993). *Morphology*. Houndmills: Palgrave.
- Lepsius, K. (1880). *Nubische Grammatik mit einer Einleitung über die Völker und Sprachen Afrika's*. Berlin: Verlag von Wilhelm Hertz.
- Newmeyer, F. (2004). Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook* 4: 181–234.
- Rizzi, L. (1990). *Relativized Minimality*. Cambridge, Mass.: MIT Press.
- Sapir, E. (1921). *Language*. New York: Harcourt Brace and Co.
- Schlegel, A. (1817). *Über dramatische Kunst und Litteratur, Grundzüge einer Kultur- und Völkergeschichte Alteuropas*. 2nd edition, 3 volumes. Heidelberg: Mohr and Winter.
- Schleicher, A. (1861-2). *Compendium der vergleichenden Grammatik der indogermanischen Sprachen. Kurzer Abriss einer Laut- und Formenlehre der indogermanischen Ursprache, des Altindischen, Altiranischen, Altgriechischen, Altitalischen, Altkeltischen, Altslawischen, Litauischen und Altdeutschen*, 2 volumes. Weimar: Böhlau.
- Schmidt, W. (1926). *Die Sprachfamilien und Sprachenkreise der Erde*. Heidelberg: Carl Winter.
- Smith, N. (1989). *The twitter machine: Reflections on language*. Oxford: Blackwell.
- Tomalin, M. (2006). *Linguistics and the Formal Sciences. The Origins of Generative Grammar*. Cambridge: CUP.