

Fiat Lingua

Title: Using Language Invention to Teach Typology and Cross-Linguistic Universals

Author: Matt Pearson

MS Date: 03-17-2017

FL Date: 04-01-2017

FL Number: FL-000043-00

Citation: Pearson, Matt. 2017. "Using Language Invention to Teach Typology and Cross-Linguistic Universals." FL-000043-00, *Fiat Lingua*, <<http://fiatlingua.org>>. Web. 01 April 2017.

Copyright: © 2017 Matt Pearson. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.



<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Using Language Invention to Teach Typology and Cross-Linguistic Universals

Matt Pearson

pearsonm@reed.edu
Reed College

Abstract

I discuss a project where students learn about language typology by creating a naturalistic constructed language. Students review cross-linguistic variation in natural languages (in areas such as phoneme inventory, word order, case alignment, etc.), and then determine which properties their invented language will have. Decisions are made at random by spinning a wheel. Attached to the wheel is a pie chart, where the size of each slice represents the percentage of the world's languages possessing a given setting for some structural parameter or set of parameters. Crucially, each decision constrains subsequent decisions in accordance with known implicational universals: e.g., in determining whether the language has prepositions or postpositions, the pie chart is adjusted based on verb-object order in the language, as decided by a previous spin of the wheel.

1. Overview

In this paper I describe a linguistics class project where students collaborate to create a typologically plausible constructed language (*conlang*), using a semi-randomized procedure to determine structural features such as phoneme inventory, word order, and case alignment. The goal of the project is to investigate typology in a hands-on, creative way, using language invention to teach students about cross-linguistic variation and the nature of implicational universals and markedness hierarchies (Greenberg 1963, Hawkins 1983, Croft 2002, et al.). By working together to invent a conlang that conforms to known statistical tendencies, students gain exposure to the idea that natural languages do not vary randomly in their structure, but instead follow predictable patterns.

The idea for this project originated with Susan Curtiss as part of a UCLA undergraduate honors course on mental grammar. When I was a graduate student at UCLA in the 1990s, I worked as a teaching assistant for this course and helped to develop the project as the basis for an independent typology-themed discussion section, following a model established by a previous TA. In describing the project here, I have refined our original procedure somewhat to reflect the increased availability of numerical data from large language samples, most especially the World Atlas of Language Structures (WALS) online database (Dryer and Haspelmath 2013).

In section 2 I lay out the general parameters for the exercise, and discuss some of the sources that instructors can use to generate questions and topics related to morpho-syntactic variation. In section 3 I describe how the structural parameters of the conlang are determined using a semi-randomized procedure that reproduces the statistical patterns expressed by implicational universals. In section 4 I briefly discuss how students work to flesh out the conlang by adding vocabulary, cultural details, etc.

2. Focus of the project

The project may be carried out over several class periods, or even an entire semester. It can form a unit or mini-course within a larger discussion-based class, or the discussion section activity within a lecture course. It can also be adapted for use outside the classroom as part of an extra-curricular activity, independent study, or typology workshop.

The project is divided into units, with the number of units depending on the topic of the class and the time available. Each unit focuses on a domain of cross-linguistic variation for which language universals have been proposed. At the beginning of each unit, the instructor and students review data from natural languages to learn about the range of cross-linguistic variation, along with any implicational generalizations which have been proposed to capture patterns of covariation. Some types of generalizations which might be covered, with an example of each, are listed in (1):

- (1)
 - a. *Phonological implicational universals*: If a language has voiceless nasals and approximants, it will also have their voiced counterparts (Maddieson 1984).
 - b. *Word order correlations*: If a language has OV order, it is highly likely to have postpositions; if it has VO order, it is highly likely to have prepositions (Dryer 1992).
 - c. *Implicational universals related to grammatical categories*: If a language has a grammatical gender system, it will also make a number distinction (Greenberg 1963).
 - d. *Markedness asymmetries*: If a language has a paucal number, it will also have a dual number; if a language has a dual number, it will also make a singular/plural distinction. In languages that make grammatical number distinctions morphologically, plural marking involves at least as much morphology as singular marking, and dual/paucal marking involves at least as much morphology as plural marking. These generalizations implicate a markedness hierarchy for number: PAUCAL > DUAL > PLURAL > SINGULAR. (Croft 2002, Corbett 2000)

After learning about the patterns of cross-linguistic variation and markedness found in natural languages, students must then decide which grammatical features their conlang will have. Essentially this involves generating a series of 'design questions' which must be answered – or, in Generative Grammar terms, a series of *parameters* which must be set for the conlang. In (2) below I list some possible topics for units, along with sample design questions for each unit. This list is by no means exhaustive; it is merely meant to suggest the range of decisions that must be made in developing a grammatically fleshed-out conlang.

- (2)
 - a. *Phoneme inventory* (Maddieson 1984):
 - Which vowels will the language have and how will they be arranged in the vowel space? (E.g., will the language have a 3-vowel system, a 5-vowel system, etc.?)
 - Which places and manners of articulation will be distinguished among the consonants?
 - Will [VOICE] be a contrastive feature of consonants? Will [NASAL] be a contrastive feature for vowels?
 - Will length be a distinctive feature of consonants and/or vowels?
 - Will the language have tones?

- b. *Syllable structure and phonotactic constraints* (Blevins 1995):
- Will the language permit codas? (I.e., will the language include both open and closed syllables, or open syllables only?)
 - Will the language permit complex onsets? If so, what combinations of consonants are allowed?
 - If coda consonants are permitted, will there be restrictions on which classes of consonants can occur in coda position? Will the language allow complex codas?
- c. *Morphological types* (Comrie 1989, Baker 1996, Nichols 1986):
- Will the language be analytic or synthetic? If it is synthetic, will it include features of polysynthesis (such as noun incorporation)?
 - If the language is (poly)synthetic, will it have strictly agglutinating morphology, or will it also include fusional and non-concatenative morphology (e.g., templatic morphology)?
 - If the language is (poly)synthetic, will it be primarily suffixing, primarily prefixing, or a combination of both?
 - Will the language have reduplication? If so, what functions will be associated with reduplication (plurality, imperfective aspect, etc.)?
 - To express argument-structure dependencies (case roles), will the language make use of head marking (indexation/agreement), dependent marking (morphological case), both, or neither?
- d. *Lexical and functional categories* (Schachter and Shopen 2007, Harley and Ritter 2002):
- Will the language include a category of adjectives, as distinct from nouns and verbs?
 - Will the language have a category of adpositions, or will it express relational concepts by other means?
 - What sort of pronominal/phi-feature system will the language have? Which person, number, and gender/animacy distinctions will be made by the pronouns and agreement markers?
 - Will the language have articles?
 - Will the language have a tense category? If so, what kinds of tense distinctions will it make?
- e. *Basic constituent order* (Greenberg 1963, Hawkins 1983, Dryer 1992, 2007):
- Will the language have OV or VO as its unmarked order for verb and object, or will both orders be allowed (non-configurationality)?
 - If the language is VO, will the subject precede or follow the verb (SVO, VSO, VOS)?
 - If the language has an adposition category, will it be prepositional or postpositional?
 - If the language has an adjective category, will attributive adjectives precede the modified noun (Adj-N) or follow the modified noun (N-Adj) within the noun phrase?
 - Will relative clauses precede the modified noun (Rel-N) or follow the modified noun (N-Rel)?
 - In possessive noun phrases, will the possessor precede the possessed noun (Poss-N) or follow the possessed noun (N-Poss)?
 - Will the language employ wh-movement or wh-in-situ in questions?

- f. *Marking syntactic dependencies* (Comrie 1978, Palmer 1994, Andrews 2007):
- If the language has morphological case and/or agreement for core arguments, will it have an accusative alignment, an ergative alignment, a split-S alignment, or a tripartite marking system?
 - If the language is ergative, will it be strictly ergative, split-ergative based on tense/aspect, split-ergative based on a person/animacy hierarchy, etc.?
 - If the language has case marking, what non-core cases, if any, will it have (dative, instrumental, locative, etc.)?
 - If the language has grammatical number (and gender) categories, will attributive adjectives and other modifiers agree in features with the noun they modify?

There are a number of useful resources for generating topics and design questions and identifying related language universals. The WALS database website, *wals.info* (Dryer and Haspelmath 2013), includes a comprehensive list of articles on topics related to cross-linguistic variation. More in-depth discussions of some of these topics can be found in typology textbooks and reference works such as Maddieson (1984), Shopen (1985, 2007), Comrie (1989), Payne (1997), Croft (2002), Velupillai (2012), and Moravcsik (2013). The Lingua Descriptive Studies questionnaire (Comrie and Smith 1977) provides a comprehensive list of topics and questions related to grammatical structure. Although it is intended as an outline for field linguists writing a reference grammar of a natural language, this questionnaire can easily be repurposed for determining the grammatical features of a conlang. A similar questionnaire, but more condensed and focused on morpho-syntactic topics, is included in Payne (1997). Finally, the University of Konstanz's Universals Archive offers an extensive database of proposed language universals, with citations to works in which those universals are proposed or discussed. This archive can be accessed at <http://typo.uni-konstanz.de/archive/intro/index.php>.

3. Deciding on the features of the language

As originally conceived by Susan Curtiss, the major structural properties of the conlang are decided at random by spinning a wheel. The intention behind this choice was, in part, to introduce a game-like element of chance into the creation of the language. By leaving certain decisions up to chance, students are forced to exercise their creativity within predetermined structural constraints, avoiding any temptation to make their conlang either overly familiar and English-like, or overly alien and exotic (either of these would defeat the purpose of an exercise focused on learning about natural language typology).

At UCLA we used a carnival-style wheel built for this purpose by Curtiss's husband. The wheel was mounted vertically on a wooden frame with a flexible pointer attached to the top. When the wheel was spun, the pointer would brush against nails protruding from the circumference of the wheel until it came to rest at a particular point (cf. the spinning wheels from the game shows *Wheel of Fortune* and *The Price is Right*). For instructors who lack the resources to build or purchase a carnival wheel, a large spinner from a board game like *Twister* can also be used, or an online application such as Wheel Decide: <http://wheeldecide.com>. (Other random selection methods are also possible, such as rolling a set of multi-sided dice from a role-playing game. Here, however, I focus on using a wheel for this purpose.)

For each grammatical parameter to be specified, a pie chart is attached to the spinner wheel with slices representing features or settings for the parameter in question. The size of each slice is determined by the rough proportion of the world's languages that have that parameter setting (in combination with other parameter settings, where relevant). When the project was originally developed, these proportions were estimated. Now, however, statistics from

the WALS database (Dryer and Haspelmath 2013) or similar sources may be used for this purpose. For each decision to be made, a student volunteer comes to the front of the room and spins the wheel. Whichever slice of the pie chart the pointer lands on determines the parameter setting that the conlang will have.

For example, suppose that the students are studying case/agreement systems and need to determine which core *alignment system* their conlang will have—that is, how subjects of intransitive clauses (S), subjects of transitive clauses (A), and objects of transitive clauses (P) will be differentiated. Among natural languages, the four most common alignments are *accusative* (A and S are marked the same way while P is marked differently), *ergative* (P and S have the same marking while A is marked differently), *tripartite* (A, S, and P are all marked differently), and *split-S* (more ‘agentive’ S arguments are marked the same way as A arguments, while more ‘patientive’ S arguments are marked the same way as P arguments). (See Comrie 1978, Dixon 1994, Palmer 1994, Coon et al. 2017.) The relative prevalence of each alignment type among the world’s languages is suggested by the numbers in (3). The *N* values in the second column of the table are taken from the WALS sample of languages with morphological marking of core case roles (Dryer and Haspelmath 2013). In the third column these *N* values have been converted into approximate whole-number percentages (for the sake of simplicity, languages with alignments other than accusative, ergative, tripartite, and split-S are excluded). Using these percentages, we can construct a pie chart where each slice represents a different alignment type, as shown in Figure 1. This chart is attached to the wheel, the wheel is spun, and whichever slice of the pie the pointer lands on determines which alignment the conlang will have.

(3)

	<i>N</i>	<i>approx %</i>
nominative-accusative (A+S vs. P)	52	57
ergative-absolutive (A vs. S+P)	32	35
tripartite (A vs. S vs. P)	4	4
split-S (A+S _A vs. P+S _P)	4	4
TOTAL	92	100

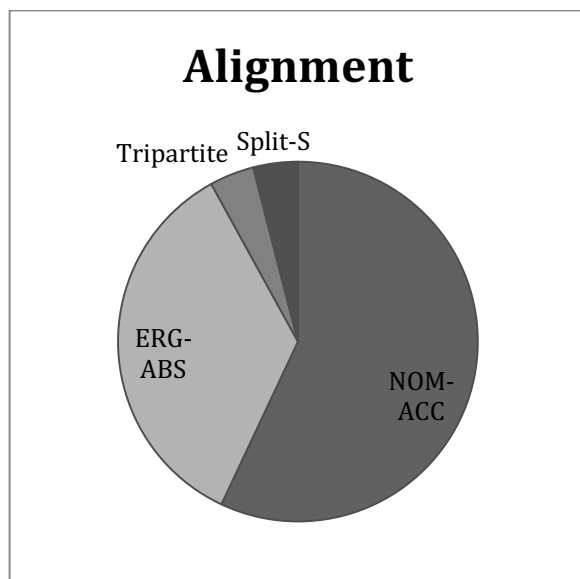


Figure 1. Spinner wheel for determining core case alignment

The probability that the conlang will end up with a given feature as a result of a random spin is determined by the relative size of the corresponding pie slice. In the case of the alignment parameter, the wheel in Figure 1 ensures a greater chance that the pointer will land on the accusative setting rather than the ergative setting, and a greater chance that it will land on the ergative setting rather than the tripartite or split-S settings. In this way, the probability that the conlang will end up with a given alignment roughly mirrors the probability that a randomly chosen natural language will have that alignment.

Crucially, as each parameter of the conlang is set, the probabilities governing subsequent decisions are adjusted based on known statistical correlations. In other words, each spin of the wheel determines which pie charts will be used for subsequent spins, providing a graphic illustration of how parameter settings in natural languages are interdependent. For example, suppose that previous spins of the wheel have determined that the conlang has a default order for verb and object in basic transitive clauses, either OV or VO, and that it has an adposition category. Will the conlang have prepositions or postpositions? Dryer (1992) and others have shown that, overwhelmingly, OV languages are postpositional while VO languages are prepositional. The table in (4) gives data supporting this generalization, with *N* values taken from WALs (Dryer and Haspelmath 2013). The number and percentage of OV languages in the WALs database with postpositions (Postp) is vastly larger than the number and percentage with prepositions (Prep); conversely, the VO languages with prepositions vastly outnumber those with postpositions.

(4)

	<i>N</i>	<i>approx</i> %
OV and postpositions	472	97
OV and prepositions	14	3
TOTAL	486	100
VO and postpositions	42	8
VO and prepositions	456	92
TOTAL	498	100

The strong statistical correlation between verb-object order and adposition type can be expressed as a pair of biconditional universals: **OV ↔ Postp** and **VO ↔ Prep**.

As a consequence of this correlation, the pie chart for determining adposition type in the conlang will look very different depending on how the OV/VO parameter had been set by previous spins of the wheel. If the conlang has OV order, the pie chart for determining adposition type would look like Figure 2 below, resulting in a 97% chance that a random spin of the wheel will yield postpositional order. On the other hand, if the parameter has been set to VO, the pie chart for determining adposition type would look like Figure 3. Here, a random spin of the wheel would have a 92% chance of landing on the preposition setting.

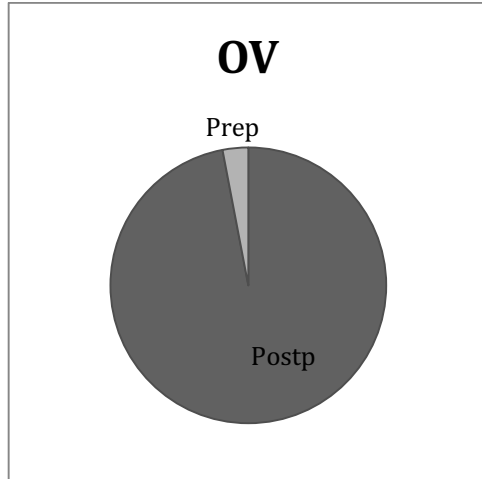


Figure 2. Spinner wheel for determining adposition type (language is OV)

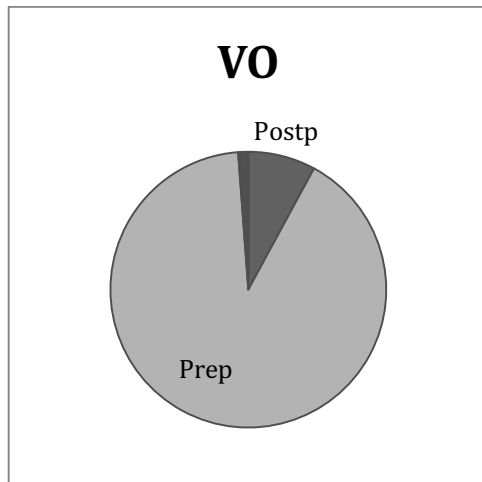


Figure 3. Spinner wheel for determining adposition type (language is VO)

A somewhat different type of word order correlation is exemplified by the table (5), showing the relationship between OV/VO order and the position of a relative clause with respect to the noun that it modifies (*N* values from Dryer and Haspelmath 2013). As (5) shows, VO languages overwhelmingly place the relative clause after the noun (N-Rel). OV languages, on the other hand, show no such correlation: they are roughly equally likely to have either order (N-Rel or Rel-N) (cf. Dryer 1992).

(5)

	<i>N</i>	<i>approx</i> %
OV and Rel-N	132	54
OV and N-Rel	113	46
TOTAL	245	100
VO and Rel-N	5	1
VO and N-Rel	416	99
TOTAL	421	100

In the case of adposition type, the correlation with verb-object order is bidirectional. If a language has OV order it is highly likely to have postpositions, and conversely, if it has postpositions it is highly likely to have OV order; likewise for VO order and prepositions. In the case of relative clauses, however, the correlation with verb-object order is only partial. If a language has VO order, it is overwhelmingly likely to have postnominal relative clauses (N-Rel), whereas no prediction can be made regarding the placement of relative clauses in OV languages. On the other hand, if a language has prenominal relative clauses (Rel-N), it is overwhelmingly likely to also have OV order, while no prediction on the order of verb and object can be made for languages with postnominal relatives. This can be expressed as a pair of unidirectional implicational universals: $VO \rightarrow N\text{-Rel}$ and $Rel\text{-N} \rightarrow OV$. In both cases the converse implication does not hold.

Figures 4 and 5 below show the pie charts, derived from the data in (5), which would be used when spinning the wheel to determine the order of noun and relative clause in the conlang. If the conlang is OV, there is a roughly equal chance that the pointer will land on N-Rel (46%) or on Rel-N (54%). If the conlang is VO, however, the probability that the pointer will land on N-Rel is overwhelming (99%).

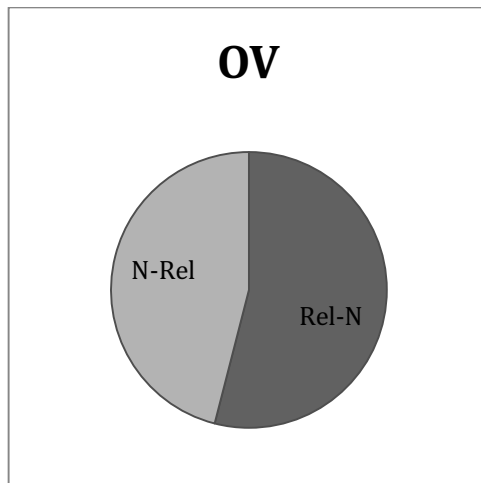


Figure 4. Spinner wheel for determining order of noun and relative clause (language is OV)

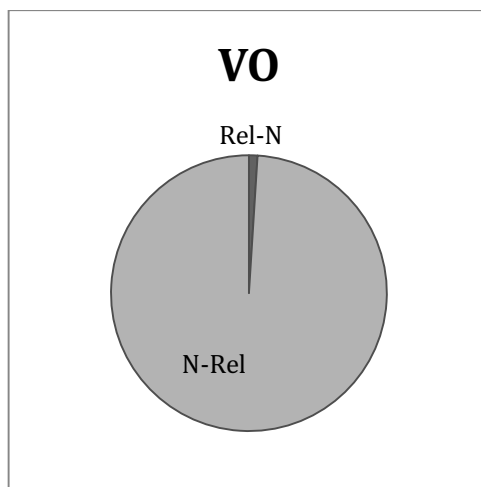


Figure 5. Spinner wheel for determining order of noun and relative clause (language is VO)

This exercise gives students the opportunity to investigate language universals in an original and hands-on manner. Typological research has shown that natural languages do not vary arbitrarily in their structure, but can instead be grouped into types (word order types, morphological types, etc.) based on shared formal properties. By exploring how the setting of each structural parameter in their ‘naturalistic’ conlang influences subsequent parameter settings, students learn how the statistical tendencies described by implicational universals give a language its overall grammatical ‘shape’. Furthermore, they are invited to speculate about *why* certain grammatical features tend to cluster together in natural languages while others do not—whether this reflects external functional constraints on language design, arbitrary properties of Universal Grammar (UG), or other factors.

4. Fleshing out the conlang

A language is more than just a set of categories and parameter settings: it also has a vocabulary, a cultural context, etc. It is in working together to flesh out of these aspects of the conlang that students are able to exercise their collective creativity.

When we carried out this exercise at UCLA, grammatical morphemes were typically invented in class, as needed. For example, if successive spins of the wheel determined that the language had ergative and absolutive case suffixes, the instructor would solicit suggestions for the phonological shape of the suffixes from the students. We would also coin lexical roots as needed during class sessions so that we could practice the features of the conlang by translating example sentences from English.

To further expand the vocabulary, students were required to propose five new roots per week (with definitions) and submit their suggestions to the instructor. Back when this exercise was developed, we asked students to include a list of roots as part of their weekly homework assignments. Nowadays, to avoid students suggesting different lexical items for the same concept, vocabulary suggestions could be uploaded to an online platform, either a course website or a wiki. For each round of submissions, proposed roots are reviewed by the instructor, or a designated student, to make sure they conform to the phonological constraints of the language (e.g., to make sure that only permitted phonemes and syllable types are used). Approved roots are then added to the dictionary for the language, which is updated and distributed to students once a week.

In addition, instructors can assign students to invent derivational morphemes (or derivational processes such as reduplication and ablaut), and practice creating new lexical items from pre-existing roots. Including productive derivation allows the vocabulary to expand quickly by giving students and instructors the means to coin new words on the fly.

Finally, students can be invited to expand the grammar and cultural background of the conlang through term paper projects. For example, students may be required to invent and describe some grammatical feature of the conlang which was not specified as part of the group project (the description could take the form of a sample chapter from a reference grammar of the language, which explains the feature concisely and illustrates it with glossed example sentences). Alternatively, students could be asked to construct a sample text in the language (e.g., a folk tale, dialogue, etc.), including a free English translation, interlinear glosses, and supplemental vocabulary. The need to write texts in the conlang, or translate texts from English, may lead students to propose and discuss additional grammatical constructions not covered in class.

5. Summary

I have described a collaborative project where students work together to invent a typologically plausible language. The phonological, morphological, and syntactic features of the language are determined at random by spinning a wheel to decide among a set of alternatives (e.g., for core argument alignment, the alternatives might be *accusative*, *ergative*, *tripartite*, and *split-S*). In each case the alternatives are weighted in accordance with their prevalence among the world's languages, as extrapolated from numerical data taken from the WALS database and other large language samples. In this way, the probability that the conlang will end up with a particular grammatical feature as a result of a random spin approximates the probability that an arbitrarily chosen natural language will have that feature.

In cases where different grammatical features correlate with one another, the result of a given spin will determine how the alternatives for subsequent spins are weighted. For example, since adposition type correlates with the default order of verb and object, the probability that the conlang is assigned prepositional (vs. postpositional) order as a result of a random spin of the wheel is adjusted based on whether the conlang is OV or VO, as determined by a previous spin. Through this exercise, typology students are able to study implicational universals firsthand by seeing how those universals play out in constraining the development of a 'naturalistic' conlang grammar.

References

- Andrews, A. 2007. The major functions of the noun phrase. *Language Typology and Syntactic Description* (2nd ed.), vol. I: *Clause Structure*, ed. T. Shopen, pp. 132-222. Cambridge: Cambridge University Press.
- Baker, M. 1996. *The Polysynthesis Parameter*. Oxford: Oxford University Press.
- Blevins, J. 1995. The syllable in phonological theory. *The Handbook of Phonological Theory*, ed. J. Goldsmith, pp. 206-244. Oxford: Blackwell.
- Comrie, B. 1978. Ergativity. *Syntactic Typology*, ed. W.P. Lehmann, pp. 329-394. Austin: University of Texas Press.
- Comrie, B. 1989. *Language Universals and Linguistic Typology* (2nd ed.). Chicago: University of Chicago Press.
- Comrie, B. and N. Smith. 1977. Lingua descriptive studies: Questionnaire. *Lingua* 42: 1-72.
- Coon, J., D. Massam, and L. Travis, eds. 2017. *The Oxford Handbook of Ergativity*. Oxford: Oxford University Press.
- Corbett, G. 2000. *Number*. Cambridge: Cambridge University Press.
- Croft, W. 2002. *Typology and Universals* (2nd ed.). Cambridge: Cambridge University Press.
- Dixon, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.
- Dryer, M. 1992. The Greenbergian word order correlations. *Language* 68: 81-138.
- Dryer, M. 2007. Word order. *Language Typology and Syntactic Description* (2nd ed.), vol. I: *Clause Structure*, ed. T. Shopen, pp. 61-130. Cambridge: Cambridge University Press.
- Dryer, M. and M. Haspelmath, eds. 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [<http://wals.info>]

- Greenberg, J. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of Human Language*, ed. J. Greenberg, pp. 73-113. Cambridge, MA: MIT Press.
- Harley, H. and E. Ritter. 2002. A feature-geometric analysis of person and number. *Language* 78: 482-526.
- Hawkins, J. 1983. *Word Order Universals*. New York: Academic Press.
- Keenan, E. 1985. Relative clauses. *Language Typology and Syntactic Description* (1st ed.), vol. II: *Complex Constructions*, ed. T. Shopen, pp. 141-170. Cambridge: Cambridge University Press.
- Maddieson, I. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.
- Moravcsik, E. 2013. *Introducing Language Typology*. Cambridge: Cambridge University Press.
- Nichols, J. 1986. Head-marking and dependent-marking grammar. *Language* 62: 56-119.
- Palmer, F.R. 1994. *Grammatical Roles and Relations*. Cambridge: Cambridge University Press.
- Payne, T. 1997. *Describing Morphosyntax: A Guide for Field Linguists*. Cambridge: Cambridge University Press.
- Schachter, P. and T. Shopen. 2007. Part-of-speech systems. *Language Typology and Syntactic Description* (2nd ed.), vol. I: *Clause Structure*, ed. T. Shopen, pp. 1-60. Cambridge: Cambridge University Press.
- Shopen, T., ed. 1985. *Language Typology and Syntactic Description*. Cambridge: Cambridge University Press.
- Shopen, T., ed. 2007. *Language Typology and Syntactic Description* (2nd ed.). Cambridge: Cambridge University Press.
- Velupillai, V. 2012. *An Introduction to Linguistic Typology*. Amsterdam: John Benjamins.