Language Universals: Their Discovery and Verbalization by Machine

Vladimir Pericliev

Institute of Mathematics and Informatics, Bulgaria

Abstract

The task of discovering language universals from data is computationally complex and hence difficult to solve manually. To handle this problem I introduce a computational system that can discover all universals in a given database and verbalize them in English. Additionally, if this database has been previously analyzed by a human, the system can compare the discoveries of the human and machine agent and produce a whole article with its results. The system is illustrated on Greenberg's data on word order and the phonological UPSID-451 database. Finally, the system is evaluated

Vladimir Pericliev
Institute of Mathematics and Informatics, bl. 8 1113 Sofia, Bulgaria
Email: peri@math.bas.bg

against six criteria for a successful machine scientific discovery tool.

Keywords: universals, word order universals, phonological universals, machine discovery

1. Introduction

A language universal is usually defined as a property that is valid for all (or most) of the languages of the world (for some refinements of this definition, cf. Pericliev 2012).

To the general public in typology, we owe this idea to Greenberg, and specifically to his influential paper on the order of meaningful elements from 1966 (Greenberg 1966). In actual fact, however, as we learn from Greenberg himself, who has always been scrupulous to acknowledge debt to other linguists, he borrowed the idea of implicational universal and related notions and *discovered for himself* its implications. Greenberg actually borrowed it from the Prague school of structuralism via Roman Jakobson whom Greenberg met at the New York Linguistic Circle. In his paper *The influence of Word and the Linguistic Circle of New York on my intellectual development* (Greenberg 1994), he acknowledges his interest to universal aspects of languages to be due on the one hand to the psychologist Osgood, noting that what would interest a psychologist would be what is common to all languages, and to "the influence of Jakobson, who talked about universal implicational relationships" (p. 23). The story, told by Greenberg himself, is the following:

By 1953-1954 then, I had clearly been influenced by the Prague structuralism that I encountered at Columbia and in the Linguistic Circle. In a sense, it was still a blooming, buzzing confusion like that of the infant as described in a famous passage by Henry James the psychologist. Somehow, typology, linguistic change, marking and universals must be connected. However, it was not until I did my paper on word order, first given at the Behavioral Sciences Center at Stanford in 1959, where also during the same academic year the memorandum on Language Universals was written by Jenkins, Osgood and me in preparation for the Dobbs Ferry Conference on Language Universals held in 1961, that the interconnection between marking, typology and universals began to take form. Put briefly, we can state them as follows. In the relation between the marked and the unmarked, whenever there is a universal implication, the unmarked is the implied. In a typological scheme, the non-existence of a particular type is logically equivalent to a universal, usually an implicational one. For example, the non-existence of languages which were VSO and post-positional could be stated in the following fashion. If a language is VSO, this implies that it is prepositional. Although these conditions are easy to state, they required a number of years to gradually mature in my mind. I recall that at one point, as the key role of implicational universals became clear to me, I had what German psychologists called *Aha-Erlebnis*. So this was what Jakobson was driving at all these years! (Greenberg 1994: 24)

Putting aside the question of scientific priority, Greenberg made important contributions to the study of language universals in

typology (but see arguments against universals in Evans & Levinson 2009). First, his paper initiated a number of works intended to provide a unitary explanation of the empirically observed universals in word order (e.g., Hawkins 1983; Dryer 1991, 1992). Secondly, the quest for universals was extended from word order to other areas of linguistics like phonology (e.g., Ferguson 1966, 1974; Maddieson 1984), semantics (e.g., Wierzbicka 1992), etc. Thirdly, Greenberg's paper set a more rigorous standard for the work in the area, obligatorily comprising the archiving of data, or explicitly stating the database used to make the observations, and an explicit listing (numeration) of the universals, thus allowing both the test of the proposed universals against the data and further research on the same data.

In the empirical approach to universals, advocated by Greenberg, the analyst is faced with a database and he/she needs to discover all the universals valid in the database. This task, however, may be computationally complex, especially when the samples are big. In this paper, I describe a sophisticated computer program which discovers all universals holding in a database and verbalizes them in English. Additionally, if the database has been previously analyzed by a linguist, the system can evaluate the findings of the linguist and produce a whole scientific article with the result of the evaluation.

The paper is organized as follows. Section 2 is a brief sketch of the program, and Section 3 gives some examples from word order and phonological universals. Section 4 assesses the computer system with respect to several criteria that a successful discovery system should satisfy and Section 5 concludes.

2. UNIVAUTO: A Brief Description

Below is a sketch of the system, called UNIVAUTO (UNIVersals AUthoring TOol). A detailed description may be found in Pericliev (2010).

UNIVAUTO accepts as input the following manually prepared information:

1. A *database* (= a table), usually comprising a sizable number of languages, described in terms of some properties (feature-value pairs), as well as a list of the abbreviations used in the database. The program also knows their "English names", or what the abbreviations used for feature values stand for (e.g., AuxV means 'Auxiliary before Verb', SOV means 'Subject-Object-Verb' in that order, etc.). A special value '*' can occur in a database, designating either that the corresponding feature is inapplicable for a language or that the value for that feature is unknown.

2. A *human agent's discoveries*, arising from the same database, stated in terms of the used abbreviations.

3. *Other information*. Aside from these two basic sources of information, the input includes also information on: the origin of database (the full citation of work where the database is given); reference name(s) of database; language families and geographical areas to which the languages in the database belong; etc.

The system supports various queries. Thus, the user may require different: (i) logical types of universals (unconditional or implicational, incl. equivalences, or bi-directional universals, with

two or more variables), (ii) minimum number of supporting languages, (iii) percentage of validity and (iv) statistical significance. The user can also choose the minimum number of (v) language families and (vi) geographical areas the supporting languages should belong to.

UNIVAUTO is a large program, comprising two basic modules: one in charge of the discoveries of the program, called UNIV(ersals), and the other in charge of the verbalization of these discoveries, called AU(thoring)TO(ol).

UNIV can discover various non-redundant logical patterns, or universals (cf. Pericliev 2012), supported in user-specified thresholds of languages, language families and geographical areas, percentage of validity and statistical significance. Importantly, given the discoveries of another, human agent, UNIV employs a diagnostic program to find (eventual) errors in the humanly proposed universals. Currently, the system identifies as problems the following categories:

- *Restriction Problem*: Universals found by human analyst that are below a user-selected threshold of positive evidence and/or percentage of validity and/or statistical significance. E.g., the user may specify that the program find highly significant universals with at least 4 supporting languages, valid in at least 80% of the relevant languages.
- *Uncertainty Problem*: Universals found by human analyst that tacitly assume a value for some linguistic property which is actually unknown or inapplicable, and marked with '*' in the database investigated.
- *Falsity Problem*: Universals found by human analyst that are false or logically implied by simpler universals (= redundant universals).

All of these problems are illustrated in the text generated by our program on Greenbergian word order universals and listed in Section 3.1.

The discoveries of UNIV fall into two types: (1) a list of new universals, and (2) a list of problems (sub-categorized as above).

UNIV assesses the "scientific merit" of its discoveries in order to decide whether to generate a report or not. It uses a natural and simple numeric method: UNIV's discoveries (novel universals plus problems) are judged worthy of generating a report if they are at least as many in number as the number of the published discoveries of the human agent studying the same database.

The authoring module AUTO follows a fixed scenario for its discourse composition, whose basic *components* are:

(1) Statement of title
(2) Introduction of goal
(3) Elaboration of goal
(4) Description of the investigated data and the human discoveries
(5) Explaining the problems in the human discoveries
(6) Statement of the machine discoveries
(7) Conclusion.

The details of this scenario, however, will vary in accordance with a number of parameters, related to the specific query to the system and the corresponding discoveries made. We cannot go into details here, and will only mention that for its surface generation, AUTO employs a hybrid approach, using both templates and rules, which are randomly chosen among a set of alternatives in order to ensure intra-textual variability.

3. Some Examples

3.1. Word Order Universals

UNIVAUTO was applied to the 30-language sample from the classical paper on the order of meaningful elements by Greenberg (Greenberg 1966). The system generated an article, which was published in a peer-reviewed journal with no post-editing, and without disclosing the "machine" origin of the text (Pericliev 1999a).

By way of illustration, below I give some excerpts from Pericliev (1999a) in italics. Comments may follow in angular brackets. The query to the system was to find non-statistical (or exceptionless) universals which are supported by at least four positive examples. The generated text criticizes some of the findings of Greenberg and lists 59 new universals, which he failed to discover in the database.

Further implicational universals in Greenberg's data

The goal of this article is to study the implicational universals in the 30 languages sample of Greenberg 1966 and compare the results of the two studies.
Further implicational universals in Greenberg's data

The goal of this article is to study the implicational universals in the 30 languages sample of Greenberg 1966 and compare the results of the two studies.

<The generated text continues with description of what an implicational universal is, a table of Greenberg's 30 languages sample, accompanied by the abbreviations used, and a listing of the

universals he found. His universals, verbalized by our program, are listed with their numeration in the original publication. An alpha-numeric numeration means that an originally complex universal has been split into elementary ones of the form "If A then B". Cf. text below>

Below we restate, in our wording, the implicational universals of the type studied here which Greenberg 1966 found in these data, preserving for ease of reference their original numeration. If the original formulation expresses a complex implicational statement, it is split into elementary propositions of the type "If P then Q":

[3].   *If a language has dominant order VSO then this language has prepositions.*

[12].   *If a language has dominant order VSO then this language has the question word or phrase placed first in an interrogative word question.*

[13].   *If a language has rigid SOV order then the subordinate verb precedes the main verb.*

[15-a].   *In languages with dominant order VSO the main verb precedes the subordinate verb.*

[15-b].   *In languages with dominant order SVO the main verb precedes the subordinate verb.*

[16-a].   *In languages with dominant order VSO the inflected auxiliary precedes the verb.*

[16-b].   *If a language has rigid SOV order then the verb precedes the inflected auxiliary.*

[16-c].   *If a language has nonrigid SOV order then the verb precedes the inflected auxiliary.*

[21-a].   *If in a language the adjective precedes the adverb*

then the noun precedes the adjective.

[22-a].   *If in a language the standard marker group precedes the adjective then this language is with postpositions.*

[23-a].   *If in a language the proper noun precedes the common noun then the noun precedes the genitive.*

[27-a].   *If a language has suffixes then this language has postpositions.*

[27-b].   *If a language has prefixes then this language is with prepositions.*

We confirmed the validity of universals [12, 13, 15-a, 15-b, 21-a, 22-a, 27-a].

Universal [27-b] is also true, however it violates our restriction pertaining to the occurrence of at least 4 positive examples in the dataset. [27-b] is supported in 1 language (Thai).

Universals [16-a, 16-b, 16-c] are uncertain, rather than indisputably valid in the database investigated, since they assume properties in languages, which are actually marked in the database as "unknown or inapplicable" (notated with "" in Table 4.1). Universal [16-a] would hold only if the feature AuxV/VAux is applicable for Berber, Hebrew, and Maori and in these languages the inflected auxiliary precedes the verb. Universal [16-b] would hold only if the feature AuxV/VAux is applicable for Burmese and Japanese and in these languages the verb precedes the inflected auxiliary. Universal [16-c] would hold only if the feature AuxV/VAux is applicable for Loritja and in this language the verb precedes the inflected auxiliary.*

Universal [23-a] is false. It is falsified in Basque, Burmese, Burushaski, Finnish, Japanese, Norwegian, Nubian, and Turkish, in which the proper noun precedes the common noun but in which the

noun does not precede the genitive.

We found the following previously undiscovered universals in the data.

Universal 1. If in a language the adjective precedes the adverb then the main verb precedes the subordinate verb.

Examples of this universal are 8 languages: Fulani, Guarani, Hebrew, Malay, Swahili, Thai, Yoruba, and Zapotec.

< There follows a listing of new universals found by the system.>

Universal 58. If a language has an initial yes-no question particle then this language has the question word or phrase placed first in an interrogative word question.

<There follows a conclusion which is a summary of the results.>

The universals presented in the generated text are of historical interest only. To-date there are much larger word order databases (e.g., Hawkins 1983, Dryer 1992) that would provide the empirical material for drawing more adequate conclusions about ordering in the world languages. The system's discoveries nevertheless are instructive in showing the deficiency of human, manual search for universals even in a rather small sample, comprising only 30 languages, described in terms of 15 features. Thus, we see that the solution reached by Greenberg is neither complete (i.e., giving all universals) nor sound (i.e., giving only the correct universals).

3.2. Phonological Universals

The UCLA Phonological Segment Inventory Database (UPSID) is

compiled by Maddieson and colleagues at UCLA and is one of the most detailed collections of segment inventories of the world languages. Originally, the database consisted of 371 languages (Maddieson 1984). A later corrected and expanded version (Maddieson & Precoda 1991, Maddieson 1991) contains already a 451 language sample, and I use this later version, usually referred to as UPSID-451, for the computations.

UPSID-451 contains phonologically contrastive segments of the world languages. For each language in the database, a segment inventory is included containing segments that have lexically contrastive function.

UPSID-451 is compiled on a genetic principle, classifying the languages into 18 major genetic groupings (= language families). These language families fall into 4 large geographical areas: Eurasia, Americas, Africa, Australia.

UPSID-451 can be assessed as a very representative sample. The primary goal of the database is to provide a sample from which statistically valid statements concerning frequency and co-occurrence can be drawn. The database is thus a reliable empirical source for a computational investigation.

In our quest for plausible universals, I ran UNIVAUTO on the UPSID-451 data, requiring that the system discover implications of the logical form "If Segment A, then Segment B" such that these implications are:

(i)   valid in at least 90 percent of the relevant languages;
(ii) statistically highly significant;
(iii) valid in two or more language families;
(iv) valid in two or more geographical areas.

Below are listed the first five universals found and verbalized by UNIVAUTO:

Universal 1. The presence of a voiced dental/alveolar plosive in a language implies the presence of a voiceless dental/alveolar plosive.

Positive evidence for this pattern is provided by 82 languages belonging to 14 families from 4 geographical areas. The counter-examples in UPSID-451 are Khalkha, Berta, Eyak, Archi, Avar, Lak, Rutul, Adzera, and Mbabaram. The universal is supported in 82/91 or 90% of the languages in UPSID-451.

Universal 2. The presence of a voiced dental/alveolar plosive in a language implies the presence of a voiced bilabial plosive.

This universal is supported in 89 languages from 16 families situated in 4 areas. The counter-examples in UPSID-451 are Sentani and Eyak. The percentage of validity of the pattern is 89/91 or 98%.

Universal 3. The presence of a prenasalized voiced dental/alveolar plosive in a language implies the presence of a prenasalized voiced velar plosive.

Examples of this universal are 16 languages, members of 9 language families from 4 geographical areas. There is one counter-example, viz. Lai. The percentage of validity of the pattern is 16/17 or 94%.

Universal 4. The presence of a prenasalized voiced dental/alveolar plosive in a language implies the presence of a prenasalized voiced bilabial plosive.

> *Positive evidence for this pattern is provided by 16 languages belonging to 9 families from 4 geographical areas. There is one counter-example, viz. Mazatec. The percentage of validity of the pattern is 16/17 or 94%.*
> *Universal 5. If a language has a voiceless dental/alveolar plosive then it also has a voiceless velar plosive.*
> *This universal is supported in 150 languages from 16 families situated in 4 areas. The exceptions to the pattern are Klao and Zuni. The percentage of validity of the pattern is 150/152 or 99%.*

The system found and verbalized 146 highly significant universals of type "If Segment A, then Segment B" that are valid in at least 90% of the languages in UPSID-451 and are supported by languages from at least two different language families from at least two different geographical areas. Of these 146 universals, 50 or 1/3 are exceptionless.

The vast majority of listed patterns do not appear in Maddieson 1984 (or any other source known to me). These patterns seem to have remained unnoticed by previous researchers, showing, again, the computational complexity of the task of discovering universals. Thus, to find all valid patterns (relative to a database) of the form "If Segment A, then Segment B" implies, first, finding all logically possible combinations from all segments used in the database, and secondly, testing all such hypotheses against the database. More technically, this reduces to finding all ordered pairs (= implications) of segments from the segments used and testing them. The number of ordered pairs of segments from a set of N segments is computed by the formula: $N^2 - N$. Thus, given that the number of segments used in UPSID-451 is N = 919 the number of potential hypotheses to

construct and subsequently test equals $919^2 - 919 = 843,642.$ Additionally, the found putative universals must be statistically significant to minimize mere chance as the reason for their occurrence, a task also difficult to achieve manually. Most previous research on phonological universals in contrast seems to have been appeased with those universals that just *happen* to be noticed by individual linguists, instead of trying an exhaustive (machine) search of the databases studied, and not infrequently posited universals without explicit concern for their statistical significance.

4. An Evaluation of the System

Some common criteria for evaluating discovery systems are the generation of novel, interesting, plausible, and intelligible knowledge, and it has been suggested that a successful system should ideally have all these capacities (Valdes-Perez 1999). The features of portability and insightfulness, which were found to be common to four linguistic discovery systems, were also found to be advantageous to discovery systems (Pericliev 1999b). Below I describe UNIVAUTO along these six dimensions. (Cf. also Colton, Bundy & Walsh 2000 for an interesting similar discussion concerning basically systems in the domain of mathematics.)

4.1. Novelty

UNIVAUTO has so far produced around 60 pages of text, covering about 250 new universals from the fields of word order and phonology. It has found (cf. Pericliev 1999a, 2000) that two of the proposed word order universals in the classical article by Greenberg

(1966) are actually false and that seven others are exceptionless relative to the database investigated rather than statistical, as claimed by Greenberg.

Three other of Greenberg's ordering universals were shown to tacitly assume feature values for some languages which are actually unknown to the database.

All these circumstances have remained unnoticed by previous human researchers, and ironically, some of the problematic universals are widely disseminated in the linguistic community (cf. e.g., the *complete* enumeration of Greenberg's (1966) ordering universals in *The Linguistics Encyclopedia*, London and N.Y., 1991).

Inspecting two further word order databases from Greenberg (1966) and Hawkins (1983), which are really small 24x4 tables, the system also managed to find patterns that have escaped these authors, considered to be the authorities in the field.

Similarly, many novel phonological universals were found in the UPSID database in comparison with Maddieson's (1984) findings, as well as some problems in these and other related proposals in the literature (lack of statistical significance and/or low level of validity and/or insufficiently diverse language support). (Cf. Pericliev 2008)

Three design properties of the system enhance the chances of finding novel knowledge. The first is the system's ability to explicitly check its own discoveries against those of a human agent exploring the same data. More generally, this strategy is not impractical in a linguistic discovery system on universals in view of the availability of universals archives, such as the Konstanz Archive. The second is the exhaustive search of a combinatorial space that the system performs. Such comprehensive searches of combinatorial spaces, that are furthermore dense with solutions, are known to be very difficult, if not completely beyond the reach of a human

investigator, a trite circumstance in computer science (but, unfortunately, not so in many domain sciences as linguistics). As a corollary of the exhaustive search, the system can make meta-scientific claims to the effect that "These are all universals of the studied type (relative to the database)". The third design property is the ability of the system to handle diverse queries (esp. those concerning different logical types of universals), some of which may not have been seriously posed or pursued before.

4.2. Interestingness

The interestingness of UNIVAUTO's findings is partly derived from the interestingness of the task it automates. Indeed, linguistics has always considered the discovery or falsification of a universal an achievement.

From a purely design perspective, the system attempts to enhance the discovery of interesting universals by outputting only the stronger claims and discarding the weaker ones (cf. Pericliev 2012). Thus, if Universal 1 logically implies Universal 2, the first is retained and the second is ignored. E.g., "All languages have stops" implies "If a language has a fricative, it also has a stop" and the second claim must therefore be dismissed as a pseudo-universal. (Ironically, this claim has been actually made more than 60 years ago in a celebrated book by Jakobson (1941), another linguistic luminary, and has never been refuted.)

4.3. Plausibility

The plausibility of posited universals has been a major concern for

UNIVAUTO. Universals are inductive generalizations from an observed sample to all human languages and as such they need substantial corroboration. The system disposes with two principled mechanisms to this end. The first is the mechanism ensuring statistical plausibility, allowing the user to specify a significance threshold for the system's inferences. It is embodied in two diverse methods, the chi-square test and the permutation test, which can alternatively be used. The second plausibility mechanism pertains to the need for qualitatively different languages to provide support for a hypothetical universal for it to be outputted by the program. The specific measure of "typological diversity" of the supporting languages is chosen by the user of the system, by selecting the minimum number of language families and geographical areas to which the supporting languages must belong.

The plausibility of (eventual) criticisms of a human agent's discoveries is even less problematic. Indeed, one can definitely (and not only plausibly) say when a proposition is false relative to a known database, and that is exactly what the system does.

4.4. Intelligibility

With some discovery systems, the user/designer may encounter difficulties in interpreting the program's findings. With other systems, typically those that model previously defined domain-specific problems, and hence systems searching conventional problem spaces, the findings would as a rule be more intelligible. However, intelligibility is a matter of degree and UNIVAUTO seems unique in producing an understandable English text to describe its discoveries.

UNIVAUTO thus both states in English its discoveries (new universals + problems) and the supporting evidence that makes these

discoveries plausible/valid. Additionally, it provides a general context into which it places these discoveries (in the introductory parts of the generated text), as well as a summary of the findings (in the conclusion part of the generated text). The readability and self-contained nature of the texts the system normally produces must not be overstated. Some users may prefer to use the output as a "skeleton article" to be subsequently enlarged and edited to fit further stylistic and linguistic needs.

4.5. Portability

Some discovery systems model general scientific tasks (for induction, classification, explanation, etc.) and would therefore be readily portable to diverse problems in diverse scientific domains. UNIVAUTO is such a system. It mimics the general task of discovery of (logic) patterns from data, and hence would be applicable not only to language universals discovery, where the objects described in the data are languages, but to any database describing any type of objects, be they linguistic or not. This however applies primarily to its discovery module. The text generation module, as it stands, is less flexible and most probably unportable to a domain outside of universals.

4.6. Insightfulness

The degree of formalization discovery programs require may result in our deeper understanding of the tasks modeled, esp. if the sciences from which the task is originally taken are not sufficiently formalized. Another source of insightfulness may be the outcomes of discovery programs, in the case when they make conspicuous some overlooked

aspects of the results. Both the implementation and use of UNIVAUTO have triggered a number of linguistically important insights.

Some of these are worth mentioning here.

First, in the application of the system to linguistic typologies it was consistently found that a set of non-statistical (in contrast to statistical) universals exists that describes all and only the actually attested types, whereas previous influential authors (Hawkins 1983), although strong proponents of exceptionless universals, have claimed them insufficient to do the job. This consistency of the system's results could not be chance of course, so that it was only a short step finding the explanation. Indeed, a linguistic typology is equivalent to a propositional function, and therefore, as known from propositional logic, for any propositional function there exists a propositional expression that generates it. As a corollary, for any linguistic typology there exists a set of non-statistical universals, describing all and only its attested types (Pericliev 2002).

And, secondly, exploring the 451 language database UPSID with UNIVAUTO has led to the formulation of a phonological principle to the effect that if Phoneme 1 implies Phoneme 2, then both phonemes share at least one feature and, besides, Phoneme 2 never has more features than Phoneme 1. This formulation was made possible only *after* the system's discovery of all universals of this type valid in the database. The subsequent (machine-aided) representation of the phonemes in terms of their feature structure highlighted this statistically significant pattern, holding in 94.5 per cent of the cases (Pericliev 2008).

5. Conclusion

The task of discovering language universals from data is

computationally complex and needs automation. I presented a computational system that can effectively perform the task, verbalizing its discoveries. The system has produced linguistically interesting results in the fields of word order and phonology. The system was evaluated against several criteria commonly accepted as necessary for a discovery tool to be successful.

References

Colton, S., A. Bundy, T. Walsh. 2000. On the Notion of Interestingness in Automated Mathematical Discovery. *International Journal of Human-Computer Studies* 53.3, 351-376.

Dryer, M. 1991. SVO Languages and the OV:VO Typology. *Journal of Linguistics* 27, 443-482.

Dryer, M. 1992. The Greenbergian Word Order Correlations. *Language* 68, 81-138.

Evans, N. & S. Levinson. 2009. The Myth of Language Universals: Language Diversity and its Importance for Cognitive Science. *Behavioral and Brain Sciences* 32.5, 429-492.

Ferguson, Ch. 1966. Assumptions about Nasals: A Sample Study in Phonological Universals. In J. Greenberg (ed.), *Universals of Language* 53-60. Cambridge, MA: MIT Press.

Ferguson, Ch. 1974. Universals of Nasality. *Working Papers on Languages Universals* 14, 1-16.

Greenberg, J. 1966. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In J. Greenberg (ed.), *Universals of Language* 73-113. Cambridge, MA: MIT Press.

Greenberg, J. 1994. The Influence of *WORD* and the Linguistic Circle of New York on my Intellectual Development. *Word* 45, 19-25.

Hawkins, J. 1983. *Word Order Universals*. New York: Academic Press.

Jakobson, R. 1941. *Kindersprache, Aphasie, und allgemeine Lautgesetze*. Uppsala: Almqvist & Wilksell.

Maddieson, I. 1984. *Patterns of Sounds*. Cambridge: CUP.

Maddieson, I. 1991. Testing the Universality of Phonological Generalizations with a Phonetically Specified Segment Database: Results and Limitations. *Phonetica* 48, 193-206.

Maddieson, I. & K. Precoda. 1991. Updating UPSID. *UCLA Working Papers in Phonetics* 74, 104-114.

Pericliev, V. 1999a. Further Implicational Universals in Greenberg's Data (a computer generated article). *Contrastive Linguistics* 24, 40-51.

Pericliev, V. 1999b. The Prospects for Machine Discovery in Linguistics. *Foundations of Science* 4.4, 463-482.

Pericliev, V. 2000. More Statistical Implicational Universals in Greenberg's Data (another computer-generated article). *Contrastive Linguistics* 25, 115-125.

Pericliev, V. 2002. Economy in Formulating Typological Generalizations. *Linguistic Typology* 6.1, 49-68.

Pericliev, V. 2008. Implicational Phonological Universals. *Folia Linguistica* 42.1, 195-225.

Pericliev, V. 2010. *Machine-Aided Linguistic Discovery: An Introduction and Some Examples*. London & Oakville: Equinox.

Pericliev, V. 2012. Formulating and Misformulating Language Universals. *Folia Linguistica* 46.1, 211-232.

Valdes-Perez, R. 1999. Principles of Human-Computer Collaboration

for Knowledge Discovery in Science. *Artificial Intelligence* 107, 335-346.

Wierzbicka, A. 1992. *Semantics, Culture and Cognition: Universal Human Concepts in Culture — Specific Configurations*. New York & Oxford: OUP.