On Basic Word Order: The Comparative Framework, Meta-Definition and the Theoretical Motivation

Yanding Guo

Shanghai International Studies University, China
Email: 0223100577@shisu.edu.cn

Abstract—The basic word order is a core term in linguistic typology, and its different theoretical definitions have always been controversial. Firstly, this paper reviews the definitions of basic word order (and dominant order which is extremely related with basic word order) from various schools of thought and compares their definitions through the two levels of "inter-constituent" and "intra-constituent". Secondly, a meta-definition of basic word order that does not deny or exclude other definitions are provided, namely, the basic word order is the order affect other word order(s) or the word order that holds relevance to predicting other word orders. Thirdly, this paper introduces two research approaches in the typological study of basic word order: causality-based research and correlation-based research. The former focus on the phylogenetical relation of word orders while the later focus on the statistical correlation between word orders. Depending on the different research approaches, the orientation of the definition of basic word order will also differ.

Keywords—word order typology, word order, basic word order, dominant word order

I. INTRODUCTION

Basic word order is a core term in linguistic typology, but its definition and theoretical issues are controversial. This paper discusses the concept of basic word order by reviewing the theoretical origins of basic word order and proposing a macroscopic framework to compare the definitions of basic word order from various perspectives.

The organization of this paper is as follows: Section II reviews the theoretical origins of basic word order, including Schmidt's [1] macroscopic view and Greenberg [2]'s basic word order and dominant word order, pointing out that basic word order is the fundamental starting point of the study of word order universals and that dominant word order has multiple meanings. Section III proposes two criteria, "inter-constituent" and "intra-constituent", to place basic word order in a comparable framework and presents a meta-definition of basic word order. Section IV discusses the relationship between the definition of basic word order and the two macroscopic research paradigms of basic word order typology.

II. ORIGINS OF BASIC WORD ORDER

A. Basic Word Order

As Song [3] argued, Linguistic typology typically uses "basic word order" to investigate word order, but Newmeyer [4] criticized there are several definitions of basic word order in typology. To discuss basic word order, we need to trace its origins, first reviewing its use before modern typology and then discussing basic word order in linguistic typology in last decades.

B. Non-typological Basic Word Order

"Basic word order" is a core term in word order typology, so we first need to discuss non-typological basic word order. Non-typological basic word order has two typical applications: (1) fieldwork on a single language and (2) cross-linguistic comparison or contact studies.

In fieldwork, basic word order often refers to the word order of the primary grammatical elements of a transitive clause, namely, the subject (S), verb (V), and object (O). For example, when describing the Dulong language, Sun [5] argued that its basic sentence order was subject-verb or subject-object-verb. This type of basic word order is an intuitive judgment of the "basicness" of sentence constituents and does not have a strict definition. As a result, it is hard to be classified into basic word order in the typological sense.

In cross-linguistic studies, basic word order is used to describe the word order of certain structures in the target language, including but not limited to S, V, and O. Holmes [6] argued that Basque was influenced by early Spanish and had a basic word order of adverb-subject-object-verb. This type of basic word order does not focus on the "basicness" of sentence constituents (adverbs are not "basic" compared to S, V, and O), but rather summarizes the word order of a language in a broad sense. Such basic word order is only an empirical generalization, without a strict definition, and is not a basic word order in the typological sense.

C. The Basic Word Order in Typology

The basic word order in typology, both prior to Greenberg and Greenbergian definition, refers to the researches share a similar theoretical assumption in basic word order. This assumption is typically reflected in the existence of correlations between certain word orders. It is worth emphasizing that this type of study does not necessarily use the specific term "basic word order.", especially when related to "dominant word order".

Greenberg [2]'s basic word order typology was inspired by research from the 19th century, especially from Schmidt [1]. In particular, as Kroeber [7] introduced, Schmidt [1] found a correlation between the position of the genitive noun and possessive pronoun and the position of the accusative and adjective in a sentence. Languages that use prepositions have a subject-genitive word order, while languages that use postpositions have the opposite order.

The core idea of basic word order in typology comes from the inspiration of Schmidt and others on Greenberg, namely that there is a correlation between syntactic elements.

D. Greenberg's Dominant Order

Greenberg [2] suggested that certain factors of languages

are closely related and proposed three basic syntactic parameters: the first is prepositions/postpositions; the second is the order of S, V, O in declarative sentences with a nominal subject and object; and the second is the order of adjectives and nouns.

However, the definition of dominant order is more complex. Greenberg [2] 's dominant order is not a consistent term, which can be seen from the following discussion.

*"The vast majority of languages have several variant orders but a single **dominant** one. Logically, there are six possible orders: SVO, SOV, VSO, VOS, OSV, and OVS. Of these six, however, only three normally occur as **dominant** orders".*

In the first sentence, there is only one dominant order, while in the second sentence, there are three. If dominant order were a consistent term, it would lead to a contradiction.

This contradiction is out of the basic paradigm of typology. Since the basic paradigm of typology is cross-linguistic investigation, which assumes sufficient examination of the language's internal structure and selection of the most representative value for that language's structure among multiple potential values, and then the distribution of these values will reveal some universal of language structure.

However, this paradigm encounters two problems: first, how do we choose a specific order in the languages that usually have multiple optional inflections? Second, how to describe the distribution of orders across languages? Greenberg's dominant order actually concludes solutions for these problems but not clearly argued.

This paper argued the word order obtained by assigning a value to a certain syntactic parameter within a language should be more specifically called the **internal dominant order**. Then, once a sufficient number of languages' parameters have been determined, it is necessary to describe the distribution of these parameters among the languages, such as how many languages in the world use Subject-Verb-Object (SVO) word order. Therefore, the description of the distribution of various levels of a cross-linguistic parameter should be more specifically called the **external dominant order.** In the Greenberg's discussion above, the first dominant order is internal dominant while the second one is external dominant order.

III. BASIC WORD ORDER FRAMEWORK

A. Inter-constituent and Intra-constituent Order

Greenberg's dominant order actually includes information from at least two levels. The internal dominant order is essentially a way of judging the basic word order, while the external dominant order is a descriptive operation for the distribution of word order across languages.

This paper proposes that basic word order is a two-tiered concept that must be understood to distinguish between inter-constituent and intra-constituent word order. These two terms are defined as follows:

1) Inter-constituent order:

The word order selected as the typological commonality benchmark is a set of word orders, each of which has multiple potential values. For example, S, V, O (six potential values); A, N (two potential values); preposition/postposition (two potential values).

2) Intra-constituent order:

For each specific structure in specific language, the process of assigning a definite value from potential values is determined by certain standards. For example, in Chinese, the value of SVO for S, V, O is determined by frequency, and the value of AN for A, N is determined by frequency.

In Greenberg's study, basic word order was simply inter-constituent order, and Greenberg used the concept of (internal) dominant word order to determine the value of Intra-constituent order.

However, in the subsequent studies of many scholars, the distinction between inter-constituent and intra-constituent word order is no longer explicitly made, which means that Greenberg's (internal) dominant word order is essentially the intra-constituent word order part of the generalized basic word order, which will be discussed in detail in the following part on how scholars' definitions can be unified.

B. The Comparative Framework of Basic Word Order

This section will review the definitions of basic word order by multiple scholars. For each scholar, their definition will be introduced first, followed by a summary table, and finally an evaluation of the definition will be provided.

1) Hawkins' Two-Tier Definition

Hawkins [8] discusses both inter- and intra- constituent word order. For example, the basic word order of English verbs is SVO. Hawkins uses the term "doubling" to name the binary inter-constituent word order and "doublet" to name the binary Intra-constituent word order. Hawkins selects a wide range of parameters for inter-constituent word order, including adposition, AN/NA, GN/NG, and VO/OV, while providing a strict definition for intra- constituent word order that includes three criteria: (1) one value has a higher frequency than the other; (2) one value appears more frequently in the grammatical system; and (3) one value is unmarked compared to the other. Thus, we can summarize Hawkins' definition as Table 1:

Table 1. The definition of basic word order by Hawkins

Term	Source	Inter-constituent	Intra-constituent
Basic word order	Hawkins [8]	V, O; A, N; G, N, etc.	text frequency, grammatical system frequency, and markedness

Hawkins believes that the word order inter-constituent and intra-constituent should be discussed separately. Regarding the inter-constituent word order, there is no restriction on using Subject-Verb-Object (SVO) as the basic word order, and a wide range of other basic word orders can be selected. Regarding the word order intra-constituent, the standard is very strict. Hawkins' strict standard is designed to serve his "universal without exception" theory. His research results in complex and universal implications without exception, and the rich selection of parameters is used to increase constraints, while the strict parameter assignment is to avoid controversial issues.

2) Tomlin's investigation-based definition

In a book dedicated to basic word order, Tomlin [9] specifically discusses basic word order, especially providing cross-linguistic distribution of SVO, which provides more data for further research on word order and language

databases. Regarding the word order intra-constituent, Tomlin explicitly states that the primary issue in typological research is to provide a statistical analysis of the six mathematically possible word order frequencies of the core components (S, V, O) in transitive clauses for a representative sample of the world's languages. In addition, Tomlin also points out that if there is no clear evidence, the viewpoint of reference grammar should not be questioned, but if someone raises a question, it also needs to be. Therefore, Tomlin's definition can be summarized as Table 2:

Table 2. The definition of basic word order by Tomlin

Term	Source	Inter-constituent	Intra-constituent
Basic word order	Tomlin [9]	S, V, O	the order in basic transitive clause

From a structural perspective, Tomlin's discussion actually reflects the view of some typologists who consider the SVO word order as the so-called basic word order without further discussion, just as in the second part of the discussion of what "basic" means, some researchers assume that the basic word order in typology is the order between S, V, and O. Due to the limited amount and types of example sentences presented in Tomlin's investigation, basic transitive sentences are relatively reliable evidence. Tomlin's definition also serves the theory, as he proposes the functional principle to summarize the distributional regularities of SVO, such as the Animated First Principle (AFP). This principle states that in basic transitive sentences, the most "animated" NP in the clause will precede other NPs.

3) Siewierska's prototypical transitive definition

Similarly, Siewierska [10] defines the basic word order in typology from the prototype transitive sentence, which appears in a neutral, independent, non-referential clause with a full noun phrase as a participant. The subject is definite and agentive human, while the object is semantically affected and the verb presents an action, not a statement or event. In other words, it is the word order in the structure of prototype transitive sentences defined by Hopper and Thompson [11]. Therefore, it can be summarized as Table 3:

Table 3. The definition of basic word order by Siewierska

Term	Source	Inter-constituent	Intra-constituent
Basic word order	Siewierska [10]	S, V, O	order in prototypical transitive clause

In terms of inter-constituent word order, Siewierska selected SVO, while in terms of intra-constituent word order, she chose prototypical transitive sentences. However, subsequent research has shown that her reasons are not consistent with Tomlin's. Siewierska and Bakker [12] further focuses on the interaction between word order and the configurational patterns of grammatical cases in a language instead of simple distribution, and argues that there are word order strategies, case strategies, and mixed strategies when configuring both syntactic roles (S-V-O) and semantic roles (A-V-P). The most relevant configurational pattern is the sentence used for encoding prototypical transitive events, because these sentences require differentiation between A and P, which is why Siewierska places prototypical transitive sentences in the core position of the basic word order.

4) Comrie's definition of Greenberg-faithness

Comrie [13] introduced the word order parameter as a collective term for the syntactic word order between structures, which includes the word order of the main components of a clause (S, V, O) and the word order of the components related to noun phrases (e.g., A-N/N-A). Comrie believed that in many cases, assigning a basic word order was not a problem, but there were special cases, such as free word order languages and S/O split languages, where it was difficult to assign a value. Some materials that were not suitable as evidence needed to be excluded [13]. Therefore, it can be summarized as Table 4:

Table 4. The definition of basic word order by Comrie

Term	Source	Inter-constituent	Intra-constituent
Basic word order	Comrie [13]	main elements in clause (S, V, O) The components related to noun phrases	Some inappropriate examples should be excluded

Comrie's introduction is consistent with Greenberg's [2] in that word order between structures is not limited to SVO, but rather a parameter used in cross-linguistic studies. First, a word order typology is established, which includes basic sentence order (SVO) and NP-related order, followed by the assignment of specific values. However, Comrie's definition does not positively discuss word order within structures, but instead provides examples of difficult cases. In this regard, Comrie only presents challenges without specific solutions, possibly due to a faithful representation of Greenberg [6], who also did not provide a method for determining preferred word order within structures, and only generally asserted the existence of a single preferred word order.

5) Whaley's intra-constituent focused definition

Whaley [14] proposed a method for determining basic word order. First, he suggested that introspection by native speakers is a helpful approach. Then, he detailed three principles: text frequency, which Whaley acknowledged has potential biases due to the influence of different genres on frequency reliability, information structure, which affects the number and location of arguments, and markedness, which refers to the explicit markings at the phonological and morphological levels. The second principle is markedness, which explicitly addresses explicit markers at the phonological and morphological levels. The third principle is pragmatic neutrality, whereby patterns in pragmatically neutral contexts are more reliable than those in pragmatically specific contexts. Therefore, Table 5 can be summarized:

Table 5. The definition of basic word order by Whaley

Term	Source	Inter-constituent	Intra-constituent
Basic word order	Whaley [14]	S, V, O	Native speaker's introspection; frequency; markedness; neutral text.

Regarding word order between structures, Whaley did not discuss what the basic word order is. However, in all chapters related to basic word order, the only order presented was Subject-Verb-Object (SVO) order. With regards to word order between structures, Whaley was a pioneer in summarizing general methods of distinguishing word order within structures.

6) Dryer's "Basic" abandoned definition

Dryer's [15] definition in World Atlas of Language Structures (WALS) has a strong impact on contemporary typology. In WALS, Dryer does not use the term "basic word order" to reflect the fact that WALS data is entirely based on frequency, as frequency is often the only reliable information. The empirical rule used is that if the count in a corpus shows that the value of one word order parameter is more than twice that of another value, then the word order is considered dominant, otherwise there is no dominant word order [13] (see Table 6).

Table 6. The definition of "basic word order" by Dryer [15]

Term	Source	Inter-constituent	Intra-constituent
Dominant order	Dryer [15]	the features word order in WALS	the type-based frequency

This definition is highly worthy of discussion: is the preferred word order here the basic word order? If so, what is the word order inter-constituent? In fact, all the preferred word orders provided by WALS are word orders inter-constituent that have not yet been specifically studied, and researchers choose the appropriate inter-constituent word order based on their needs. However, due to the fact that some studies cannot accommodate frequency standards, such as Tomlin [9], Siewierska [10], etc., these studies cannot be carried out based on the data provided by WALS because their theoretical assumptions contradict WALS. Therefore, some studies, while disagreeing with the text frequency standard, use data from WALS as a source, which is inappropriate.

C. The Meta-Definition of Basic Word Order

The meta-definition of basic word order is proposed in this paper to better understand, inherit, and develop the achievements of predecessors. Therefore, the meta-definition proposed in this paper is based on the following three principles:

i) not denying the specific operations of predecessors on basic word order, such as frequency standards, secondary materials, and other issues;

ii) contributing to understand and connect different schools' definitions of basic word order;

iii) contributing to connect the future achievements of language typology with existing achievements.

Thus, the meta-definition of basic word order proposed in this paper is as follows:

Basic word order: The order of certain or multiple grammatical elements with assigned specific values that affects the word order of other structures in language universality or has relevance to predicting other word orders.

In detail, Basic word order includes two levels: the selection of participants and the assignment of values. SVO is a typical basic word order in participant selection, but other word orders can also be basic word orders. The standards for assigning values generally include frequency, prototype transitive sentence standard, and neutral pragmatic style, among others.

IV. Two Research Approaches in Basic Word Order Typology

In basic word order typology, this paper argues that there are two research modes: causality-based research and correlation-based research.

A. Causal Patterns and Their Foundations in Typology

Causal patterns refer to a research paradigm that seeks typological evidence under the theoretical assumption of causal relationships. From a perspective of language emergence, human language needs to encode the objective world, and the most fundamental aspect of the objective world involves transitive events with two participants. Therefore, SVO word order, which encodes the transitive clause with the subject, verb, and object, may have been the earliest word order and may have influenced the word order of other structures, as some scholars believe.

Tomlin [9] and Siewierska [10] argue from the perspective of basic transitive clauses and prototypical transitive clauses that the basic word order only has S, V, and O as elements, which are the basic constituents of a transitive clause. This word order is seen as the cause that influences other language phenomena. For example, if a language uses the same sequence as the temporal sequence to encode transitive clauses, that is, encodes AVP events in the sequence of SVO word order, then there is no need for case marking to distinguish the agent and patient. Conversely, if a language does not follow SVO word order, it is likely to confuse the agent and patient and requires case marking to distinguish them. This is also an expression of the language strategies of word order, case marking, and mixing strategies proposed by Siewierska and Bakker [12]. Thus, it can be inferred that languages with case marking do not need to rely on word order to distinguish argument structures, while languages without case marking must rely on word order for argument structure recognition. Therefore, the word order of S and O in case-marking languages is relatively free, while the flexibility of word order in non-case-marking languages is relatively low, and mixed word orders are intermediate between the two. Whether it is the qualitative evidence presented by Siewierska and Bakker [12] or the quantitative research based on corpus data by Levshina [16], the regularity that more marking and more flexible word order is a reasonable causal inference is evident.

Therefore, the commonalities of word order under causal patterns often reflect two principles: (1) in the inter-constituent level, the scope of constituents is relatively small, generally the basic word order of a clause, that is, SVO word order; (2) in the intra-constituent level, the assignment of constituents is strict, and frequency criteria are generally not used to obtain better validation results. As mentioned above, the reason for the former is that SVO is considered the most basic constituent of language from the perspective of language emergence, and other constituents are like "branches" that are greatly influenced by the "trunk". The reason for the latter is that scholars consider from the perspective of language emergence that there is no necessary connection between frequency and importance, and low-frequency language rules may also have a significant impact on other language rules.

B. Correlation Patterns and Their Data-Driven Approach

Correlation pattern is a research paradigm that proposes explanations or causal inferences based on typological evidence without theoretical assumptions, or the only assumption is there are some correlations among word orders. In contrast, the word order in correlation patterns prioritizes the search for related relationships. Thus, the basic word order becomes a reference point for observing other language

phenomena. Namely, one of the most significant correlation parameters.

The correlation patterns have statistical significance supported by existing data, and the corresponding interpretations or causal inferences are made based on the data. The language typology studies based on large-scale language corpora, such as Hawkins [8] and Dryer [15], and the language classification studies based on syntactic annotated corpora and cluster analysis techniques, such as Liu [17], are representative of this type of pattern. Not only in the word order typology, but Wu and Jin [18] also pointed out that language typology research on correlation involves cross-domain regional distribution and the correlation between language and natural, human, social, and economic factors. It is not difficult to see that the study of word order with the goal of pursuing related relationships can only be established on the premise of building large-scale language corpora, large-scale language families, and other non-linguistic databases. Therefore, correlation patterns are essentially data-driven.

Although these patterns have many problems worthy of discussion, this paper only focuses on their impact on the term "basic word order." The basic word order under correlation patterns is just a set of variables for establishing correlation relationships, and its theoretical basis is secondary. Therefore, unlike causal patterns, the status of basic word order in correlation patterns is not high and can even be discarded. Of course, if the study of correlation patterns still adopts basic word order, it often reflects the following principles: (1) in the word order between structures, participant items including but not limited to S, V, and O are often selected. (2) In the word order within structures, frequency standards are generally used. The reason for the former is that correlation patterns first seek correlation without assuming or having to assume that a language participant item is "core." Regarding the latter, it mainly aims to improve the reproducibility of research and the exchangeability of data, which is conducive to replicating existing academic achievements and connecting different language databases. Furthermore, more statistical methods can be applied.

It must be emphasized that causal patterns and correlation patterns are not mutually exclusive. Many times, the theoretical assumptions of causal patterns play a critical role in correlation patterns' search for participant items, explanations, and causal inferences. The statistical methods of correlation patterns can provide data support for the research results of causal patterns and connect with other databases during interdisciplinary and interdisciplinary research. Causal patterns and correlation patterns often alternate and promote each other rather than exist in isolation.

V. CONCLUSION

This paper reviews the origin of basic word order theory, discusses how basic word order affects other language facts theoretically, or serves as a starting point for observing other language facts, and proposes that Greenberg's definition of the preferred word order has a complex meaning. Secondly, this paper reviews and discusses different definitions of various schools and proposes a unified framework and meta-definition for observing basic word order. Finally, based on the theoretical foundations of different research, it explains why there are two major definition divergences in basic word order and provides a perspective for the linguistic typology to observe a unified basic word order.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] P. W. Schmidt, *The Language Families and Language Circles of the World*, Heidelberg: Carl Winter's Universitätsbuchhandlung, 1926. (in German).

[2] J. H. Greenberg, *Universals of Language*, 2nd ed. Cambridge: M. I. T. Press, 1963, ch. 5.

[3] J. J. Song, *Word Order*, Cambridge: Cambridge university press, 2012, p. 14.

[4] F. J. Newmeyer, *Language Form and Language Function*, Cambridge: M. I. T. Press, 1988, ch. 6, p. 331.

[5] H. K. Sun, "The state of Du Long-language," *Minority Languages of China*, vol. 1, no. 4., pp. 292–303, Dec. 1979. (in Chinese)

[6] U. T. Holmes and D. M. Crabb, "A comparative study of word order in old Spanish and old French prose works," *Language*, vol. 32, no. 2, 332, Apr. 1956.

[7] A. L. Kroeber, "Methods and principles: *Die Sprachfamilien und Sprachenkreise der Erde*. P. W. S Chmidt, S. V. D. (Heidelberg, 1926. 596 pp. and atlas.)," *American Anthropologist*, vol. 30, no. 4, pp. 693–696, Oct. 1928.

[8] J. A. Hawkins, *Word Order Universals*, New York: Academic Press, 1983, ch. 1, p. 13

[9] P. J. Hopper and S. A. Thompson, "Transitivity in grammar and discourse," *Language*, vol. 56, no. 2, pp. 251–299, Jun. 1980.

[10] R. S. Tomlin, *Basic Word Order: Functional Principles*, London: Croom Helm, 1986, ch. 2 & 3, pp. 17–34.

[11] A. Siewierska, *Word Order Rules*, London: Croom Helm, 1988, ch. 1, p. 8.

[12] A. Siewierska and D. Bakker, "Case and alternative strategies: Word order and agreement marking," in *The Oxford Handbook of Case*, A Malchukov and A. M. Spencer, Eds. Oxford: Oxford University Press, 2008, ch. 26, pp. 458–469.

[13] B. Comrie, *Language Universals and Linguistic Typology*, Oxford: Blackwell, 1889.

[14] L. Whaley, *Introduction to Typology: The Unity and Diversity of Language*, Thousand Oaks, California, 1997, ch. 6, pp. 96–106.

[15] M. S. Dryer, "Determining dominant word order," in *The World Atlas of Language Structures Online*, M. S. Dryer and M. Haspelmath, Eds. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013.

[16] N. Levshina, "Token-based typology and word order entropy: A study based on Universal Dependencies," *Linguistic Typology*, vol. 23, no. 3, pp. 533–572, Nov. 2019.

[17] H. Liu, "Dependency direction as a means of word-order typology: A method based on dependency treebanks," *Lingua*, vol. 120, no. 6, pp. 1567–1578, Jun. 2010.

[18] J. Wu and L. Jin, "Correlations in linguistic typology," *Foreign Language Teaching and Research*, vol. 49, no. 5, pp. 710–800, 2017.