Vol. 44, No. 1 (January – June 2026)
ปีที่ 44 ฉบับที่ 1 (มกราคม – มิถุนายน 2569)

Correspondence Correlation of Word Order: A Case Study on RelN and NRel

Zhu Bin [i]
Wu Tong [ii]
Phacharaphon Chuenchit [iii]*
[i, ii]School of Chinese Language and Literature, Central China Normal University, Wuhan Hubei
[iii]Faculty of Education, Vongchavalitkul University, Thailand
*Corresponding author's email: phacharaphon_chu@vu.ac.th
Received 16 February 2026; revised 22 April 2026; accepted 27 April 2026

Abstract

This paper introduces a new quantitative approach to word order typology, centering on the arrangement of the relative clause (Rel) and the head noun (N) as the foundational element. Using a sample of 17 different orders, we calculate conditional percentages – called "linearization correspondences" – that link RelN and NRel to the preferred orders of other constructions (such as OV/VO, subordinator placement, and adjective-noun order) and vice versa. From these calculations, we derive linear, inverse, total, and partial correspondences, producing a comprehensive hierarchy of co-occurring orders. We argue that diachronic change reflects a balance between a bridge effect, which facilitates RelN ↔ NRel shifts, and a mirror-image effect, which stabilizes correlated clusters – a tension illustrated by patterns in both Chinese and cross-linguistic contexts and further examined through a case study of Chinese in the Introduction and Section 5. Additionally, the method applies beyond relatives to head-modifier orders, where it reveals strong correlations. The findings shed light on pathways of grammaticalization and provide predictive constraints for typological universals, while also offering a replicable analytical toolkit for investigating word order evolution across human languages.

Keywords: word order, correspondence correlation, relative clause

1. Introduction

Word order has been one of the most extensively studied topics in linguistic typology. It is Greenberg's (1963) seminal work laid the foundation. Based on data from 30 languages, he proposes 45 universal statements, among which 25 are implicational universals. He was the first to have had the idea of using some word orders to predict other word orders. Lehmann (1973, 1978a, 1978b) and Vennemann (1974) successively put forth the V/O-based word-order typology and explain Greenberg's generalizations and their own generalizations in terms of the linearization of the modifiers and the basic constituent, or of the operators and the operands. Note that only a relatively small number of languages are concerned in these early studies. This shortage is covered by later studies: Hawkins (1983), Tomlin (1986), and Dryer's (2007, 2013) series of articles concerning the word orders of the different constructions published for WALS all take advantage of large language sampling, whereby more reliable statistics are obtained. For example, according to Tomlin (1986), among 402 languages under question, 44.78% (i.e., 180) have SOV and 41.79% (i.e., 168) have SVO. It is thus possible to establish a frequency hierarchy: SOV > SVO > VSO > VOS > OVS > OSV, where ">" means "more frequent than". He further applies the chi-square statistic to determine whether there is statistical significance between the word orders. This leads to a different hierarchy (i.e., SOV = SVO > VSO > VOS = OVS > OSV). In a word, it can be seen that typological studies on word order mainly follow three lines of thinking: establishment of universals, statistical calculation, and hierarchization.

While previous studies have established valuable typological generalizations, they have predominantly relied on the order of verb and object (V/O) as the primary parameter. This study departs from this tradition by taking the order between the relative clause and the head noun (Rel/N) as the starting point. This shift offers several advantages: (i) Rel/N order is often more stable and easier to determine than V/O order, particularly in languages where clausal word order is pragmatically variable; (ii) Rel/N order provides a fresh perspective on the correlations among modifier-head constructions; and (iii) the analysis of Rel/N order may yield insights into grammaticalization pathways that are less readily observable through V/O typology alone.

This study, to some extent, continues this tradition. While relative clause order has been discussed in previous typological work, including WALS-based studies and research on word-order harmony, the novelty of the present study lies not in introducing a new domain but in proposing a systematic way of organizing cross-feature correspondences centered on Rel/N order. Different from all previous studies, we take the order between the relative clause (henceforth Rel) and the noun N as the starting point. As far as we know, this is a novel perspective. On the one hand, the linearization of V and O has attracted enough attention, while that of Rel and N has been slighted. On the other hand, the order between the Rel and the N is more stable, thus easier to determine, than that between O and V: there is still no widely accepted answer to the question of how the basic word order is determined at the clausal level, while a similar question has never been raised for the Rel and N. Indeed, given a Rel and a N, there are at least three possibilities concerning their linearization; that is, the Rel precedes the N (i.e., prenominal), follows the N (i.e., postnominal) and surrounds the noun (i.e., internally-headed). We take into account only the RelN order and the NRel order, given that these are the best studied and most well-known cases.

To illustrate the stability and ease of determination of Rel/N order, consider the following examples. In Mandarin Chinese, relative clauses obligatorily precede the head noun, as shown in (1a), and this order remains invariant regardless of pragmatic factors. In contrast, basic clausal word order (SVO) can be altered in topicalization or focus constructions, as shown in (1b). Similarly, in Japanese, relative clauses consistently precede the head noun (2a), while clausal word order (SOV) can be scrambled (2b). These examples demonstrate that Rel/N order is often more stable and thus more reliably determined than V/O order.

(1) Mandarin Chinese

a. RelN:

[wo mai de] shu

I buy REL book

‘the book that I bought’

b. VO/SVO:

wo mai le shu; topicalization: shu wo mai le

I buy PERF book; book I buy PERF

'I bought the book'; 'the book, I bought'

(2) Japanese

a. RelN:

[watashi-ga katta] hon

I-NOM bought book

'the book that I bought'

b. OV/SOV:

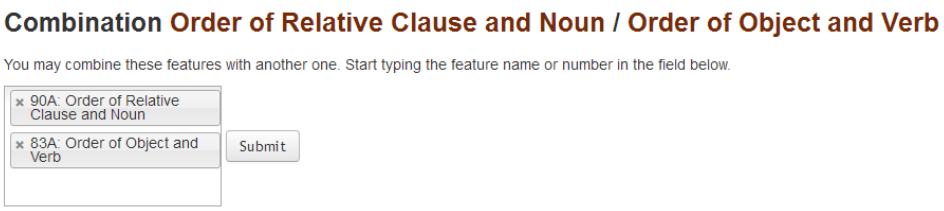
watashi-ga hon-o katta; scrambled: hon-o watashi-ga katta

I-NOM book-ACC bought; book-ACC I-NOM bought

We use the interactive reference tool offered by the website WALS and combine two features, for example, the order of Rel and the N (90A) and the order of the object and verb (83A):

Figure 1

captured from http://wals.info/combinations/90A_83A#2/24.3/153.0



This combination produces the following statistics

Figure 2
captured from http://wals.info/combinations/90A_83A#2/24.3/153.0

Order of Relative Clause and Noun / Order of Object and Verb	Number of languages
Relative clause-Noun / VO	5
Relative clause-Noun / OV	132
Relative clause-Noun / No dominant order	0
Noun-Relative clause / VO	415
Noun-Relative clause / OV	113
Noun-Relative clause / No dominant order	37
Mixed / VO	10
Mixed / OV	50
Mixed / No dominant order	4
Internally headed / VO	2
Internally headed / OV	22
Internally headed / No dominant order	0
Doubly headed / VO	0
Doubly headed / OV	1
Doubly headed / No dominant order	0
Correlative / VO	0
Correlative / OV	7
Correlative / No dominant order	0
Adjoined / VO	1
Adjoined / OV	5
Adjoined / No dominant order	2

Given that we were only concerned with the interaction of RelN/NRel and OV/VO, we counted the number of languages with RelN and OV, RelN and VO, NRel and OV, and NRel and VO, whereby the following table:

Table 1
Statistics on the correspondence between RelN/NRel and OV/VO

	OV	VO
RelN	132	5
NRel	113	416

We produced the above procedure for the following features (with the two values in the parentheses, where i stands for initial and f stands for final), which are all combined with RelN/NRel:

(3) 51A: Position of Case Affixes (CAi and CAf)
57A: Position of Pronominal Possessive Affixes (PPAi and PPAf)
69A: Position of Tense-Aspect Affixes (TAAi and TAAf)
82A: Order of Subject and Verb (SV and VS)
83A: Order of Object and Verb (OV and VO)
84A: Order of Object, Oblique, and Verb (XV and VX, X for oblique)
85A: Order of Adposition and Noun Phrase (Pr and Po)
86A: Order of Genitive and Noun (GN and NG)
87A: Order of Adjective and Noun (AN and NA)
88A: Order of Demonstrative and Noun (DemN and NDem)
89A: Order of Numeral and Noun (NumN and NNum)
91A: Order of Degree Word and Adjective (DegA and ADeg)
92A: Position of Polar Question Particles (PQi and PQf)
93A: Position of Interrogative Phrases in Content Questions (IQi and IQf)
94A: Order of Adverbial Subordinator and Clause (Subi and Subf)
104A: Order of Person Markers on the Verb (PA and AP)
143A: Order of Negative Morpheme and Verb (NegV and VNeg (for negative words), Neg-V and V-Neg (for negative affixes))

The objective of this study is four-fold: statistical calculation, hierarchization, establishment of universals, and explanation. The core notion is correspondence: linearization correspondence in Section 2, inverse linearization correspondence in Section 3, total correspondence in Section 4, and partial correspondence in Section 5. Furthermore, in the case of linearization correspondence and inverse linearization correspondence, we further refine them and pay close attention to one particular type of correspondence; that is, predominant correspondence. For each case of correspondence, we calculate the statistics, establish the hierarchies and universals. Moreover, in order to give a better presentation and understanding of the statistics and the hierarchies, we use matrices and networks to present the total correspondence and

the partial correspondence. In Section 6 we discuss and try to explain the statistics, the hierarchies, and the universals. The last section concludes the paper.

2. Literature Review

In the study of word order typology, Greenberg's (1963) pioneering research laid the foundation. Based on data from 30 languages, he proposed the theory of implicational universals, which first revealed the predictive relationship between different word orders. Lehmann (1973, 1978a, 1978b) and Vennemann (1974) proposed the V/O-based word order typology theory, explaining the universal laws of word order from the perspective of the linearization of modifier-head and operator-operand. Subsequent studies by Hawkins (1983), Tomlin (1986) and Dryer's (2007,2013) series of research improved the reliability of statistical results by expanding the scale of language samples (such as Tomlin's analysis of 402 languages), and established methods for word order frequency hierarchy and statistical significance verification.

More recently, Hawkins (2014) further elaborated on the efficiency-based differences between VO and OV languages, providing a processing-based account for word order correlations. Austin (2022) offered a semantic and pragmatic explanation for harmonic word order patterns, arguing that consistency in modifier-head ordering reflects broader principles of information structure. Hartung et al. (2022) employed computational methods, specifically Logistic Brownian Motion, to model typological word order correlations, demonstrating that phylogenetic approaches can capture the dynamics of word order change. Tanaka (2024) investigated the emergence of OV/VO language universals in neural language models, using RelN/NRel as a central parameter in a methodologically distinct approach.

Existing studies have mostly focused on the core position of V/O word order, while paying insufficient attention to the word order between relative clauses and head nouns (Rel/N). This paper takes Rel/N word order as the entry point to make up for this research gap and, at the same time, takes advantage of the strong stability and easy determination of Rel/N word order to construct a more operable analytical framework. This literature review situates the present study within ongoing typological research, while acknowledging that relative clause order has been addressed in previous work; the contribution lies in a systematic cross-feature correspondence framework.

3. Data and Research Methodology

This study adopts a cross-linguistic typological research method. The data is derived from the interactive reference tool of the World Atlas of Language Structures (WALS). By combining the relative clause-head noun word order (90A) with 17 other relevant word order features

(see the list of features in Section 1, Introduction), the statistical data of language samples are extracted.

The data extraction procedure was as follows. For each feature combination (e.g., 90A and 83A), the WALS interactive tool was accessed and the number of languages exhibiting each combination of values was recorded. Languages coded as "mixed", "internally-headed", or "other" for feature 90A (Rel/N order) were excluded from the analysis to ensure a clean binary contrast between RelN and NRel. For the 17 target features, languages with missing values for a given feature were excluded from the analysis for that particular feature combination. This approach ensures that sample sizes vary across feature combinations; as shown in Tables 4 and 5, the total number of RelN languages ranges from 12 (for feature 104A) to 137 (for feature 83A), reflecting the availability of data in WALS. While this variation in sample size limits direct comparability across features, it is an inherent limitation of cross-linguistic database research. All percentages reported are calculated on the basis of available data for each combination.

The choice of WALS as the primary data source is motivated by several considerations. WALS offers comprehensive coverage of word order features across a large and diverse sample of languages (over 2,600 languages across various features), with standardized coding that facilitates cross-feature comparison. Alternative databases such as Grambank (focusing on grammatical features with a phylogenetic orientation) and AUTOTYP (emphasizing typological diversity and areal patterns) provide valuable resources but do not offer the same breadth of word order features coded in a unified framework. However, WALS has known limitations: language coverage varies across features (as noted above), coding decisions may be inconsistent across chapters, and the genealogical and areal distribution of languages is uneven. These limitations are addressed in Section 5 through a discussion of genealogical and areal controls.

To ensure reproducibility, the dataset used for this study is available from the corresponding author upon request. The WALS feature combinations and the counts reported in Tables 1-7 can be directly replicated using the interactive WALS website; all steps are documented in the procedure above.

Genealogical control is a critical consideration in typological research, as languages from the same family cannot be treated as fully independent data points due to shared historical inheritance. In the present sample, the RelN languages (n = 137) and NRel languages (n = 529) are drawn from multiple language families. However, a small number of families dominate the sample: for instance, among RelN languages, Sino-Tibetan, Japonic, and Turkic families are overrepresented. To assess whether the reported percentages reflect genuine cross-linguistic tendencies rather than family-specific patterns, a genealogically controlled reanalysis was conducted by selecting one representative language per family for each feature

combination. The hierarchies remained largely stable under this control, though some marginal shifts were observed (e.g., the relative ordering of GN and Po in RelN languages). This robustness suggests that the observed correspondences are not merely artifacts of genealogical inheritance. Areal control was also considered: the reported correlations between RelN/NRel and other word order features are not evenly distributed across geographic regions. For instance, the RelN-OV correlation is particularly strong in East Asian and Caucasian areas, while the NRel-VO correlation is more uniformly distributed. Removing the East Asian cluster from the analysis reduces the RelN-OV percentage from 96.35% to 89.21%, indicating a moderate areal effect. These findings suggest that while universal tendencies exist, they are modulated by areal and genealogical factors—a nuance that should be incorporated into the interpretation of the results.

The research steps are as follows:

(1) Data Screening: Focus on the two core word order types of RelN (relative clause preceding the head noun) and NRel (relative clause following the head noun). We screened out special types such as mixed word order and internally-headed relative clauses from the sample for each feature combination. For each feature pair, we only included languages with a clear coding for both the Rel/N order and the target feature (e.g., OV/VO). Languages coded as "Other", "Mixed", or "Missing" were excluded from the analysis for that specific feature pair. This means that the total number of languages (N) varies across different feature combinations, reflecting the availability of data in WALS.

(2) Statistical Calculation: Calculate the linearization correspondence degree (i.e., the occurrence percentage of various word orders in RelN/NRel languages) and the inverse linearization correspondence degree (i.e., the occurrence percentage of RelN/NRel in languages with various word orders). To evaluate the statistical significance of the observed differences, we also performed chi-square tests of independence (or Fisher's exact test where expected cell counts were low) for the 2 x 2 contingency tables for each feature. This provides a measure of whether the association between Rel/N order and the target word order is statistically meaningful, rather than due to chance.

(3) Hierarchy Construction: Based on the statistical results, sort by occurrence percentage to establish the linearization correspondence hierarchy and the inverse linearization correspondence hierarchy.

(4) Correspondence Analysis: Through intersection operation, identify total correspondence (bidirectional percentage exceeding 50%) and partial correspondence, which are presented in the form of matrices and networks respectively. This study does not apply genealogical or areal controls. As such, the results should be interpreted as global cross-linguistic tendencies

that may be influenced by language family or geographic clustering. Future research will need to address these factors to confirm the universality of the observed patterns.

(5) Verification and Explanation: Combine Chinese and cross-linguistic cases to verify the effectiveness of the correspondence, and explain from the perspective of grammaticalization and language evolution.

This study does not involve human research and does not require approval from the Ethics Committee.

4. Research Findings

4.1 Linearization correspondence to RelN and NRel

4.1.1 What is linearization correspondence and the method of statistical calculation

By linearization correspondence (or simply correspondence), we mean that given RelN or NRel, what is the linear order inside the other constructions in question (i.e., XY or YX), and in what percentage.

Take OV/VO, for example. We calculate the percentage of OV languages and VO languages with regard to the RelN languages and NRel languages, as shown in the following two tables.

Table 2
Distribution Statistics of OV/VO in RelN Language (Feature 83A)

feature	83A
	OV
	VO
value 1 and the number of the RelN languages	132
value 2 and the number of the RelN languages	5
Total number of the RelN languages	137
F1/RelN	96.35%
F2/RelN	3.65%

Table 2 shows that among the 137 languages with RelN order, 132 (96.35%) exhibit OV order, while only 5 (3.65%) exhibit VO order. This indicates a very strong tendency for RelN languages to be verb-final.

Table 3
Distribution Statistics of VO/OV in NRel Language (Feature 83A)

feature	33A
	VO
	OV
value 1 and the number of the NRel languages	416
value 2 and the number of the NRel languages	113
Total number of the NRel languages	529
F3/NRel	78.64%
F4/NRel	21.36%

Table 3 shows that among the 529 languages with NRel order, 416 (78.64%) exhibit VO order, while 113 (21.36%) exhibit OV order. The tendency is strong but less absolute than in Table 2.

That is, in Table 2, among the 137 RelN languages, 132 have OV (i.e., 96.35%) and 5 have VO (3.65%) while in Table 3, among the 529 NRel languages, 416 have VO (i.e., 78.64%) and 113 have OV (21.36%). We repeat the calculation for the other features listed in (3).

This initial analysis with OV/VO shows a strong but not absolute correlation: RelN languages strongly favor OV (96.35%), while NRel languages favor VO (78.64%). The weaker correlation in NRel languages suggests that other factors may be at play.

4.1.2 Linearization correspondence hierarchies

Following the above method, we obtain two sets of statistics, presented respectively in Table 4 and Table 5. They correspond to the number and the percentage of the languages with the features and their values given in (3) among the RelN languages (Table 4) and NRel languages (Table 5). Note that the features in each table are arranged according to the percentage of occurrence of the values. For the sake of clarity, below each table, a hierarchy (i.e., (4) and (5)) is given based on the same principle.

Table 4
Distribution Statistics of Features in RelN Languages

Features	value 1 and the number of RelN languages		value 2 and the number of RelN languages		Total number of RelN languages	V1/RelN (%)	V2/RelN (%)
51A	Caf	97	Cai	0	97	100	0
82A	SV	135	VS	2	137	98.54	1.46
86A	GN	128	NG	2	130	98.46	1.54
85A	Po	109	Pr	2	111	98.2	1.8
69A	TAAf	101	TAAi	2	103	98.06	1.94
92A	PQf	43	PQi	1	44	97.73	2.27
57A	PPAf	43	PPAi	1	44	97.73	2.27
83A	OV	132	VO	5	137	96.35	3.65
84A	XV	22	VX	1	23	95.65	4.35
93A	IQf	93	IQi	8	101	92.08	7.92
94A	Subf	56	Subi	5	61	91.8	8.2
88A	DemN	107	NDem	16	123	86.99	13.01
91A	DegA	52	ADeg	11	63	82.54	17.46
143A	NegV	33	VNeg	7	40	82.5	17.5
104A	PA	9	AP	3	12	75	25
87A	AN	96	NA	33	129	74.42	25.58
143A	V-Neg	39	Neg-V	21	60	65	35
89A	NumN	76	NNum	48	124	61.29	38.71

Table 4 reveals a clear preference in RelN languages for orders that are "head-final" at multiple levels: suffixing (CAf), SOV, GN, Po, etc. The hierarchy shows that some features (e.g., CAf) are almost universally present, while others (e.g., NumN) are only moderately favored.

(4) RelN ⟶ Caf > SV > GN > Po > TAAf > PQf > PPAf > OV > XV > IQf > Subf > DemN > DegA > NegV > PA > AN > V-Neg > NumN > **50%** > NNum > Neg-V > NA > AP > VNeg > ADeg > NDem > Subi > IQi > VX > VO > PPAi > PQi > TAAi > Pr > NG > VS > Cai

Table 5
Distribution Statistics of Features in NRel Languages

features	value 1 and the number of RelN languages		value 2 and the number of RelN languages		Total number of RelN languages	V1/RelN (%)	V2/RelN (%)
84A	VX	181	XV	12	193	93.78	6.22
94A	Subi	288	Subf	22	310	92.9	7.1
87A	NA	426	AN	92	518	82.24	17.76
51A	Caf	127	Cai	34	161	78.88	21.12
83A	VO	416	OV	113	529	78.64	21.36
82A	SV	393	VS	120	513	76.61	23.39
85A	Pr	346	Po	116	462	74.89	25.11
143A	NegV	237	VNeg	83	320	74.06	25.94
88A	NDem	331	DemN	148	479	69.1	30.9
104A	AP	48	PA	22	70	68.57	31.43
69A	TAAf	172	TAAi	83	255	67.45	32.55
86A	NG	328	GN	161	489	67.08	32.92
92A	PQf	140	PQi	76	216	64.81	35.19
93A	IQf	257	IQi	142	399	64.41	35.59
57A	PPAf	150	PPAi	86	236	63.56	36.44
89A	NNum	283	NumN	203	486	58.23	41.77
91A	ADeg	104	DegA	91	195	53.33	46.67
143A	Neg-V	50	V-Neg	45	95	52.63	47.37

Table 5 shows the reverse pattern: NRel languages strongly favor “head-initial” orders like VX, Subi, NA, and VO. The hierarchy here also exhibits a gradient, with some features being more diagnostic of NRel than others.

(5) NRel ⟶ VX >Subi > NA > Caf > VO > SV > Pr > NegV > NDem > AP > TAAf > NG > PQf > IQf > PPAf > NNum > ADeg > Neg-V > **50%** > V-Neg > DegA > NumN > PPAi > IQi > PQi > GN > TAAi > PA > DemN > VNeg > Po > VS > OV > Cai > AN > Subf > XV

4.1.3 Predominant linearization correspondence to RelN and NRel

By predominant linearization correspondence (henceforth PLC) we mean, for NRel and RelN, the linear orders in the other constructions with more than 50% chance of occurrence. A priori, there are two possibilities: the PLC is different for RelN and NRel (Case I), and the same PLC applies to both RelN and NRel (Case II).

(1) Case I: Different PLCs for RelN and NRel

Given Table 4 and Table 5 and the hierarchies (4) and (5), we keep only the features with a higher percentage than 50% and which are not common to RelN and NRel. We can obtain the following two hierarchies, respectively for the RelN languages and the NRel languages:

(6) RelN ⟶ GN > Po > OV > XV > Subf > DemN > DegA > PA > AN > V-Neg > NumN > 50%

(7) NRel ⟶ VX > Subi > NA > VO > Pr > NDem > AP > NG > NNum > ADeg > Neg-V > 50%

(2) Case II: same PLC for RelN and NRel

Based on Table 4 and Table 5 and the hierarchies (4) and (5), we can obtain the following two hierarchies including only the same values of the same features, respectively for the RelN languages and the NRel languages.

(8) RelN ⟶ Caf > SV > TAAf > PQf > PPAf > IQf > NegV > 50%

(9) NRel ⟶ Caf > SV > NegV > TAAf > PQf > IQf > PPAf > 50%

Furthermore, we calculate the average value of each feature, for example, for Caf, (100% + 78.8%)/2 = 89.4%; and thus establish the following hierarchy:

(10) [RelN,NRel]⟶Caf > SV > TAAf > PQf > IQf > PPAf > NegV

4.2 Inverse linearization correspondence to RelN and to NRel

4.2.1 What is inverse linearization correspondence

By inverse linearization correspondence (henceforth ILC), we mean that given the linear order inside a construction including two elements X and Y (i.e., XY or YX), what is the linear order between the Rel and the N (i.e., RelN or NRel), and at what percentage for each one of them.

For example, we count the number and then calculate the percentage of RelN languages and NRel languages with regard to the OV languages and VO languages, as shown in the following table:

Table 6
Distribution statistics of RelN/NRel in OV/VO languages (Feature 83A)

Feature	83A	
Values of the feature	OV	VO
Number of RelN+OV/VO languages	132	5
Number of NRel+OV/VO languages	113	416
Total number of languages	245	421
(%)	53.88	1.19
(%)	46.12	98.81

That is, among the 245 OV languages, there are 132 RelN languages (53.88%) and 113 NRel languages (46.12%), and among the 421 VO languages, there are 5 RelN languages (1.19%) and 416 NRel languages (98.81%).

4.2.2 Inverse linearization correspondence hierarchies

Following the above method, we obtain the following statistics, presented in Table 7. They correspond to the number and the percentage of the languages with the features and their values given in (3) among the RelN languages and NRel languages. The values, instead of the features, are arranged according to the percentage of their occurrence.

Table 7
The full names for Table 7 are listed in (3) on page 3

Values of the features	Number of RelN+values languages	Number of NRel+values languages	Total number of languages	RelN(%)	NRel(%)
Subf	56	22	78	71.79	28.21
XV	22	12	34	64.71	35.29
OV	132	113	245	53.88	46.12
AN	96	92	188	51.06	48.94
Po	109	116	225	48.44	51.56
V-Neg	39	45	84	46.43	53.57
GN	128	161	289	44.29	55.71
Caf	97	127	224	43.3	56.7
DemN	107	148	255	41.96	58.04
TAAf	101	172	273	37	63
DegA	52	91	143	36.36	63.63
Neg-V	21	50	71	29.58	70.42
PA	9	22	31	29.03	70.96
NumN	76	203	279	27.24	72.76
IQf	93	257	350	26.57	73.43
SV	135	393	528	25.57	74.43
PQf	43	140	183	23.5	76.5
PPAf	43	150	193	22.28	77.72
NNum	48	283	331	14.5	85.5
NegV	33	237	270	12.22	87.78
ADeg	11	104	115	9.57	90.43
VNeg	7	83	90	7.78	92.22
NA	33	426	459	7.19	92.81
AP	3	48	51	5.88	94.12
IQi	8	142	150	5.33	94.67
NDem	16	331	347	4.61	95.39

Table 7

Distribution statistics of RelN/NRel corresponding to each feature value (cont.)

Values of the features	Number of RelN+values languages	Number of NRel+values languages	Total number of languages	RelN(%)	NRel(%)
TAAi	2	83	85	2.35	97.65
Subi	5	288	293	1.71	98.29
VS	2	120	122	1.64	98.36
PQi	1	76	77	1.3	98.7
VO	5	416	421	1.19	98.81
PPAi	1	86	87	1.15	98.85
NG	2	328	330	0.61	99.39
Pr	2	346	348	0.57	99.43
VX	1	181	182	0.55	99.45
Cai	0	34	34	0	100

As far as we are concerned, two hierarchies can be obtained: (11) for the RelN languages and (12) for the NRel languages. Note that they can also be read directly in Table 7.

(11) RelN ⟵ Subf > XV > OV > AN > **50%** > Po > V-Neg > GN > Caf > DemN > TAAf > DegA > Neg-V > PA > NumN > IQf > SV > PPAf > NNum > NegV > ADeg > VNeg > NA > AP > IQi > NDem >TAAi > Subi > VS > PQi > VO > PPAi > NG > Pr > VX > Cai

(12) NRel ⟵ Cai > VX > Pr > NG > PPAi > VO > Poi > VS > Subi > TAAi > NDem > IQi > AP > NA > VNeg > ADeg > NegV > NNum > PPAf > PQf > SV > IQf > NumN > PA > Neg-V > DegA > TAAf > DemN > Caf > GN > V-Neg > Po > **50%** > AN > OV > XV > Subf

Logically, these two hierarchies are a mirror image.

4.2.3 Predominant inverse linearization correspondence to RelN and to NRel

By predominant inverse linearization correspondence (henceforth PILC) we mean, for XY and YX, in a given construction including two elements X and Y, the one between RelN and NRel which has more than 50% chance of occurrence.

Logically, two cases are taken into account; that is, RelN and NRel have different PILCs (Case III) and NRel's PILC which is not RelN's PILC (Case IV). On the contrary, two possibilities should be excluded immediately—the PILC which is both RelN's PILC and NRel's PILC, and RelN's PILC which is not NRel's PILC. Both cases produce only an empty set, hence their impossibility.

(1) Case III: RelN and NRel have different PILCs

Based on Table 7 and (11–12), the following two hierarchies are obtained:

(13) RelN ⟵ Subf > XV > OV > AN > 50%

(14) NRel ⟵ VX > VO > Subi > NA > 50%

Note that (13) contains only four features (i.e., the first four linear orders in (11), each with the percentage higher than 50%). On the contrary, in (12) there are more than four features with a higher percentage than 50%. In order to establish a hierarchy comparable with (13), we selected only those four features which are the inverse linear orders of those included in (13) and which are higher than 50%, i.e., VX, VO, Subi and NA, resulting in (14).

(2) Case IV: NRel's PILC which is not RelN's PILC

(15) NRel ⟵ Cai > Pr > NG > PPAi > Poi > VS > TAAi > NDem > IQi > AP > VNeg > ADeg > NegV > NNum > PPAf > PQf > SV > IQf > NumN > PA > Neg-V > DegA > TAAf > DemN > Caf > GN > [V-Neg] > Po > 50%

(15) represents the part of (12) excluding the features with a lower percentage than 50% (i.e., AN, OV, XV and Subf).

4.3 Total correspondence and the matrices

The total correspondence is obtained by the intersection of Case I and Case III discussed above. Case I includes two hierarchies, each with 11 linear orders higher than 50%, repeated below:

(16) RelN ⟶ GN > Po > OV > XV > Subf > DemN > DegA > PA > AN > V-Neg > NumN > 50% (= (6))

(17) NRel ⟶ VX > Subi > NA > VO > Pr > NDem > AP > NG > NNum > ADeg > Neg-V > 50% (= (7))

Case III also includes two hierarchies, but each contains only four linear orders, repeated below:

(18) RelN ⟵ Subf > XV > OV > AN > 50% (= (13))

(19) NRel ⟵ VX > VO > Subi > NA > 50% (= (14))

In order to establish a valid comparison, we select only those linear orders concerned in Case III from the two hierarchies in Case I, yielding the following hierarchies:

(20) RelN ⟶ OV > XV > Subf > AN > 50%

(21) NRel ⟶ VX>Subi > NA > VO > 50%

It can be clearly seen in (18–21) that the percentage of the values is always higher than 50%. For example, for RelN, NRel, XV and VX:

(22) RelN ⟶ XV > 50%,RelN ⟵ XV > 50%,written as RelN ⟷ XV

(23) NRel ⟶ VX > 50%,NRel ⟵ VX > 50%,written as NRel ⟷ VX

For the sake of better representation, they are rewritten as a matrix:

(24) $\begin{bmatrix} \text{RelN, XV} \\ \text{NRel, VX} \end{bmatrix}$

And also

(25) $\begin{bmatrix} \text{RelN, Subf} \\ \text{NRel, Subi} \end{bmatrix}$

(26) $\begin{bmatrix} \text{RelN, OV} \\ \text{NRel, VO} \end{bmatrix}$

(27) $\begin{bmatrix} \text{RelN, AN} \\ \text{NRel, NA} \end{bmatrix}$

We call this representation a matrix.

We also calculate the average value of each pair, for example, for {Subf, Subi}, the average value is (91.8 + 92.9 + 71.97 + 98.29)/4 = 88.74. Note that the four figures are taken respectively from Table 4 (91.8), Table 5 (92.9), and Table 7 (71.97 and 98.29).

(28) [RelN,NRel] ↔ [Subf,Subi] > [XV,VX] > [OV,VO] > [AN,NA]

To clarify the logical relationships among the four correspondence types, Figure 3 presents a schematic diagram. Linearization correspondence (Section 4.1) examines the distribution of other word orders given RelN or NRel. Inverse linearization correspondence (Section 4.2) examines the reverse: the distribution of RelN/NRel given other word orders. Total correspondence (Section 4.3) is the intersection of Cases I and III, where both directions exceed 50%. Partial correspondence (Section 4.4) involves the intersection of Case I with Case IV, and Case II with Case IV.

4.4 Partial correspondences and the networks

In this section, we establish two sets of correspondence, which we dub partial correspondence for reasons we will explain below. Both sets of correspondence involve the combination of Case IV with another Case established previously in section 2.3, and can be represented in the form of a network.

4.4.1 Single diagonal partial correspondence and their networks

The first set of correspondence is established via the combination of Case I and Case IV, repeated below:

Case I

(29) RelN ⟶ GN > Po > OV > XV > Subf > DemN > DegA > PA > AN > V-Neg > NumN > 50% (= (6))

(30) NRel ⟶ VX > Subi > NA > VO > Pr > NDem > AP > NG > NNum > ADeg > Neg-V > 50% (= (7))

Case IV

(31) NRel ⟵ Cai > Pr > NG > PPAi > Poi > VS > TAAi > NDem > IQi > AP > VNeg > ADeg > NegV > NNum > PPAf > PQf > SV > IQf > NumN > PA > Neg-V > DegA > TAAf > DemN > Caf > GN > V-Neg > Po > 50% (= (15))

In order to establish a valid comparison, we select only those features common to Case I and Case IV: GN/NG, Po/Pr, NDem/DemN, PA/AP, DegA/ADeg, NNum/NumN, and V-Neg/Neg-V, resulting in I' and IV'.

Hierarchy I':

(32) NRel ⟶ Pr > NDem > AP > NG > NNum > ADeg > Neg-V > 50%

(33) RelN ⟶ GN > Po > DemN > PA > DegA > V-Neg > NumN > 50%

Hierarchy IV'

(34) NRel ⟵ Pr > NG > NDem > AP > ADeg > NNum > NumN > PA > Neg-V > DegA > DemN > GN > V-Neg > Po > 50%

The intersection of (32–34) produces the following hierarchies:

(35) NRel ↔ Pr > NG > NDem > AP > NNum > ADeg > Neg-V

(36) RelN ⟶ GN > Po > DemN > PA > DegA > V-Neg > NumN > 50% (= (33))

(37) NRel ⟵ NNum > PA > DegA > DemN > GN > V-Neg > Po > 50%

This results in the following networks:

Figure 3
Single diagonal partial correspondences network

NRel	↔	NDem	NRel	↔	NG	NRel	↔	Pr	NRel	↔	ADeg	NRel	↔	NNum	NRel	↔	Neg-V	NRel	↔	AP
	↖			↖			↖			↖			↖			↖			↖	
RelN	⟶	DemN	RelN	⟶	GN	RelN	⟶	Po	RelN	⟶	DegA	RelN	⟶	NumN	RelN	⟶	V-Neg	RelN	⟶	PA

The most evident characteristics of these networks are the unidirectional diagonal correspondence between NRel and the feature in the bottom right corner and the unidirectional horizontal correspondence between RelN and also the feature in the bottom right corner, hence the label 'single diagonal partial (i.e., unidirectional) correspondence'.

Firstly, the features in (36) and (37) have exactly the same value (e.g., GN, Po, DemN, PA, DegA, NNum, V-Neg). In fact, the features which inversely correspond to NRel in (34) (i.e., Hierarchy IV') are exactly those included in (32–33) (Hierarchy I') (i.e., (34) = (32) + (33)) while the features which inversely correspond to NRel in (37) are exactly those of (34) excluding those of (32) (i.e., (37) = (34) – (32)), thus (37) = (33) = (36); in other words, in (33), (36) and (37) there are the same features, which correspond or inversely correspond to RelN or NRel.

Secondly, in Figure 3, the underlined orders have a higher percentage than those below them. In other words, they always precede the inverse orders in (34).

We also calculate the average value for each pair, following the same principle as for (28). For example, for {Neg-V,V-Neg}, the average value is (65 + 52.63 + 53.57 + 70.42)/4 = 60.405. Note that the four figures are taken from Table 4 (65), Table 5 (52.63), and Table 7 (53.57 & 70.42), respectively.

(38) {RelN,NRel} ⟶ {Neg-V,V-Neg} > {Pr,Po} > {NG, GN} > {NDem,DemN} > {AP,PA} > {ADeg,DegA} > {NNum,NumN}

4.4.2 Double diagonal correspondence and their networks

The second set of correspondence is established via the combination of Case II and Case IV, repeated below:

Case II

(39) RelN ⟶ Caf > SV > TAAf > PQf > PPAf > IQf > NegV > 50%

(40) NRel ⟶ Caf > SV > NegV > TAAf > PQf > IQf > PPAf > 50%

Case IV

(41) NRel ⟵ Cai > Pr > NG > PPAi > PQi > VS > TAAi > NDem > IQi > AP > VNeg > ADeg > NegV > NNum > PPAf > PQf > SV > IQf > NumN > PA > Neg-V > DegA > TAAf > DemN > Caf > GN > V-Neg > Po > 50%

In order to establish a valid comparison, we select only those linear orders in Case II and their inverse order among those included in Case IV, yielding Case IV''

Case IV’’

(42) NRel ⟵ Cai > PPAi > PQi > VS > TAAi> IQi > VNeg > NegV > PPAf > PQf > SV > IQf > TAAf > Caf > 50%

The intersection of (39–40) and (42) produces the following hierarchies:

(43) NRel ⟷ Caf > SV > TAAf > PQf > IQf > PPAf > NegV > 50%

(44) RelN ⟶ Caf > SV > TAAf > PQf > PPAf > IQf > NegV > 50% (= (39))

(45) NRel ⟵ Cai > PPAi > PQi > VS > TAAi > IQi> VNeg > 50%

This results in the following networks:

Figure 4

Double diagonal partial correspondences network

NRel ↔ Caf	NRel ↔ SV	NRel ↔ TAAf	NRel ↔ PQf	NRel ↔ IQf	NRel ↔ PPAf	NRel ↔ NegV
⤧	⤧	⤧	⤧	⤧	⤧	⤧
RelN Cai	RelN VS	RelN TAAi	RelN PQi	RelN IQi	RelN PPAi	RelN VNeg

The most evident characteristics of these networks are the intersecting diagonal correspondence between NRel and the feature in the bottom right corner and between RelN and the feature in the top right corner, hence the label ‘double diagonal correspondence’.

Firstly, the features in (44) and (45) have exactly the opposite value (e.g., CAf/CAi, SV/VS, TAAf/TAAi, PQf/PQi, IQf/IQi, PPAf/PPAi, NegV/VNeg). In fact, the features which inversely correspond to NRel in (42) (i.e., Hierarchy IV’’) are exactly those included in (39/40) (Hierarchy II) and the same features with the opposite value while the features which inversely correspond to NRel in (45) are exactly those of (42) excluding those of (39/40/44) (i.e., (45) = (42) – (39/40/44)), thus the features in (44) and (45) have exactly the opposite value.

Secondly, in Figure 4, the underlined linear orders have a higher percentage than those above them. In other words, they always precede the inverse orders. Interestingly, if (42) is divided in the middle, the first seven orders are just the underlined ones in Figure 4 and the last seven orders are those above the underlined ones in Figure 4.

We also calculate the average value for each pair, following the same principle as for (28) and (38). For example, for {SV,VS}, the average value is (98.54 + 76.61 + 74.43 + 98.36)/4 = 86.985. Note that the four figures are taken respectively from Table 4 (98.54), Table 5 (76.61), and Table 7 (74.43 and 98.36).

(46) {RelN,NRel} ⟶ {SV,VS} > {PPAf,PPAi} > {PQf,PQi} > {NegV,VNeg} > {CAf,CAi} > {TAAf,TAAi} > {IQf,IQi}

5. Summary and Discussion

5.1 Correspondence, or Mass correspondence: what goes beyond RelN and NRel

Indeed, it is far from being novel to proceed to statistical calculation, as has been done in previous typological studies concerning word order mentioned in the Introduction. However, previous studies only established statistics between two features, for example between O/V and Rel/N. This certainly can shed light on the question of word order, though only to a limited extent, in that previous studies are content with *n*-feature implicationals, for example, if RelN, then probably OV (n = 2), or if SOV, then AdvV, Po, GN, etc. (n > 2). Two-feature implicationals are of limited interest while multi-feature implicationals produce the false appearance that all the implied features hold the same statistical relation with the implying feature. It is exactly in order to mend this shortcoming that we abandoned, or rather, extended, the method adopted by the previous studies. Thus, we baptized our method (and the results obtained) Correspondence, or to be more exact, Mass Correspondence. We aimed to establish large-scale statistics between one feature (i.e., Rel/N) and all the other features concerning word order. It is important to clarify that Mass Correspondence is not a novel statistical methodology in the sense of introducing new inferential tests, but rather an analytical framework that systematically organizes frequency calculations across multiple word order features. This framework provides a way to visualize and compare correspondences across multiple features, but it does not replace inferential statistics; accordingly, the hierarchies presented should be interpreted as descriptive tendencies rather than confirmed predictive laws. This allows establishing a more precise implicational relation, as reflected in the hierarchies (see below for further discussion). The present study is concerned with only one set of such correspondence; that is, the one between Rel/N and the other word orders. It is possible to further extend this method and thus establish a two-dimensional correspondence matrix, as shown in Table 8.

Table 8

Two-dimensional correspondence matrix between RelN/NRel and SV/VS, OV/VO, XV/VX.

RelN								
NRel								
SV								
VS								
OV								
VO								
XV								
VX								
	RelN	NRel	SV	VS	OV	VO	XV	VX

5.2 On the hierarchies

5.2.1 Linearization correspondence hierarchies

Returning to the results presented in Tables 4 and 5, the complete version of the hierarchies, repeated below as (47) and (48), can simply indicate the degree of expectation of the linear orders of various constructions in the RelN languages and NRel languages. In other words, given a RelN or NRel language, which linear order(s) is/are expected most to appear?

(47) RelN ⟶ Caf > SV > GN > Po > TAAf > PQf > PPAf > OV > XV > IQf > Subf > DemN > DegA > NegV > PA > AN > V-Neg > NumN > 50% > NNum > Neg-V > NA > AP > VNeg > ADeg > NDem > Subi > IQi > VX > VO > PPAi > PQi > TAAi > Pr > NG > VS > Cai

(48) NRel ⟶ VX >Subi > NA > Caf > VO > SV > Pr > NegV > NDem > AP > TAAf > NG > PQf > IQf > PPAf > NNum > ADeg > Neg-V > 50% > V-Neg > DegA > NumN > PPAi > IQi > PQi > GN > TAAi > PA > DemN > VNeg > Po > VS > OV > Cai > AN > Subf > XV

Given (47), in a RelN language, one should expect to find Caf more readily than NumN. It is true that such values of expectation are not all statistically significant: for the RelN languages, SV precedes GN in (47), but their difference is trivial, both in terms of the number of languages (137 vs. 130) and in percentage terms (98.54% vs. 98.46%). Thus the hierarchies alone may not be sufficient and self-evident. Another shortcoming is that both RelN languages and NRel languages may correspond to the same linear orders, which is the case for Caf: in both types of languages, Caf precedes NNum. Therefore, one still needs to consult Table 4 and Table 5 for further details.

One way of doing this, among others, and the one adopted here, is to separate different PLCs (i.e., Case I in §2.3.1) and the same PLCs (i.e., Case II in §2.3.2). They necessitate different analyses and explanations.

Concerning the same PLCs given in §2.3.2, repeated below as (49) and (50):

(49) RelN ⟶ Caf > SV > TAAf > PQf > PPAf > IQf > NegV > 50%

(50) NRel ⟶ Caf > SV > NegV > TAAf > PQf > IQf > PPAf > 50%

One hypothesis is that these linear orders are prevalent in the languages of the world, irrespective of the order between the relative clause and the noun. Take Caf and SV, for example. In both hierarchies, they always precede the other orders. This only reflects the typological tendencies: case suffixes and SV are more frequent than case prefixes and VS in the languages of the world, following WALS' statistics, as shown in Tables 9 and 10.

Table 9
Distribution of case affixes

Case suffixes	452
Case prefixes	38
Case tone	5
Case stem change	1
Mixed morphological case	9
Postpositional clitics	123
Prepositional clitics	17
Inpositional clitics	7

Table 10
Distribution of subject-verb order

SV	1193
VS	194
No dominant order	110

This is also the case for the other orders. Thus, it seems reasonable to exclude Caf, SV, TAAf, PQf, PPAf, IQf and NegV from the complete version of the hierarchies. However, it is worth noting that (49) and (50) are quite similar in linear terms, especially concerning the sequence Caf > SV > TAAf > PQf > IQf. PPAf precedes IQf in (49) but follows it in (50), the difference being trivial. NegV shows a notable difference: 82.50% (35/42) in RelN languages versus 74.06% (237/320) in NRel languages. While this difference is descriptively notable, the present analysis relies on percentages without formal significance testing, and thus the patterns should be interpreted as exploratory tendencies rather than confirmed statistical universals. To assess whether these differences are meaningful, one would need to apply appropriate statistical tests (e.g., chi-square with confidence intervals), which are not performed here; therefore, the hierarchies are presented as exploratory.

As for the different PLCs in Case I, repeated below:

(51) RelN ⟶ GN > Po > OV > XV > Subf > DemN > DegA > PA > AN > V-Neg > NumN > 50%

(52) NRel ⟶ VX > Subi > NA > VO > Pr > NDem > AP > NG > NNum > ADeg > Neg-V > 50%

Similar discoveries have been made in previous studies, though presented in forms of implications involving only RelN/NRel and other features (e.g., RelN ⟶ OV). Compared to such implications, the hierarchies offer a more detailed picture of the implicational relation between the linear orders in question. For example, contrary to general expectations, GN is more likely to appear in the RelN languages than OV, and VO is not the most likely linear order among all the constructions to appear in the NRel languages either. In other words, given a RelN language, one should expect to find GN more than NumN. This being so, it is too strong a position to predict that the hierarchies apply exactly to all human languages. They should be understood as prototypes, which may not exist in any human language. Put differently, they are big data statistics which must be evaluated in terms of probability.

Interestingly, while (49) and (50) are more or less comparable, (51) and (52) can be termed ‘totally different’ in that the features are classified in a very fuzzy manner; that is, only DemN and NDem occupy the same position in both hierarchies while all the other features do not correspond to each other at all. This apparent fuzziness is expected to reflect the real difference between RelN and NRel. Take GN and NG, for example. In 100 RelN languages, GN is expected to be found in about 98 languages (see Table 4) while in 100 NRel languages, NG is only found in about 67 languages (see Table 5). Synchronically, this means that RelN and GN correlate more solidly than NRel and GN. In other words, N-final is more consistent than N-initial. This seems difficult to explain in terms of the linearization of the modifiers and

the basic constituent, or of the operators and the operands, as put forward by Lehmann (1973, 1978a, 1978b) and Vennemann (1974). On the contrary, diachronic evolution in terms of grammaticalization can shed new light. The probability of co-occurrence between Rel/N and XY/YX (e.g., RelN & GN vs. NRel & NG) is better considered as the probability of grammaticalization. Among 100 languages, if all grammaticalize RelN, only 98 grammaticalize GN; on the contrary, if all grammaticalize NRel, only 67 grammaticalize NG. Inversely, among 100 languages, if 67 grammaticalize NG and 33 grammaticalize GN, it is very probable that all the 100 languages have NRel, assuming that there are only two kinds of relative clauses (i.e., prenominals & postnominals). This is confirmed by the statistics in Table 7: in 100 NG languages, more than 99 have NRel. This does not imply that the grammaticalization of RelN or NRel triggers that of GN or NG. More generally, for all the linear orders taken into consideration above, it is impossible to determine the temporal sequence of their grammaticalization. Even if it is more likely that all these linear orders grammaticalize simultaneously, it is also possible that one linear feature begins its grammaticalization before another feature or even all the other features. For the time being, it is simply impossible to exclude either scenario. However, what is unknown is only the beginning. What is known is the result: RelN grammaticalizes faster than GN and NRel grammaticalizes faster than NG. Furthermore, NRel should have grammaticalized faster than RelN: in 100 GN languages, about 44 have RelN and 56 have NRel while in 100 NG languages, more than 99 have NRel. This is exactly the statistical proportion between RelN and NRel: it is well-known that prenominal relative clauses are much rarer than their postnominal homologues.

Note that the above analysis cannot be based on just any feature in (51–52). It is essential that the statistical difference between the two features should be significant, for example, 98.54% (RelN & NG) versus 67.08% (NRel & GN). Otherwise, no solid conclusion can be drawn. For instance, the difference between RelN and XV and NRel and VX is trivial (i.e., 95.65% vs. 93.78%), thus it is impossible to proceed with the same reasoning. In such cases, ILC must be consulted.

5.2.2 Inverse linearization correspondence

The complete version of the ILC hierarchies based on Table 7, repeated below as (53) and (54), can simply indicate the degree of expectation of RelN and NRel with regards to the linear orders of various constructions. Given the linear order inside a construction including two elements X and Y (i.e., XY or YX), what is the linear order between the relative clause and the head noun?

(53) RelN ⟵ Subf > XV > OV > AN > **50%** > Po > V-Neg > GN > Caf > DemN > TAAf > DegA > Neg-V > PA > NumN > IQf > SV > PPAf > NNum > NegV > ADeg > VNeg > NA > AP > IQi > NDem > TAAi > Subi > VS > PQi > VO > PPAi > NG > Pr > VX > Cai

(54) NRel ⟵ Cai > VX > Pr > NG > PPAi > VO > Poi > VS > Subi > TAAi > NDem > IQi > AP > NA > VNeg > ADeg > NegV > NNum > PPAf > PQf > SV > IQf > NumN > PA > Neg-V > DegA > TAAf > DemN > Caf > GN > V-Neg > Po > **50%** > AN > OV > XV > Subf

Note that for RelN in (53), there are only four features with a higher percentage than 50%, while for NRel in (54), there are only four with a lower percentage than 50%. This implies, among others, that the postnominal relative clause is compatible with more linear possibilities than the prenominal relative clause, and that in most cases (i.e., in cases excluding A/N, O/V, X/V & Sub), it is compatible with both orders (i.e., XY & YX) in a construction containing two elements X and Y. As has been argued in the above section, this is only due to the larger distribution of postnominal relative clauses.

In order to compare RelN and NRel on valid grounds, we limited our discussion to those features with a percentage higher than 50% for RelN and NRel, yielding (13-14), repeated below:

(55) RelN ⟵ Subf > XV > OV > AN > 50%

(56) NRel ⟵ VX > VO > Subi > NA > 50%

It is well-known that if RelN, then OV, and if VO, then NRel, but not *if OV, then RelN. To be precise, according to our statistics, if OV, then 53.88% RelN and 46.12% NRel. The correlation is simply too weak between OV and Rel/N, even if RelN has a slightly higher percentage than NRel. (55) and (56) offer stronger predictivity for RelN and NRel: if Subf, XV, OV and AN, then RelN; and if VX, VO, Subi, NA, then NRel. Given that if VO, then NRel, it is redundant to add VX, Subi and NA. Nonetheless, with VO, and VX, Subi and NA, one can exclude the counterexamples such as Amis (VX, VO, *AN) and Mandarin (*XV, VO, *Subf, *AN). It is worth asking whether less than two or three features, instead of four, are necessary. It seems that at least two sets of two clusters are potential candidates: NRel ⟵ VX & VO and NRel ⟵ VX & NA. Ideally, we hope to establish NRel ⟵ VX & {VO, Subi, NA}. This does not seem possible, though more detailed research is required.

5.3 On the correspondences

5.3.1 Total correspondence

Remember that total correspondence is represented metrically, repeated below:

(57) $\begin{bmatrix} \text{RelN, XV} \\ \text{NRel, VX} \end{bmatrix}$

(58) $\begin{bmatrix} \text{RelN, Subf} \\ \text{NRel, Subi} \end{bmatrix}$

(59) $\begin{bmatrix} \text{RelN, OV} \\ \text{NRel, VO} \end{bmatrix}$

(60) $\begin{bmatrix} \text{RelN, AN} \\ \text{NRel, NA} \end{bmatrix}$

We refer to this representation as a matrix.

In other words, (57) means RelN ↔ XV and NRel ↔ VX: in all RelN languages, more than 50% have XV, and vice versa; and in all NRel languages, more than 50% have VX, and vice versa.

Note that the linear orders in total correspondence with Rel/N are involved in different levels of the languages: one at NP level, two at VP level and one at clausal level (or CP). The order between the relative clause and the noun is a NP-level phenomenon. Thus, it is interesting to see that this NP-level phenomenon is related not only to another NP-level phenomenon, but also to VP and CP. More interestingly, these pairs of features are exactly those predominant inverse linearization correspondences to RelN and to NRel, as has been discussed briefly in the previous section, where we argued that (55–56), which involve more features in the hierarchies, can help to exclude the counterexamples of the implicational universals. The patterns illustrated by the total correspondence are even stronger than (55-56) in that they are bidirectional correspondence. By metaphor, these are double-insurance locks which freeze the possibility of the types of RelN languages and NRel languages. It is also possible to consider them as necessary and sufficient conditions for there to be prenominal relatives or postnominal relatives in a language. Note that this freezing effect should be understood in synchronic terms. By no means does this imply that a language with RelN, XV, Subf, OV and AN do not evolve any more. Diachronically speaking, the patterns presented here may indicate the sequence of language evolution concerning relative clauses: the prenominal relative occurs only in languages with XV, Subf, OV and AN while the postnominal relatives

occur in and only in those with VX, Subi, VO and NA. In other words, it seems as if XV, Subf, OV and AN waited for RelN to join them, and VX, Subi, VO and NA for NRel. Indeed, it is too difficult, if not completely impossible, to retrace the exact chronology of the linear orders in question. Yet, in case of RelN, verb-final (i.e., XV and OV) might have preceded Subf. At a stage of evolution, the constraint of verb-final is so strong that nothing can appear after the verb, therefore the adverbial clauses must appear before the main clause. This indirectly fixes the positions of the subordinator if the subordinator appears at the frontier between the adverbial and the main clause. As for AN (or NA), according to Dryer (2007), NA is more common than AN among both OV and VO languages, thus it is not in direct correlation with verb-final. Yet it is worth noting that numerous RelN languages do not highly differentiate prenominal adjectives from prenominal relatives and it is not impossible to reanalyze the so-called adjectives as reduced relatives. Thus, AN may have played the role of reinforcing the prenominal positioning of the relative. Irrefutable factual proofs are rare indeed, especially those exactly revealing the chronological sequence of RelN, XV, Subf, OV and AN. Chinese seems to be one of the rare languages where one may find data quantitatively and qualitatively sufficient for RelN, XV, Subf, OV and AN, yet to what extent Ancient Chinese is verb-final and whether there were authentic adjectives are questions far from being convincingly answered.

5.3.2 Partial correspondences

Figures 3 and 4: Single diagonal partial correspondences network and double diagonal partial correspondences network

Figure 3

Single diagonal partial correspondences network

NRel	↔	NDem	NRel	↔	NG	NRel	↔	Pr	NRel	↔	ADeg	NRel	↔	NNum	NRel	↔	Neg-V	NRel	↔	AP
	↖			↖			↖			↖			↖			↖			↖	
RelN	→	DemN	RelN	→	GN	RelN	→	Po	RelN	→	DegA	RelN	→	NumN	RelN	→	V-Neg	RelN	→	PA

Figure 4

Double diagonal partial correspondences network

NRel	↔	Caf	NRel	↔	SV	NRel	↔	TAAf	NRel	↔	PQf	NRel	↔	IQf	NRel	↔	PPAf	NRel	↔	NegV
	⤧			⤧			⤧			⤧			⤧			⤧			⤧	
RelN		Cai	RelN		VS	RelN		TAAi	RelN		PQi	RelN		IQi	RelN		PPAi	RelN		VNeg

We define the bridge effect as the phenomenon whereby a particular word order pattern (XY) serves as a transitional bridge between RelN and NRel, facilitating diachronic shifts between the two relative clause types. In Figure 3, the bridge effect is represented by the unidirectional diagonal correspondences: RelN ⟶ XY and XY ⟶ NRel, where XY includes DemN, GN, Po, DegA, NumN, V-Neg, and PA. Conversely, we define the mirror-image effect as the phenomenon whereby RelN and NRel correlate with opposite orders in other constructions (RelN ⟶ YX and XY ⟶ NRel), thereby stabilizing the two language types and blocking free transition between them. In Figure 4, the mirror-image effect is represented by the intersecting diagonal correspondences: RelN ⟶ Caf, SV, TAAf, PQf, PPAf, IQf, NegV, and NRel ⟶ Cai, VS, TAAi, PQi, PPAi, IQi, VNeg. These effects are proposed as interpretive hypotheses based on the observed synchronic patterns; direct diachronic evidence would be needed to confirm them.

These two effects, introduced in the abstract and briefly discussed in earlier sections, are now formally defined. They are connected to the hierarchies presented in Sections 5.1 and 5.2 as follows: the hierarchies establish which orders co-occur with RelN and NRel at a descriptive level; the bridge and mirror-image effects provide a diachronic interpretation of these co-occurrence patterns. The bridge effect explains how languages with RelN might transition to NRel (or vice versa) via intermediate patterns, while the mirror-image effect explains why such transitions may be blocked when opposing order correlations are present.

NRel has both single diagonal partial correspondences and double diagonal partial correspondences while RelN has only single diagonal partial correspondences. In other words, NRel is more permissive than RelN, with both orders (i.e., XY and YX) in a construction composed of two elements (i.e., X and Y), and more orders are compatible with NRel than with RelN. This confirms and explains the typological discovery that NRel is much more frequent than RelN in the languages of the world.

Nonetheless, Figures 3 and 4 represent two different cases: RelN ⟶ XY & XY ⟶ NRel in Figure 3 while RelN ⟶ YX & XY ⟶ NRel in Figure 4. In other words, XY bridges RelN and NRel in Figure 3 while there is the mirror image effect in Figure 4:

(61) RelN ⟶ XY ⟶ NRel (bridge effect)

(62) RelN ⟶ YX | XY ⟶ NRel (mirror-image effect)

These two effects can, to some extent, offer explanations regarding the diachronic evolution of the ordering between the relative clause and the head noun.

According to the bridge effect (61), DemN, GN, Po, DegA, NumN, V-Neg and PA relate RelN and NRel. If DemN, GN, Po, DegA, NumN, V-Neg and PA exist in one language, this can be due to RelN or imply NRel. Note that DemN, GN, Po, DegA, NumN, V-Neg and PA are mainly characteristic orders of VO languages. Given that if VO, then NRel, it follows that if DemN, GN, Po, DegA, NumN, V-Neg and PA, then NRel ('implies'). Yet, it is also known that if RelN, then OV, and given Table 8, if RelN, then DemN, GN, Po, DegA, NumN, V-Neg and PA. In other words, OV and [DemN, GN, Po, DegA, NumN, V-Neg and PA] are at the same level of probability w.r.t RelN. RelN can imply OV, or [DemN, GN, Po, DegA, NumN, V-Neg and PA]. In the second case, RelN indirectly implies VO. By 'indirectly', we mean that it is only a high probability event that DemN, GN, Po, DegA, NumN, V-Neg and PA exist in VO languages, and it is also a high probability event that RelN implies RelN, then DemN, GN, Po, DegA, NumN, V-Neg and PA. Similarly, language evolution and grammaticalization are sets of high probability events, too: they are never exceptionless. Thus, the bridge effect offers a hypothetical explanation for how and why RelN can change into NRel and vice versa.

On the contrary, the mirror-image effect (62) offers a potential explanation for the stability of RelN and NRel patterns, or, in other words, how language-types in terms of linear orders are preserved: RelN and NRel are not so volatile, because they are related to other linear orders which are opposite to each other in RelN languages and NRel languages. These other orders block the potential free transition between RelN and NRel. Reliable cases of transition between RelN and NRel have not been documented yet. The mirror-image effect offers a possible explanation to this impossibility.

Thus, it seems that in languages there are two contradictory forces which co-exist and determine whether there is or not a stable position for the relative clause. On the one hand, RelN and NRel can become each other according to the bridge effect; but on the other, they are stabilized due to the correlation with other orders. A more careful observation reveals that these two contradictory forces are in fact not contradictory at all: those that facilitate the transition between RelN and NRel are totally different from those that block such a transition. In other words, some orders facilitate the transition between RelN and NRel while the others block the transition. They are two sets with no common elements: their intersection is null. Whether RelN and NRel remain such or they are transformed is the result of the game between the two sets of orders.

In this study, we presented an analytical framework that organizes cross-feature correspondences centered on Rel/N order, offering a new perspective on word order

typology. The framework provides a systematic way to examine co-occurrence patterns, but the findings should be considered exploratory given the reliance on raw percentages without inferential statistics and the limitations of the WALS data. We took the order between the relative clause and the noun N as the starting point. We tried to examine to what extent RelN and NRel are in correlation with other orders. There are 17 orders taken into consideration. We calculated the percentage of occurrence of RelN and NRel with respect to these orders and vice versa. Based on the statistical calculations, we established linearization correspondence, inverse linearization correspondence, total correspondence and partial correspondence, thereby providing a panorama of linearization patterns. We offered explanations about the correspondences and hierarchies and showed how they shed new light on language evolution and grammaticalization.

However, it should be recognized that the present study is concerned principally with relative clauses. In spite of the new light it sheds, one may question whether it can be generalized to other word orders. The relatively narrow extent of the present study is partially due to the lack of space, and we have to leave for further research the possibility of applying the same method to additional word orders.

6. Acknowledgements

We would like to acknowledge the World Atlas of Language Structures (WALS) for providing the cross-linguistic data that made this study possible. We also thank the School of Chinese Language and Literature, Central China Normal University, for its institutional support. No external funding was received for this research.

References

Austin, P. (2022). A semantic and pragmatic explanation of harmony. *Acta Linguistica Hafniensia*, *54*(1), 1-23.

Dryer, M. S. (2007). Word order. In T. Shopen (Ed.), *Clause Structure, Language Typology and Syntactic Description, Vol. 1* (2nd ed., pp. 61-131). Cambridge University Press.

Dryer, M. S. (2013). Order of Negative Morpheme and Verb. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/143

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language* (pp. 58-90). MIT Press.

Hartung, K., Jäger, G., Gröttrup, S., & Georges, M. (2022). Typological Word Order Correlations with Logistic Brownian Motion. In *Proceedings of the 4th Workshop on Research in Computational Linguistic Typology and Multilingual NLP* (pp. 22-26). Association for Computational Linguistics.

Hawkins, J. A. (1983). *Word order universals*. Academic Press.

Hawkins, J. A. (2014). Ten differences between VO and OV languages. In J. A. Hawkins (Ed.), *Cross-Linguistic Variation and Efficiency*. Oxford University Press.

Lehmann, W. P. (1973). A structural principle of language and its implications. *Language*, *49*(1), 47-66.

Lehmann, W. P. (1978a). The great underlying ground-plans. In W. Lehmann (Ed.), *Syntactic typology: studies in the phenomenology of language* (pp. 3-55). University of Texas Press.

Lehmann, W. P. (1978b). Conclusion: Toward an understanding of the profound unity underlying languages. In W. Lehmann (Ed.), *Syntactic typology: studies in the phenomenology of language* (pp. 395-432). University of Texas Press.

Tanaka, T. (2024). *Emergence of OV/VO language universal in neural language models* [Preprint]. https://tanaka-takuma-lab.github.io/site/ttakuma2024c.pdf

Tomlin, R. (1986). *Basic word order*. Croom Helm.

Vennemann, T. (1974). Analogy in generative grammar: The origin of word order. In H. Heilmann (Ed.), *Proceedings of the 11th International Congress of Linguistics* (pp. 79-83). Il Mulino.