# A Top-down Graph-based Tool for Modeling Classical Semantic Maps: A Crosslinguistic Case Study of Supplementary Adverbs

**Zhu Liu**
Tsinghua University
liuzhu22@mails.tsinghua.edu.cn

## Abstract

Semantic map models (SMMs) construct a network-like conceptual space from cross-linguistic instances or forms, based on the connectivity hypothesis. This approach has been widely used to represent similarity and entailment relationships in cross-linguistic concept comparisons. However, most SMMs are manually built by human experts using bottom-up procedures, which are often labor-intensive and time-consuming. In this paper, we propose a novel graph-based algorithm that automatically generates conceptual spaces and SMMs in a top-down manner. The algorithm begins by creating a dense graph, which is subsequently pruned into maximum spanning trees, selected according to metrics we propose. These evaluation metrics include both intrinsic and extrinsic measures, considering factors such as network structure and the trade-off between precision and coverage. A case study on cross-linguistic supplementary adverbs demonstrates the effectiveness and efficiency of our model compared to human annotations and other automated methods. The tool is available at https://github.com/RyanLiut/SemanticMapModel.

## 1 Introduction

A linguistic form—such as a morpheme, word, or construction—can map to multiple related meanings or functions[1], which, in turn, correspond to different cross-linguistic forms. To analyze this multifunctionality, linguists employ semantic map models (SMMs) (Haspelmath, 2003). SMMs construct a network, known as conceptual space, where nodes represent distinct yet related functions, and edges reflect the semantic similarity between these

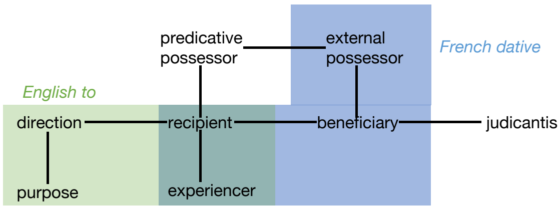


Figure 1: A semantic map of typical dative functions and the regions of English (green) to and French dative (blue).

functions. This structure adheres to the connectivity hypothesis (Croft, 2001), which asserts that functions shared by a single form should map onto a connected[2] region in the conceptual space, thus preserving the limits of linguistic variation (Greenberg, 1963). An example illustrating typical dative functions and two forms is shown in Figure 1.

Constructing the conceptual space typically involves three steps. First, the semantic domain and the corresponding cross-linguistic forms are identified. Second, linguists analyze a form-function table, determining whether a given form can express a specific function based on a multilingual corpus. Third, edges are added and adjusted by either manual annotation or automated procedures, following bottom-up or top-down approaches guided by the connectivity hypothesis. This ensures that each subgraph—representing the multiple functions of a form (i.e., the rows in the form-function table)—remains connected. Typically, this process incrementally builds the graph, edge by edge, as new instances are encountered, reflecting a bottom-up approach. In contrast, our method introduces a top-down algorithm to improve efficiency. These steps are illustrated in Figure 2.

However, this construction process has several limitations. First, manually adjusting the graph is both time-consuming and labor-intensive, espe-

[1] We use the terms *function* or *concept* rather than *sense* (conventional meanings) or *use* (contextual meanings), as it is often difficult to distinguish between conventional, contextual, or even vague meanings, particularly in the case of function words or affixal categories (Haspelmath, 2003).

[2] Any two nodes in this region are connected either directly or indirectly through other nodes.

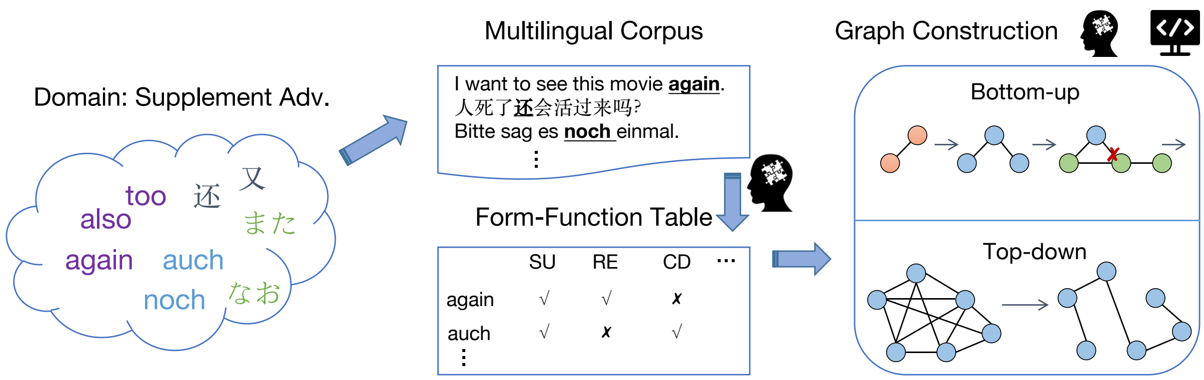


Figure 2: Three steps for constructing semnatic maps. First is to identify the semantic domain and related forms in multiple languages. Second, linguists analyze the form-function table based on the multilingual corpus. Third, a graph is constructed in either a bottom-up or top-down manner. Our method employs a top-down construction.

cially when handling large datasets. Second, although some semi-automated methods have been proposed to assist (Teng, 2015), they still adhere to a bottom-up approach, requiring substantial manual effort to determine which edges to retain or add. Third, human judgment introduces an element of subjectivity, particularly when deciding between two equally plausible connections.

To address these challenges, we propose a novel graph-based algorithm for constructing the conceptual space in a top-down manner. The algorithm begins by creating a dense graph, where the co-occurrence of functions (nodes) in the same form (colexification) serves as the weights for their corresponding edges. We then prune this graph by extracting maximum spanning trees, sorted in descending order by the total weight of their edges. The optimal trees are selected based on automated metrics that include coverage of linguistic instances and network topology. This process is considered top-down because it starts with a dense but suboptimal conceptual space and is subsequently pruned into a tree constrained by our proposed hypothesis. Therefore, it provides a global solution by considering all possible edges in the graph, rather than employing a bottom-up strategy that begins from a localized region formed by individual forms. A case study on supplementary adverbs demonstrates the model's competitive performance and efficiency compared to human-generated maps and other automated methods.

The key contributions of this paper are as follows:

- We propose a top-down graph-based approach for constructing conceptual spaces and semantic map models (SMMs).
- We design a set of metrics to evaluate the quality of the resulting networks.
- A case study on supplementary adverbs demonstrates the efficiency and effectiveness of our proposed method.
- We develop a visualization tool based on this approach to assist typological linguists in studying multifunctionality across languages conveniently.

## 2 Related Work

**Semantic Map Models** Semantic map models (SMMs) are designed to describe linguistic forms that convey different meanings, particularly within the realm of linguistic typology. These forms may include content words (Guo, 2012c; Cysouw, 2007), function words (Zhang, 2017), or constructions (Malchukov et al., 2007). The distinctions between meanings can vary, encompassing word senses (for content words), grammatical functions (for function words), or even specific word forms (Malchukov et al., 2007). SMMs are typically represented as graphs, where nodes correspond to individual functions and edges reflect the proximity between these functions. Classical SMMs (Haspelmath, 2003) connect nodes based on the connectivity hypothesis (see Subsection 3.3). In contrast, second-generation models, such as those based on multi-dimensional scaling (MDS) (Croft and Poole, 2008), introduce weighted edges determined by the frequency of co-occurrence between function nodes. Our method builds upon classical

SMMs with the enhancement of assigning weights to edges.

Generally, linguists construct Semantic Map Models (SMMs) by manually adjusting edge connections to satisfy the necessary constraints. However, this process can be labor-intensive and time-consuming, particularly when dealing with large datasets. Several methods have been proposed to automate the map-building process. For instance, the author in Teng (2015) developed an algorithm that determines when to add an edge for each form in a bottom-up manner. Nevertheless, their approach still requires human input to select the optimal connections for each form, thereby classifying it as a "computer-assisted application." Additionally, their method only addresses the addition of edges and does not account for the removal of edges for new forms. The other two studies Xiao et al. (2021); Chen and Chen (2015) concentrate on the second generation of SMMs, which is not the focus of our work.

**Supplementary Adverbs** Supplementary adverbs are universal across multiple languages, serving to modify the meaning of verbs, adjectives, or other adverbs in a sentence and thereby providing additional contextual information. These adverbs play a crucial role in communication by indicating time, degree, manner, or continuity. Researchers have conducted extensive studies on this subject, emphasizing their multifunctionality. Given this high degree of multifunctionality, the work in (Zhang, 2017) distinguishes different functions in terms of semantic features, such as *Bounded*, *Sequenced*, etc. Linguists focus on words in their respective languages, while still adopting a cross-lingual perspective. For instance, the English adverb *still* is often used to indicate the continuity or persistence of an action over time (Michaelis, 1996). In German, *noch* can denote an additional action or state, while *schon* often implies that something has occurred earlier than expected (König, 1977). In Chinese, the adverb 还 is used to express the continuation or inclusion of an action, whereas 有 can indicate an additional aspect of an action (Shen, 2001; Liu, 2000). Some research also illustrates the evolution of these related words in a diachronic (Tong, 2004) or synchronic (Guo, 2012b) context.

# 3 Preliminary

In this section, we first present some notations, then introduce the conceptual space and the semantic map.

## 3.1 Notations

In this section, we first establish some notations, followed by an introduction to the conceptual space and the semantic map. We summarize these in Table 1.

## 3.2 Nodes and Data Forms

Given a core semantic domain $\mathcal{D}$, we select a candidate set of forms $\mathcal{X}$ and a corresponding set of potential functions $\mathcal{Y}$. Each form $x \in \mathcal{X}$ may be a word (primarily function words), a morpheme, or even a construction. We then examine a corpus to calculate the frequency $f(x, y)$ with which a specific form $x$ corresponds to the function $y \in \mathcal{Y}$. Consequently, we can create a binary[3] function-form matrix $M \in \{0,1\}^{m \times n}$, where the rows represent $m$ forms and the columns represent $n$ functions:

$$M_{x,y} = \begin{cases} 1, & \text{if } f(x,y) \geq 1 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

## 3.3 Conceptual Space

A conceptual space represents the similarity between functions or concepts, adhering to the connectivity hypothesis (Croft, 2001). This hypothesis posits that any form $x$ associated with different functions must correspond to a connected region. We can formalize the conceptual space as an undirected graph $\mathcal{G}$, where the nodes represent the function space $\mathcal{Y}$, and the edges must satisfy the following hypothesis.

**Hypothesis 1** The subgraph $\mathcal{G}(\{y \mid M(x,y) = 1\})$ is connected, where $\mathcal{G}(\{y \mid M(x,y) = 1\})$ denotes the subgraph consisting of the function set associated with a form $x$.

However, Hypothesis 1 is relatively relaxed, allowing many graphs to satisfy the constraints. Therefore, we need to design metrics to evaluate the graph. We consider both intrinsic and extrinsic metrics, which are introduced in detail in Section 4.2.

[3]The binary setting is required in classical SMMs, although the value can also take other forms, such as the frequency of $x$ with respect to $y$ in a corpus, or even a probability for assigning $x$ to $y$. We leave this for further study.

| Term | Notation | Description |
|---|---|---|
| form | $x \in \mathcal{X}$ | A linguistic unit, such as function words or constructions. |
| function | $y \in \mathcal{Y}$ | Meanings or concepts represented as nodes. |
| function-form table | $M$ | A binary indicator of whether a form expresses a specific function. |
| degree of colexification | $w$ | Co-occurrence of functions, represented as weights. |
| conceptual space | $\mathcal{G}(\{y \mid M(x,y)=1\})$ | Represents the similarity between function nodes. |
| semantic map for $x'$ | $\mathcal{G}(\{y \mid M(x',y)=1\})$ | A connected boundary or region corresponding to a specific form $x'$. |

Table 1: Notations and descriptions for key terms.

### 3.4 Semantic Map

Given a conceptual space, a semantic map for a specific form $x'$ is defined as the connected region within the space, indicating the distribution of the functions that $x'$ encompasses. The map (as shown in Figure 1) illustrates the proximity of these functions, implying their diachronic evolution and providing insights into language typology.

There are several variations of Semantic Map Models (SMMs). The classical model is based on Hypothesis 1 concerning connectivity. Another version (Croft and Poole, 2008) incorporates weights for edges based on their frequency of co-occurrence between two functions in a corpus. This approach visualizes the nodes in a plot that preserves their similarity or distance according to each weight, utilizing tools such as MDS. However, this version discards Hypothesis 1, resulting in a graph that is less interpretable than the classical model. In our paper, we adopt the classical model while adding weights to the edges, aiming to leverage the advantages of both methods.

## 4 Approach

In this section, we first introduce our top-down graph-based algorithm for Semantic Map Models (SMMs). Following this, we design both intrinsic and extrinsic metrics.

### 4.1 Top-down Graph-based Algorithm

Considering the trade-off between coverage and precision, we adopt a top-down perspective: (1) to construct a dense graph (sufficient condition); and (2) to prune it into a sparser graph (necessary condition).

First, we construct a dense graph $G_0$, where each edge $e\langle y_i, y_j\rangle$ connecting two function nodes $y_i$ and $y_j$ has an associated weight $w(e)$, referred to as the unnormalized degree of association (Guo, 2012a). This weight is calculated as follows:

$$w(e\langle y_i, y_j\rangle) = M(:, y_i) \cdot M(:, y_j), \qquad (2)$$

where $M(:, y_i)$ represents the $i$th column of the matrix $M$. Thus, $w(e)$ indicates the number of co-occurrences for the two functions[4].

Next, we remove some edges from the complete graph $G_0$. We introduce three constraints into the topology of the conceptual space, as induced by Hypothesis 1:

1. The first constraint is that the final graph $G$ should be connected. This is crucial because if $G$ contains multiple unconnected components, the data can be partitioned into corresponding parts for separate analysis.

2. The second constraint is acyclicity (Zhang, 2015); cycles among the function nodes do not adequately represent entailment relations and fail to predict some wrong forms [5] and should be avoided whenever possible, unless there are no other solutions available to cover the data. Haspelmath (Haspelmath, 2003) refers to the loop as a "vacuous map" because it has no predictive power.

3. The third constraint pertains to the size of $G$; specifically, the summed weights of all edges should be maximized. This is important because these edges represent the most similar functions, which should be retained as much as possible.

These three additional constraints effectively create a maximum spanning tree using graph theory. The revised Hypothesis 2, derived from Hypothesis 1, can be formalized as follows:

**Hypothesis 2** $\mathcal{G}(\{y \mid M(x,y)=1\})$ is a maximum spanning tree, where the weights $w(e)$ on each edge $e$ are calculated using Equation 2.

[4]An alternative version (Guo, 2012a) normalizes $w$ by the occurrence frequency of either individual $y_i$ or $y_j$.

[5]For example, a sequential connection of A, B, and C can predict a form with both A and C as impossible due to the disconnected region, while a cycle cannot.

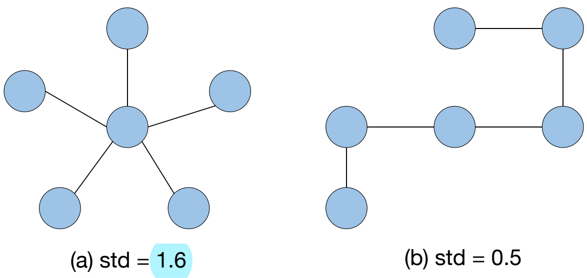


Figure 3: Network topology with different standard deviations of degrees. The left shows a star-like graph with a central node connecting other nodes, while the right shows an averaged connectivity for every node.

In practice, we obtain $n$ spanning trees sorted by their size. Each tree is then evaluated using intrinsic or extrinsic metrics $m$. Algorithm 1 illustrates this process. We utilize the Python package *networkx* to compute the maximum spanning trees [6] sorted by weights, i.e., using the function $MaxSpanningTrees()$. Additionally, we assess the model, as described in the $Evaluate()$ function in the next subsection.

**Algorithm 1** Top-down Graph-based SMMs

**Input:** $M$, number of candidate trees $n$, metrics $m$
**Output:** $G$, the set of conceptual spaces

```
G ← ∅
Construct G_0 from M with weights by Equation 2
P ← MaxSpanningTrees(G_0)
for each tree t in the top n trees in P do
    if Evaluate(t) satisfies m then
        G ← G ∪ {t}
    end if
end for
return G
```

### 4.2 Intrinsic Metrics

For a candidate graph $G$, we evaluate it based on recall, precision, and topology. Recall measures the proportion of forms that satisfy Hypothesis 1 relative to the total number of candidate forms. In contrast, precision shares the same numerator as recall, but its denominator is the number of all possible connected components given the constructed graph. Recall reflects the coverage of the actual samples, while precision ensures that the graph maintains limited predictive power. The trade-off between precision and recall restricts us in finding an optimal graph that is both sufficient (recall) and necessary (precision). For instance, a complete graph serves as a trivial solution that satisfies Hypothesis 1 with a recall of 1, but it possesses very low predictive power, as it tends to predict many noisy forms.

We also propose a new metric for the topology of the network. Specifically, we compute the standard deviation (Div_D) of all the degrees, which is the number of edges connected to a specific node. A lower standard deviation is desirable because it indicates that the edges from the nodes are more "averaged," rather than exhibiting a star-like topology, which offers less interpretability for the entailment relationship of functions. For instance, Croft (Croft, 2003) employs a Markov chain-style representation (Figure 3b) rather than a star-like network (Figure 3a) to depict the animacy hierarchy across multiple languages. Additionally, we use the size of the graph, i.e., the summed weights of all the connected edges, to indicate the important relationships that are retained. Table 2 presents brief descriptions of the different metrics.

| Metric | Description | Trend |
|---|---|---|
| Size | Summed weights of edges | ↑ |
| Recall | Coverage rate of instances | ↑ |
| Precision | Accuracy of predicted instances | ↑ |
| Div_D | Standard deviation of degrees | ↓ |
| Acc | Matched rate compared to GT | ↑ |

Table 2: Different metrics for evaluating the conceptual space. The trend shows the optimal direction for a better network.

### 4.3 Extrinsic Metrics

We also evaluate the network extrinsically, assuming we have the ground-truth semantic space constructed by linguistic experts. We calculate the accuracy (acc) as the ratio of matched edges to the total number of edges. Specifically, we represent a graph with an adjacency matrix $T$, using $T_p$ for the candidate graph and $T_g$ for the ground truth (GT). The accuracy is calculated as follows:

$$\text{acc} = \frac{\sum_{i,j} \mathbf{1}_{T_p(i,j)=T_g(i,j)}}{n^2}, \tag{3}$$

where $n$ is the number of functions.

We emphasize that several baselines are necessary to determine the lower bound of accuracy. The

[6] https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.tree.mst.SpanningTreeIterator.html

| L | G | AF | SU | RE | CO | GD | DE | IS | CD | DC | PT | SC | WH | SE | SC | IC | UE | BL | DS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ZH | 还 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| | 又 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| | 也 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| | 在 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| BO | ra | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | tarong | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| EN | also | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | too | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | again | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | still | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| DE | auch | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | noch | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| FR | aussi | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | encore | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| RU | tbzhe | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | opyat’ | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| JA | も | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | また | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | なお | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KO | 도 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 더 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 또 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| | 다시 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 아직 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| VI | cũng | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| | nữa | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | còn | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | lại | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

Table 3: Form-function table for the Supplement-related semantic domain. Here, “L” represents languages and “G” denotes grams. Abbreviations for languages and functions are detailed in Tables 4 and 5. A value of “1” indicates that the gram corresponds to the function in at least one sentence.

first baseline is a tree (LT) in which the edges do not overlap with the GT. The lower bound can be calculated as:

$$LB_{LT} = \frac{n^2 - 4 \times (n-1)}{n^2}. \tag{4}$$

The second baseline is a complete graph, which is the opposite of the first case: only the edges in the ground truth (GT) are correctly chosen. The lower bound for this scenario can be expressed as:

$$LB_C = \frac{4 \times (n-1) + n}{n^2}. \tag{5}$$

## 5 Case Study

In this section, we present a case study that utilizes Semantic Map Models (SMMs) to analyze adverbs related to "Supplement" (Guo, 2010). The author collected 28 forms (function words) across nine languages that exhibit the "Supplement" function along with 17 other related functions [7]. Based on whether a form in a corpus corresponds to a given function, the author constructed the form-function table $M$, as shown in Table 3. The full names of the languages and functions are provided in Tables 4 and 5.

| Abbr | Full |
|---|---|
| ZH | Chinese |
| BO | Tibetan |
| EN | English |
| DE | German |
| FR | French |
| RU | Russian |
| JA | Japanese |
| KO | Korean |
| VI | Vietnamese |

Table 4: ISO 639 abbreviation codes and full names for languages used in the case study

**Qualitative Analysis** We generated the conceptual space using the algorithm described earlier, with the graph exhibiting the largest weight displayed in Figure 4 (purple solid lines), alongside the ground truth (GT) graph (black dashed lines). Compared to the GT network, our method reveals several distinctive features: (1) Each edge is assigned a weight that reflects the degree of connec-

[7] We excluded the "Conjunction" function, as no words were associated with it.

| Abbr | Full |
|---|---|
| AF | Additive Focus |
| SU | Supplement |
| RE | Repetition |
| CO | Continuation |
| GD | Greater Degree |
| DE | Decrement |
| CD | Condition |
| DC | Discretional Condition |
| PT | Polarity Trigger |
| SC | Serious Condition |
| WH | Whatever |
| SE | Sequence |
| SD | Sequential Coordinator |
| IC | Inconsistency |
| UE | Unexpectedness |
| BL | Bottom Line |
| DS | Discourse Continuation |

Table 5: Abbreviation and full names for functions

tivity between two functions; (2) Our network is acyclic, as stated in Hypothesis 2, while the GT includes cycles to maintain connectivity. We argue that introducing cycles should be avoided when possible due to the lack of interpretability (Haspelmath, 2003), deferring this decision to human experts; (3) The basic topology of our network is similar to the GT; for instance, functions like "SU," "CD," and "RE" serve as critical points with multiple connecting edges.

**Quantitative Analysis** Before evaluating the graphs sorted by size in descending order, we first identify the starting index for different graph sizes, as shown in Table 6. The data indicate that the number of graphs at each size increases exponentially as the size grows, suggesting a wide variety of candidates and highlighting the importance of the evaluation metrics. Subsequently, we evaluate the top five graphs sorted by size, using a step size of 10K due to varying edge sizes, with results summarized in Table 7. We also provide two baselines: one for the complete graph (denoted as C) and another for a tree with no overlap with the GT (denoted as LT). For comparison, the evaluation of the ground truth (denoted as GT) serves as the upper bound.

The results demonstrate that our method uncovers numerous comparable conceptual spaces, with recall exceeding 85% and accuracy surpassing 90%. It is important to note that precision is relatively low for each graph due to the high number of connected subgraphs. However, we identified instances where precision exceeds that of the ground truth (GT) while maintaining a similar level of recall (e.g., index 4). This finding suggests that linguists can utilize our algorithm to generate a set of comparable candidates and subsequently adjust the graph according to their expertise. We observed four forms in the largest network of Figure 4 that violate connectivity hypothesis 1. These forms are *taro* (RE-CO-CD-DS), *still* (CO-GD-DE-CD-BL), *cũng* (AF-CD-DC-PT-UE-BL), and *còn* (RE-CO-GD-CD-PT). Notably, all share the common function CD, indicating a cycle involving this function, as shown by the GT.

| Size | 90 | 89 | 88 |
|---|---|---|---|
| Begin of Index (boi) | 0 | 1,440 | 21,744 |

Table 6: Beginning of index (boi) corresponding to different sizes for a tree.

| Index | Size↑ | Recall↑ | Precision↑ | Accuracy↑ |
|---|---|---|---|---|
| C | 286 | 1 | 0 | 50.0 |
| LT | - | - | - | 79.0 |
| GT | 91 | 1 | 0.20 | 1 |
| 0 | 90 | 85.7 | 0.17 | 92.6 |
| 1 | 89 | 82.1 | 0.21 | 91.4 |
| 2 | 89 | 82.1 | 0.44 | 90.1 |
| 3 | 88 | 82.1 | 0.34 | 91.4 |
| 4 | 88 | 78.6 | 0.50 | 88.9 |

Table 7: Evaluation of our generated graphs and baselines (denoted as complete graph C and ground truth GT). The index represents the first N maximum spanning trees, scaled by 10,000.

| Round | RG_1 | RG_2 |
|---|---|---|
| 1 | -17.8 | -22.1 |
| 2 | -21.9 | -22.4 |
| 3 | -20.5 | -19.2 |
| 4 | -23.8 | -21.7 |
| 5 | -23.1 | -24.1 |
| Mean | -21.4 | -21.9 |
| Std. Dev. | 2.13 | 1.58 |

Table 8: Pearson correlation between Div_D (diversity of degrees) and accuracy across five rounds. The mean and standard deviation for each round are also provided.

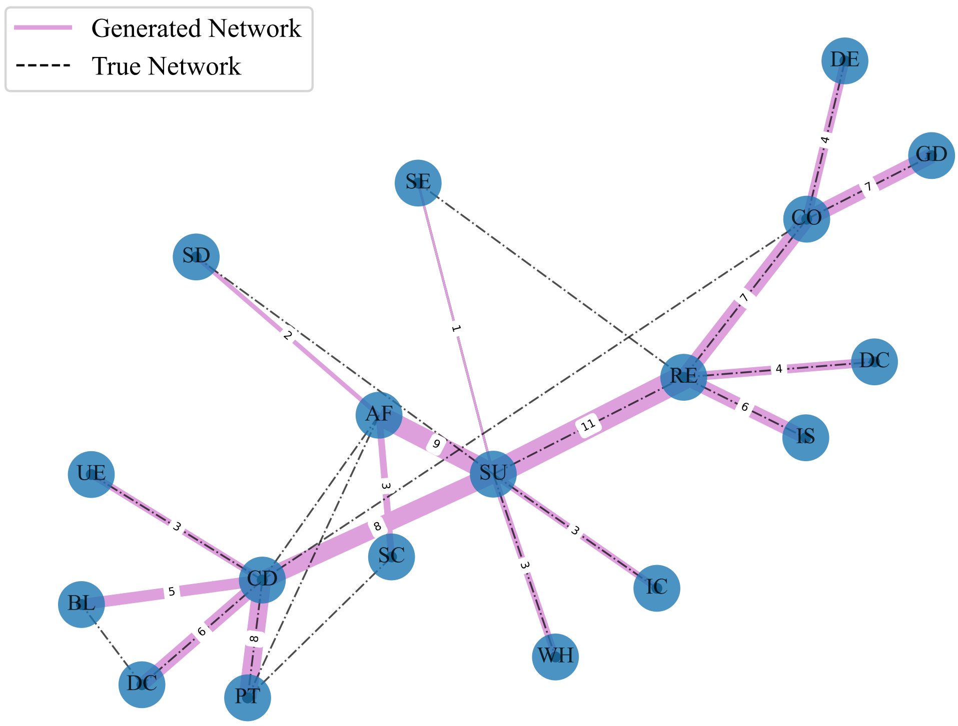


Figure 4: Tree of conceptual space with the largest size. The pink connections represent the network generated by our method, while the black dashed line indicates the ground truth as labeled by an expert. Numbers on the edge indicate the number of co-occurrences in a same word for the corresponding functions.

**Metrics Analysis** We evaluate the effectiveness of our proposed metric, Div_D. Given that tree structures typically exhibit less diversity in node degrees due to their sparse edges, we do not restrict our analysis to tree-based graphs. Instead, we generate graphs randomly in two ways: the first method produces graphs with entirely random edges, referred to as "RG_1," while the second method ("RG_2") assigns edges with a probability proportional to the original weights, favoring higher-weighted edges. We sample 1,000 graphs per round over five rounds. In each round, we compute the Pearson correlation between Div_D and accuracy, as shown in Table 8. Both cases exhibit a moderate negative correlation, indicating that star-like topologies, characterized by lower degree diversity, are less desirable for an ideal conceptual space.

## 6 Conclusion

In this paper, we propose a graph-based algorithm to generate a conceptual space and semantic map in a top-down manner. This approach is realized by instantiating the conventional connectivity hypothesis into a greedy maximum spanning tree, where the weights are determined by the number of co-occurrences. To select the optimal candidate, we introduce a novel metric—diversity of degrees—based on graph topology, in addition to classical metrics such as precision, recall, and accuracy against ground truth. Through a case study on supplementary adverbs, we demonstrate the effectiveness and efficiency of our approach. Our visualization tool provides linguists with a valuable reference that can be further refined manually, contributing to the fields of language typology, computational semantics, and related subjects.

## 7 Limitations

Our current version is preliminary and has several limitations. First, we do not account for the frequency with which a form is associated with a specific function in a corpus, an important factor in second-generation SMMs, such as MDS-based models. In future work, this frequency could be integrated into the edge weights. Second, we have not addressed the subjectivity and uncertainty involved in assigning functions to forms. The grams studied in SMMs, particularly function words like conjunctions, often exhibit highly uncertain and overlapping semantic spaces. We plan to mitigate this issue by employing soft labels instead of deterministic ones when populating the data matrix. Lastly, we have not incorporated temporal information into the evaluation of functions. While this aspect is more relevant to linguistic analysis, we aim to explore directional relationships in an automated manner in future research. Moving forward, we will validate our model with additional case studies across different languages and time periods.

## 8 Ethics Statement

We do not foresee any immediate negative ethical consequences arising from our research.

## References


Zhenyu Chen and Zhining Chen. 2015. Revealing hidden patterns in grammar through map analysis: The weighted minimum edge map. *Chinese Language*, (5):428–437.

William Croft. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press, Oxford.

William Croft. 2003. *Typology and Universals*, second edition edition. Cambridge University Press, Cambridge.

William Croft and Keith T. Poole. 2008. Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. *Theoretical Linguistics*, 34(1):1–37.

Michael Cysouw. 2007. Building semantic maps: the case of person marking. In Matti Miestamo and Bernhard Walchli, editors, *New Challenges in Typology: Broadening the Horizons and Redefining the Foundations*, pages 225–248. Mouton, Berlin.

Joseph H. Greenberg, editor. 1963. *Universals of Language*. M.I.T. Press, Cambridge, MA.

Rui Guo. 2010. A semantic map study of adverbs related to "supplement". In *Paper presented at the International Symposium for Comparative and Typological Research on Languages of China*, Hong Kong, China.

Rui Guo. 2012a. The minimum connection principles and connection degrees of concepts in semantic maps. In *Proceedings of the Conference on Multilingual and Multifunctional Grammatical Forms*, Peking University.

Rui Guo. 2012b. Synchronic semantic evolution and semantic relations of polysemous function words. *Journal of Shanxi University (Philosophy and Social Sciences Edition)*, 35(3):151–159.

Rui Guo. 2012c. The typology of adjectives and the grammatical status of chinese adjectives. *Chinese Language Learning*, (5).

Martin Haspelmath. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello, editor, *The new psychology of language*, volume 2, pages 211–243. Lawrence Erlbaum, Mahwah, NJ.

Ekkehard König. 1977. Temporal and non-temporal uses of 'noch' and 'schon' in german. *Linguistics and Philosophy*, 1(2):173–198.

Feng-hsi Liu. 2000. The scalar particle hai in chinese. *Cahiers de Linguistique - Asie Orientale*, 29(1):41–84.

Andrej Malchukov, Martin Haspelmath, and Bernard Comrie. 2007. Ditransitive constructions: A typological overview. In *Paper for the Conference on Ditransitive Constructions*, Max Planck Institute for Evolutionary Anthropology, Leipzig.

Laura A. Michaelis. 1996. Cross-world continuity and the polysemy of adverbial still. In Gilles Fauconnier and Eve Sweetser, editors, *Space, World and Grammar*. The University of Chicago Press, Chicago and London.

Jiaxuan Shen. 2001. Two constructions with hai in chinese. *Studies of the Chinese Language*, 6:483–493.

Ma Teng. 2015. Computer-aided construction of the semantic map model. In Li Xiaofan, Zhang Min, Guo Rui, and et al., editors, *Research on Semantic Maps of Multifunctional Grammatical Forms in Chinese*, pages 194–206. The Commercial Press, Beijing.

Xiao-e Tong. 2004. The semantic evolution and the semantic network of adverb hai. *Journal of Southwest University for Nationalities. Humanities and Social Sciences*, 25(8):448–452.

Shan Xiao, Yalu Liao, and Chenghui Xu. 2021. A review and prospect of research on semantic maps in chinese: 2003-2021. *Modern Linguistics*, 9(5):1334–1344.

Min Zhang. 2015. The semantic map model: Principles, operation, and application in the study of mandarin multifunctional grammatical forms. In Xiaofan Li, Min Zhang, and Rui Guo, editors, *The Semantic Map Research on Mandarin Multifunctional Grammatical Forms*, pages 3–54. Commercial Press, Hong Kong.

Ying Zhang. 2017. Semantic map approach to universals of conceptual correlations: a study on multifunctional repetitive grams. *Lingua Sinica*, 3(1):7.