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

Zhu Liu, Cunliang Kong, Ying Liu, Maosong Sun
liuzhu22@tsinghua.org.cn

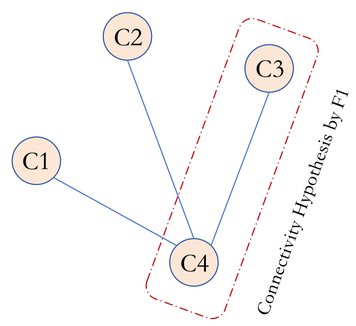
NAACL 2025

➢ What are classical Semantic Maps?

A graph model to represent similarity and entailment relationships in domain-specific concept comparions for cross-lingual forms

		C1	C2	C3	C4
L1	F1			✓	✓
L2	F2		✓		✓
L3	F3	✓			✓
L3	F4	✓			✓

Form-Concept Table
L - Languages; F - Forms; C - Concepts



Semantic Map/Conceptual Space

Node: Concept / Meaning / Function / Sense
Forms: Function words or content words / Affixes / Constructions
Edge: Co-lexification Association between nodes
Form-Concept Table: Patterns of multifunctionality/Polysemous

Connectivity Hypothesis (H1): concepts associated with a form (colexification) must be connected (minimal topological distance)
Economic Priciple: The edges should be as few as possible.

➢ Example

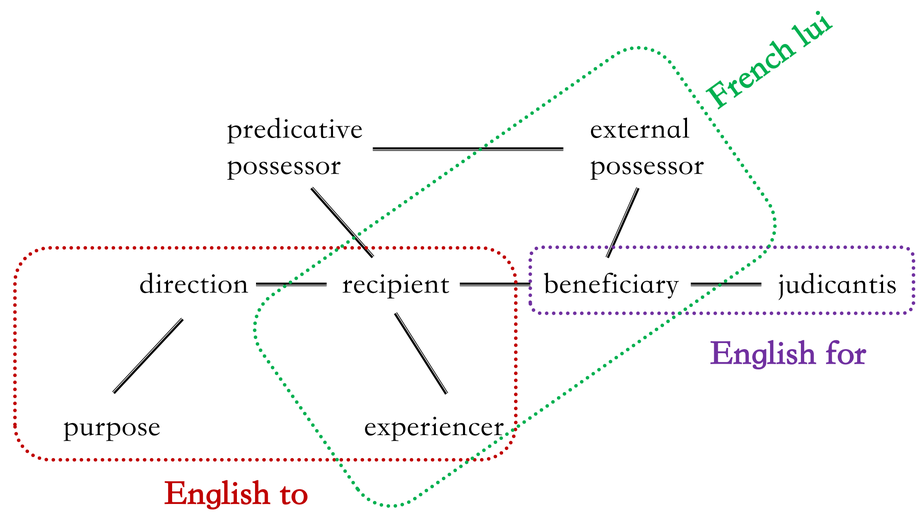
Domain: External-possession construction

English preposition to
a. Goethe went to Vienna as a student. (direction)
b. Eve gave the apple to Adam. (recipient)
c. This seems outrageous to me. (experiencer)
d. I left the party early to get home in time. (purpose)

English preposition for
a. My mom bought a bike for the son. (benefactive)
b. The math is too difficult for me. (judicantis)

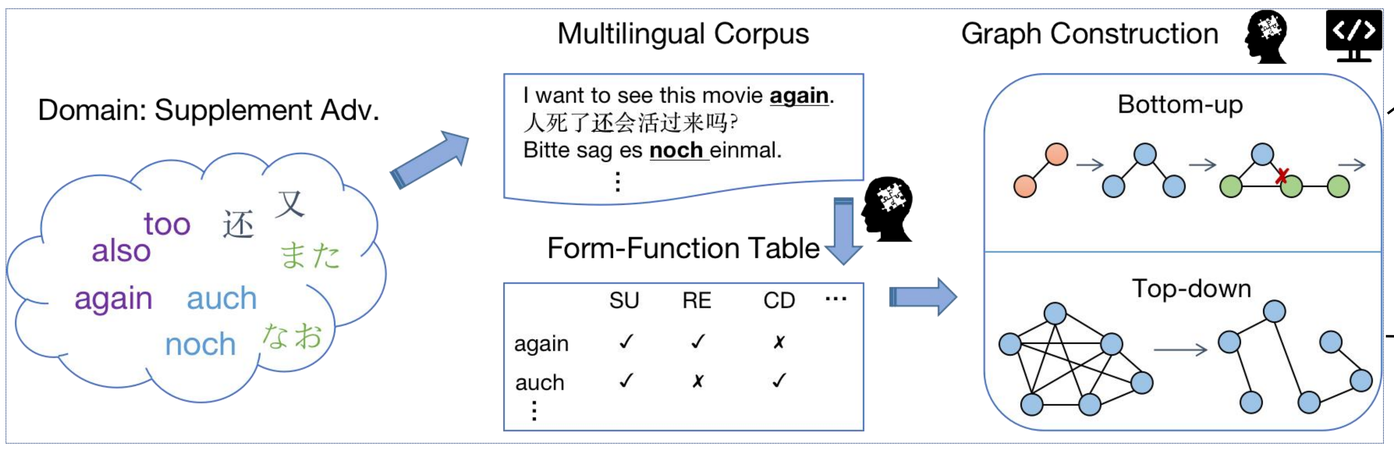
French lui (DAT-case)
a. Je lui donne le livre. [I'll give her the book] (recipient)
b. Je lui ai trouvé unemploi [I found a job for him.] (benefactive)
c. On lui a cassé la jambe [They broke his leg] (ext. possessor)
d. Ce livre lui plaît [He likes this book] (experiencer)

Russian: u, DATive-case
Chinese, Catalan, Homeric Greek, Basque ...



Haspelmath M. External possession in a European areal perspective[J]. Typological studies in language, 1999, 39: 109–136.

➢ How to build classical Semantic Maps?



Bottom-up approach (Conventional)
- Expert knowledge or prior programs
- Incrementally satisfy connectivity hypothesis (H1) case by case
- Avoid too many redundant edges when necessary
- Time-consuming and hard to scale

Top-down approach (**ours**)
- Initially generate a complete graph by calculating the number of bifunctional co-occurrences
- Convert local connectivity H1 into a global one H2
- Efficient and automated

H2: A Maximum spanning tree from the complete graph
- Overall connectivity
- Acylicity (Economic Principle & Avoid Null Predictivity)
- Maximum total edge weights (H1, proved)

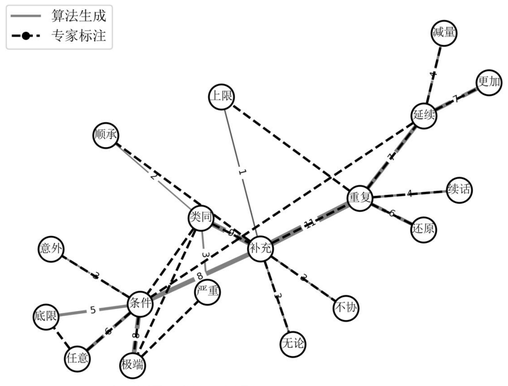
➢ Evaluation on A Case Study

Overview
- 36 languages/dialects, 82 grammatical forms and 18 functions
- Supplementary Adverbs (The core meaning is to express supplementation)
- The form-function table has been created by linguists
- The golden map (GT) in a bottom-up fashion serves as a reference standard

Metrics

Metric	Description	Trend
Size	Total edge weight sum	↔
Cov	Proportion of covered instances	↑
Avg_D	Average node degree	↔
Div_D	Standard deviation of node degrees	↓
Acc	Matched / total elements in $\mathcal{G}$	↑
Prec	Matched edges / generated edges	↑
Rec	Matched edges / annotated edges	↑
F1	Harmonic mean of Prec and Rec	↑

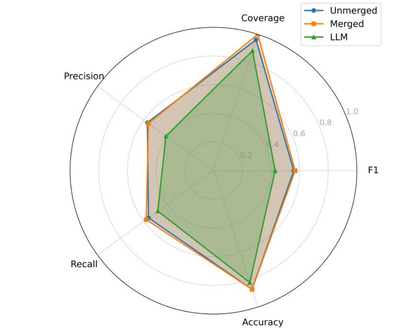
Baselines
- GT
- Ours (Unmerged vs. Merged)
- LLM (by Prompt)
- Human Judgment

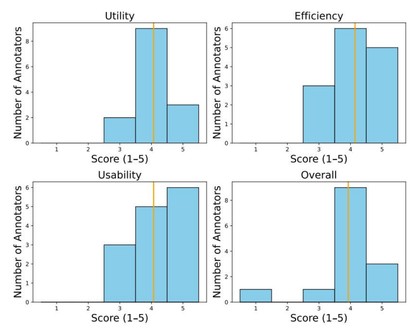


《清华语言学第八辑》,2026, P194–213

Metrics	Unmerged	Merged	LLM	GT
Size↔	186.50	**200.20**	237.30	204.30
#Edges↔	15.60	**16.70**	23.10	16.82
Div_D↓	**1.40**	1.48	2.00	1.472
Avg_D↔	1.83	**1.94**	2.58	1.97
Time(s)↓	**3.25**	3.26	379.60	-

https://arxiv.org/pdf/2507.04070





➢ Implications

- Second Language Acquisition
- Language Evolution
- .....

Paper V2