MultiSpider: Towards Benchmarking Multilingual Text-to-SQL Semantic Parsing

AAAI 2023 Main Conference

Longxu Dou[1], Yan Gao[2], Mingyang Pan[1], Dingzirui Wang[1], Wanxiang Che[1], Dechen Zhan[1], Jian-Guang Lou[2]
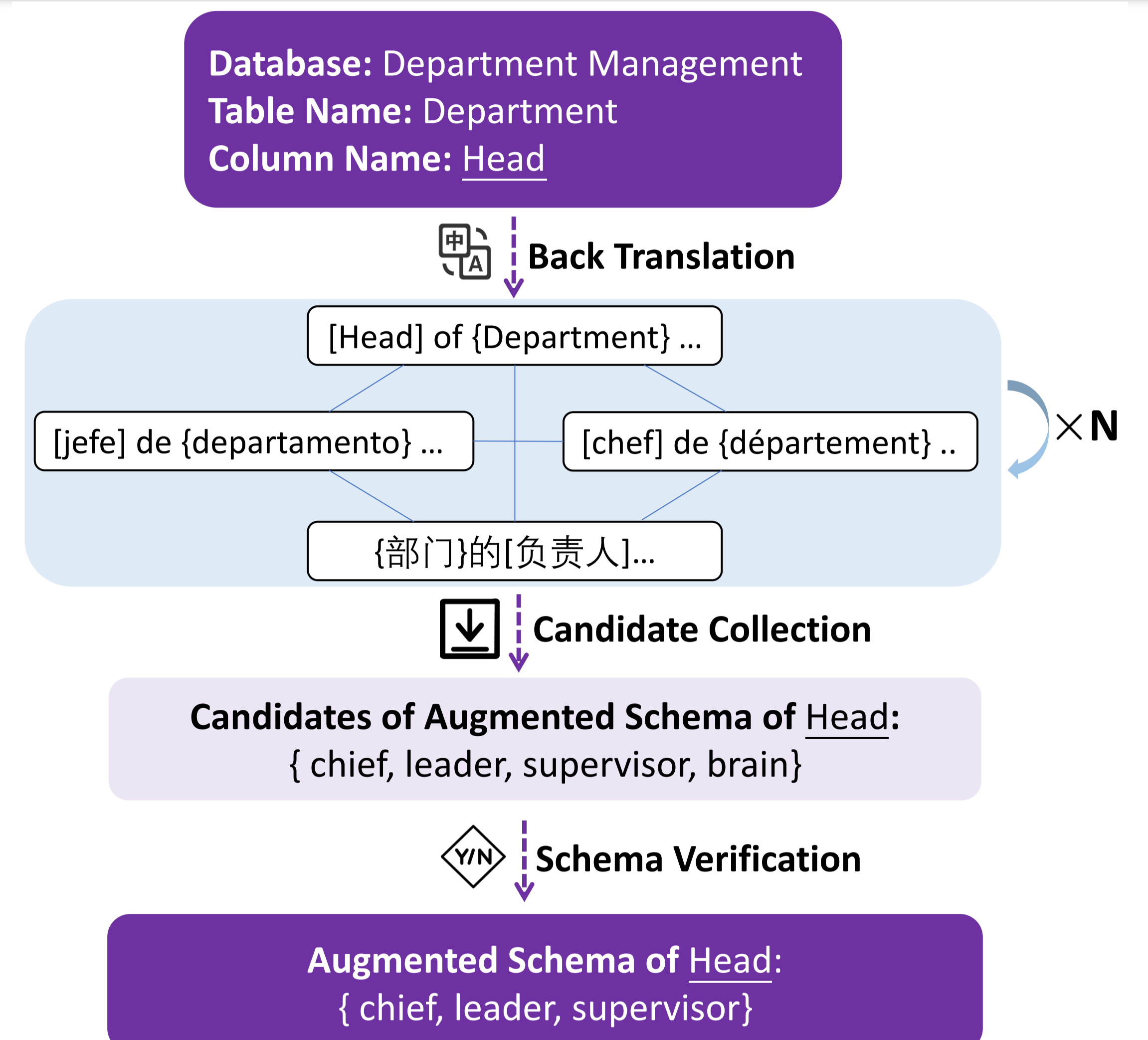
[1]Harbin Institute of Technology [2]Microsoft Research Asia

❖ Contribution: Multilingual Text-to-SQL Benchmark & Schema-Augmentation

Lang.	Question	Schema with Augmentation
English	Return the record companies of orchestras, sorted descending by the years in which they were founded.	year of founded {established year, the year of foundation}, …
German	Geben Sie die Plattenfirmen von Orchestern zurück, absteigend sortiert nach den Jahren ihrer Gründung.	Gründungsjahr {jahr der grundlage, jahr der gründung}, …
French	Listez les maisons de disques des orchestres, triées par ordre décroissant des années de leur création.	année de foundation {année de creation}, …
Spanish	¿Cuáles son las compañías discográficas de las orquestas en orden descendente de años de fundación?	año de fundación {Año Establecido, año de creación}, …
Chinese	返回按创立年份降序排列的乐团唱片公司的名称。	成立年份 {创立之年, 建立之年}, …
Japanese	創設年の降順でオーケストラのレコード会社を並べる。	創設年 {創業年, 設立年}, …
Vietnam	Liệt kê các công ty thu âm của các dàn nhạc theo thứ tự giảm dần về năm mà từng công ty được thành lập .	năm thành lập {năm sáng tạo}, …

SQL: SELECT Record_Company FROM orchestra ORDER BY Year_of_Founded DESC

SQL: SELECT レコード・レーベル FROM オーケストラ ORDER BY 創設年 DESC

Database: Department Management
Table Name: Department
Column Name: Head

Back Translation

[Head] of {Department} …

[jefe] de {departamento} … [chef] de {département} ..

×N

{部门}的[负责人]…

Candidate Collection

Candidates of Augmented Schema of Head:
{ chief, leader, supervisor, brain}

Schema Verification

Augmented Schema of Head:
{ chief, leader, supervisor}

MultiSpider: The **largest** multilingual text-to-SQL dataset, including **seven** mainstream languages. Translate both question&schema.

SAVe: a data-augmentation method from the perspective of **schema**. Back Translation + Schema Verification

❖ Dataset Analysis: High-Quality & More-Challenging

Type	Schema	Mistake	Correction
Abbreviation	aid	援助 (assistance)	作者ID (ID of the author)
	did	做了 (done)	领域ID (ID of the domain)
Jargon	body builder	造车者 (carmakers)	健美运动员 (muscle-builder)
	snatch	抢夺 (wrest)	挺举 (weightlifting)
Polysemy	player	演员 (actor)	运动员 (athlete)

Inaccurate Translation (Question)

Lexical	Spider: What capital is the largest in the us? (DB: Geo)
	CSpider: 美国最大的资本是什么? (money)
	MultiSpider: 美国最大的州会是什么? (metropolis)
Structural	Spider: List names of conductors in descending order of years of work.
	SQL: SELECT Name FROM conductor ORDER BY Year_of_Work DESC
	Google: コンダクターの名前と降順での勤務年数を示す? (List both name and year)
	MultiSpider: 勤務年数の降順での指揮者の名前は? (List only name)

We identify five **typical translation mistakes**.

Lexical Challenge	Explanation
Question:有多少不同的获胜者都参加了 "wta championships",并且都是左撇子? (How many different winners both participated in the WTA Championships and were left-handed?) Gold: SELECT count(DISTINCT winner_name) FROM matches WHERE tourney_name = 'WTA Championships' AND winner_hand = 'L'	Mention: 左撇子 Schema: 惯用手 (Slang)
Question:最小バージョン番号とそのテンプレートタイプコードは? (What the smallest version number and its template type code?) Gold: SELECT min(Version_Number) , template_type_code FROM Templates	Mention: バージョン番号 Schema: バージョンナンバー (Hiragana and Katakana)
Question: 每个国家中的被最多人讲的主流语言是什么? (What is the language spoken by the largest percentage of people in each country?) Gold: SELECT Language , CountryCode , max(Percentage) FROM countrylanguage GROUP BY CountryCode	Mention: 主流 Schema: 百分比 (Semantic Match)

Structural Challenge	Explanation
Question: 按照从老到少的顺序输出老师的姓名? (List the names of teachers in ascending order of age.) Gold: SELECT Name FROM teacher ORDER BY Age ASC	Mention: 从老到少 Operator: ORDER BY Age ASC (Dialect)
Question: 成績証明書のリリースの最も早い日付は何ですか?詳細を教えてくださ (What is the earliest date of a transcript release, and what details can you tell me?) Gold: SELECT transcript_date , other_details FROM Transcripts ORDER BY transcript_date ASC LIMIT 1	Mention: 最も早い日 Operator: ORDER BY Date ASC (Commonsense)

- The **specific language properties** like Hiragana and Katakana (Japanese).
- The **morphologically rich** language (German and French).
- The **dialect and slang sayings** require further commonsense reasoning.

❖ Experimental Results

Model	DE	ES	FR	JA	ZH	VI
Directly Predict						
mBERT	50.9	52.2	50.7	43.1	49.6	45.3
XLM-R	57.6	60.8	59.1	48.3	55.5	56.5
Translate-then-Predict						
mBERT	49.6	51.2	47.6	39.1	46.7	43.3
XLM-R	58.8	57.2	58.7	46.3	55.3	53.8
Translate-then-Train						
mBERT	49.5	51.2	51.3	38.2	45.8	49.3
XLM-R	**60.2**	**61.9**	**61.7**	**51.3**	**57.6**	**63.9**

Zero-shot

Strong PLM with strong MT yields the promising zero-shot performance.

Model	EN	DE	ES	FR	JA	ZH	VI
Monolingual Training (only use target language training data)							
mBART	57.3	39.7	41.3	37.5	45.7	55.0	42.2
mBART + SAVᴇ	58.3	42.6	42.6	51.2	46.9	56.6	43.1
RAT-SQL + XLM-R	68.6	62.5	61.7	64.1	53.1	63.4	65.9
RAT-SQL + XLM-R + SAVᴇ	68.8	63.9	62.7	65.7	54.3	66.2	66.1
Multilingual Training (use training data from multiple languages)							
mBART	58.3	42.7	45.9	42.9	52.2	57.8	43.2
mBART + SAVᴇ	59.7	46.9	47.1	43.0	54.3	61.9	45.6
RAT-SQL + XLM-R	68.8	64.8	67.4	65.3	60.2	66.1	67.1
RAT-SQL + XLM-R + SAVᴇ	70.8	66.7	69.3	67.5	61.6	67.3	67.8

Monolingual & Multilingual

The performance of Japanese is significantly behind other languages.
The absolute drop of accuracy in non-English languages is about 6.1%.
SAVe significantly improves the non-English languages (1.4%-1.9%).