



Mahidol University  
*Wisdom of the Land*

# Digital platform for language diversity in Thailand

Siripen Ungsitipoonporn



Sarawut Kraisame



Chingduang Yurayong

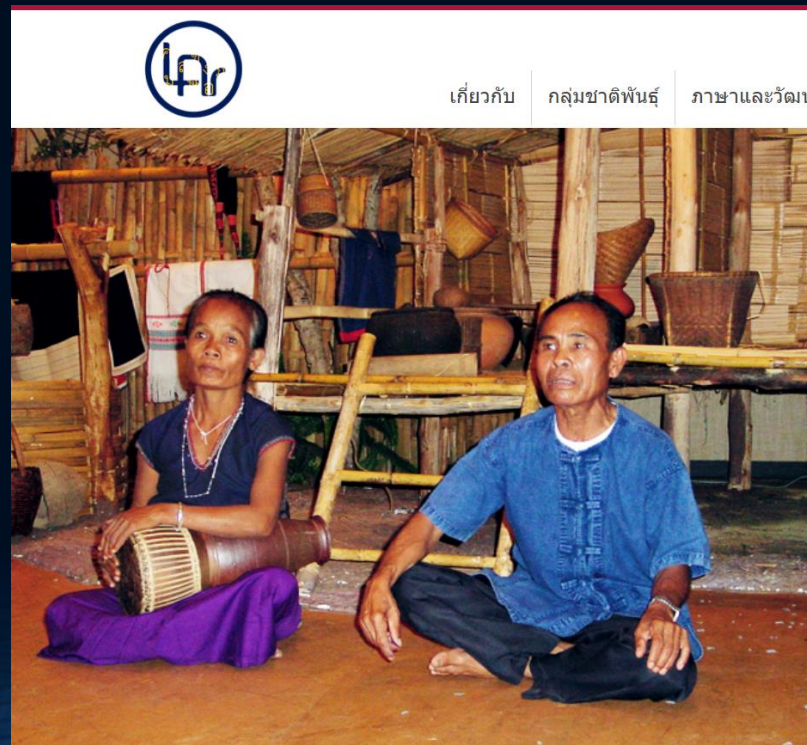


# Background

- A 40-year long tradition of language documentation at Mahidol University
- A great amount of data on languages spoken in Thailand has been collected
- A large part of the collected data still remains unannotated, unorganised and unsystematised for further uses

# Currently available language resources

- LangArchive-TH [www.langarchive-th.org](http://www.langarchive-th.org)
  - ✓ 18 linguistic-ethnic groups (Austroasiatic, Austronesian, Sino-Tibetan, Tai-Kadai)
  - ✓ 227 video files
  - ✓ 165 sound files
  - ✓ 148 text documents
  - ✓ 365 pictures
  - ✓ 90 ELAN files



## ยินดีต้อนรับสู่ LangArchive-TH

คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ในประเทศไทย

Archive of Languages and Cultures of Ethnic Groups of Thailand  
สถาบันวิจัยภาษาและวัฒนธรรมเอเซีย มหาวิทยาลัยมหิดล

สนับสนุนการอนุรักษ์ความหลากหลายทางภาษาและภูมิปัญญาทางวัฒนธรรมของกลุ่มชาติพันธุ์ และมุ่งเน้นพัฒนาคลังข้อมูลที่ชุมชนในท้องถิ่นสามารถเข้าถึงได้

ภายใต้โครงการ “คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์: ส่งเสริมการปกป้อง และดำรงรักษา ความหลากหลายทางภาษาและวัฒนธรรมในประเทศไทย” ร่วมกับ Endangered Languages Archive (ELAR), SOAS University of London ด้วยทุนสนับสนุนจาก Newton Fund

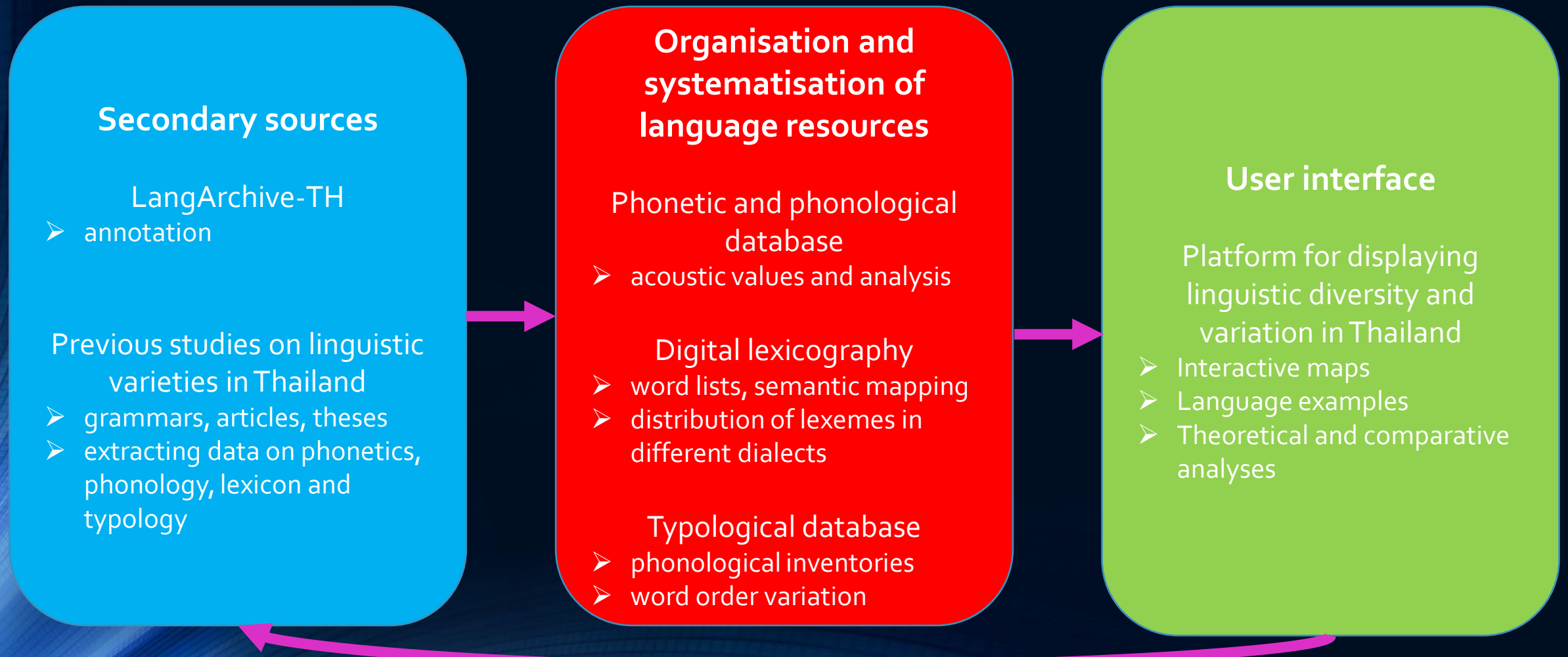


# Currently available language resources

- Previous studies on 5 major language families in Thailand (Austroasiatic, Austronesian, Hmong-Mien, Sino-Tibetan, and Tai-Kadai)
  - ✓ Grammars
  - ✓ Research publications
  - ✓ MA and PhD theses
- More than 100 varieties

Language family	Estimated number of linguistic varieties with available descriptive data	Estimate number of previous researches on the given linguistic varieties
1. Austroasiatic	28	50
2. Austronesian	9	30
3. Hmong-Mien	3	10
4. Sino-Tibetan	20	40
5. Tai-Kadai	44	100
Total	104	230

# Planned implementation



# Some models for creating digital platform

- 中国语言地图集 Zhōngguó yǔyán dìtújí [Language atlas of China] (Zhang 2012)
- Grammar Atlas of Japanese Dialects (NINJAL)
- Atlas of the Languages of Iran (Anonby, Taheri-Ardali et al. 2015-2020)
- Algonquian Linguistic Atlas (Junker et al. 2005)
- The Database of Eurasian Phonological Inventories (Nikolaev 2018)
- WALS (Dryer & Haspelmath 2013)

# Sub-project

- Word order variation among linguistic varieties spoken in Thailand

Construction	Head-initial	Head-final
1. Adjective	Noun-Adjective	Adjective-Noun
2. Demonstrative	Noun(-Classifier)-Demonstrative	Demonstrative(-Classifier)-Noun
3. Possessive	Possessed-Possessor	Possessor-Possessed
4. Animal gender modifier	Animal-Gender	Gender-Animal
5. Intensifier	Adjective-Intensifier	Intensifier-Adjective
6. Quantity	Noun-Number(-Classifier)	Number(-Classifier)-Noun
7. Temporal adverb	Verb-Temporal adverb	Temporal adverb-Verb
8. Adpositional phrase	Verb-Adpositional phrase	Adpositional phrase-Verb
9. Comparative	Adjective-Comparative-Standard	Comparative-Standard-Adjective
10. Negation	Negative-Verb	Verb-Negative
11. Direct object	Direct object-Verb	Verb-Direct object
12. Ditransitive	Direct object-Indirect object	Indirect object-Direct object
13. Existential	Exist-Noun	Noun-Exist
14. Relative clause	Noun-Relative clause	Relative clause-Noun
15. Complementiser	Clause-Complementiser	Complementiser-Clause

➤ Areal tendencies and syntactic reconstruction of proto-languages

# Sub-project

- Word order variation among linguistic varieties spoken in Thailand

			Austroasiatic	Austronesian	Tai-Kadai	Hmong-Mien	Sino-Tibetan	
			Khmu	Urak Lawoi'	Thai	Green Hmong	Red Karen	Burmese
Phrase type	Construction	Pattern	Word order	Word order	Word order	Word order	Word order	Word order
Noun	Adjective	ADJ- <b>N</b> -ADJ						
	Adnominal possession	POSS- <b>N</b> -POSS				+	+	+
	Animal gender marking	Mod- <b>N</b> -Mod				+		+
	Classifier	NUM-CLF- <b>N</b> -NUM-CLF				+		+
	Demonstrative	DEM- <b>N</b> -DEM						+
	Relative clause	REL- <b>N</b> -REL						+
Adjective	Intensifier	ADV- <b>ADJ</b> -ADV	Head	Head	Head	Head	Head	Head
	Comparative	STD- <b>ADJ</b> -STD						
Verb	Adjunct	ADV- <b>V</b> -ADV		+		+		+
	Direct object	DO- <b>V</b> -DO						+
	Ditransitive	IOBJ- <b>DOBJ</b> -IOBJ					+	+
	Negation	V- <b>NEG</b> -V					+	+
	Predicative possessive	OBJ- <b>V</b> -OBJ					+	+

➤ Areal tendencies and syntactic reconstruction of proto-languages