MAX PLANCK INSTITUTE
FOR EVOLUTIONARY ANTHROPOLOGY

Representing Semantic Networks in Concepticon

Annika Tjuka und Johann-Mattis List

- *Dept. of Linguistic and Cultural Evolution, MPI-EVA*
- *Chair for Multilingual Computational Linguistics, Uni Passau*

AGENDA

1 Introduction to Conception

2 Challenge: Network-like Concept Lists

3 Best Case Example: Partial Colexifications (List 2023)

4 Conclusion

Introduction to Concepticon

Concepticon

A resource of concept and word lists that offers standardized concept sets and links to glosses. It serves as a reference catalog for historical and typological language comparison.

- Includes cross-linguistically comparable concepts such as HAND, TREE, YOU, or GIVE.
- Concept lists are used to elicit the glosses for the concepts across languages.
- Usually small lists of up to 300 concepts compiled by historical linguists and linguistic field workers.

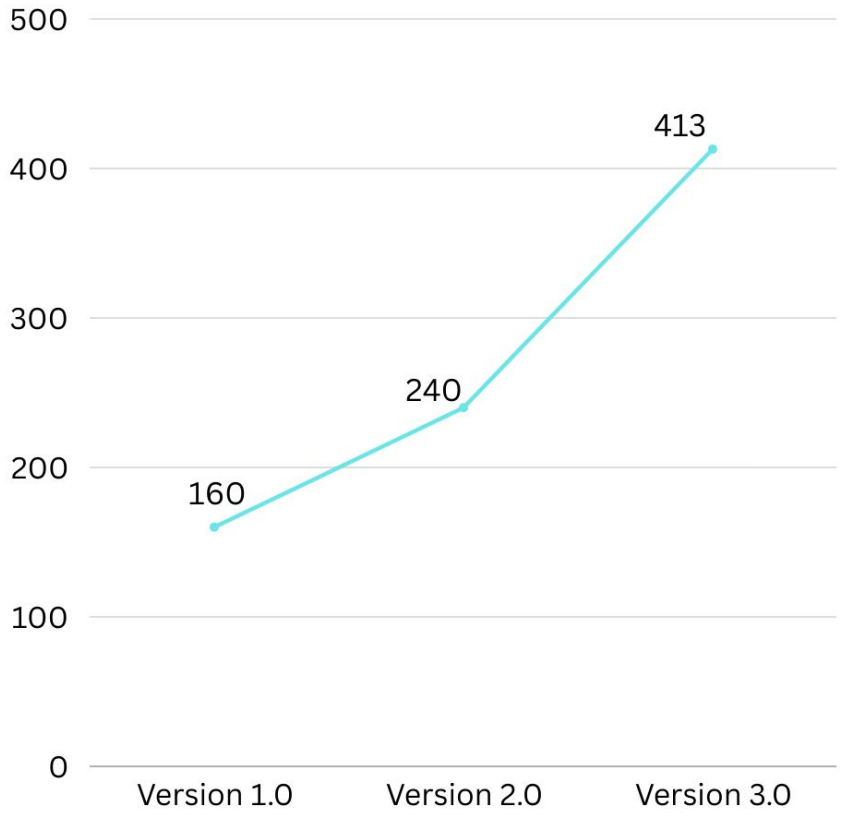
CLLD web app: https://concepticon.clld.org/

GitHub repository: https://github.com/concepticon/concepticon-data

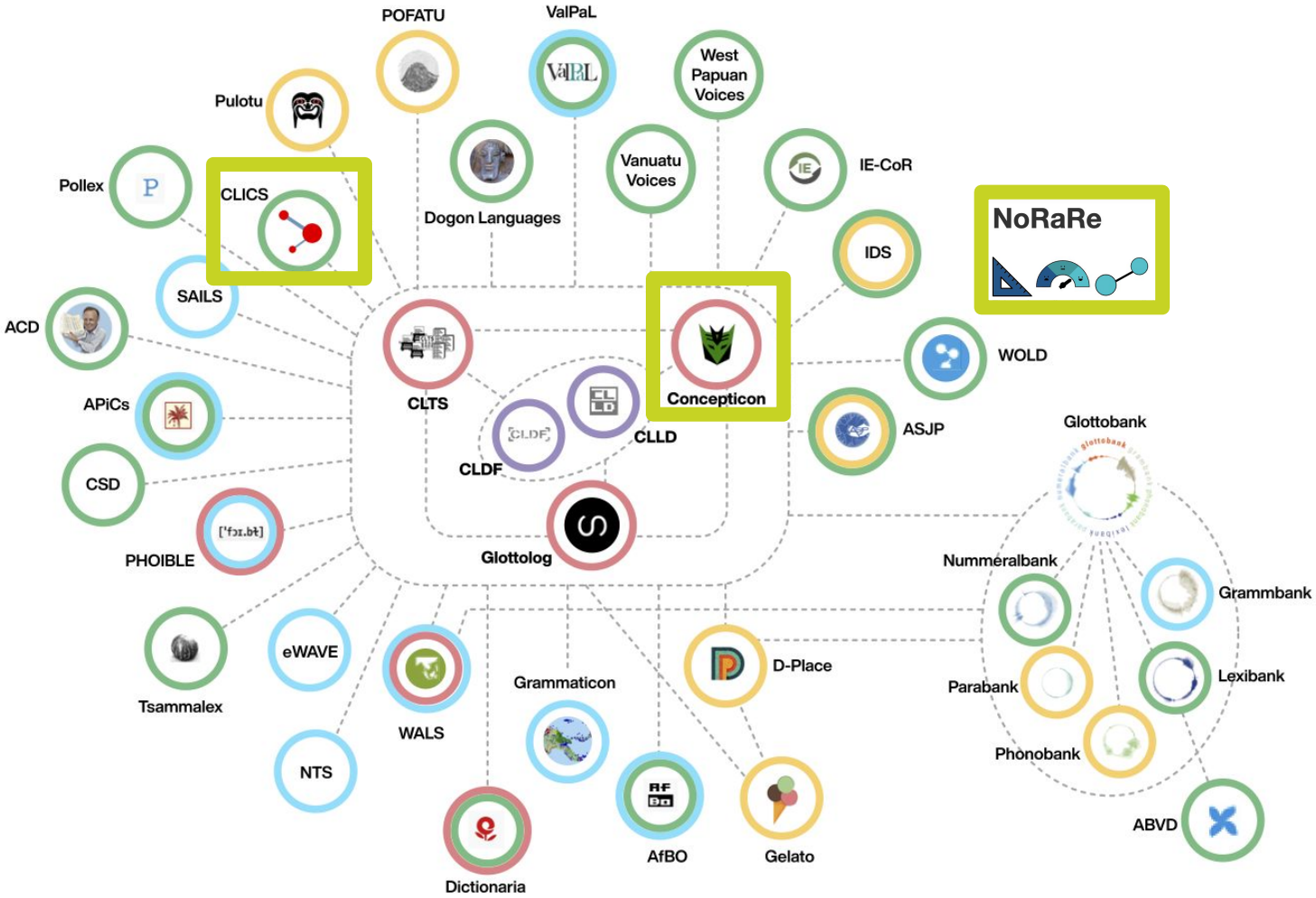
Concepticon

- The Concepticon was established in 2015 and the first major release was in 2016 with 162 concept lists (Concepticon version 1.0: List et al. 2016a,b).

- The most recent major release with 413 concept lists was in 2023 (Concepticon version 3.0: Tjuka et al. 2023; List et al. 2022).

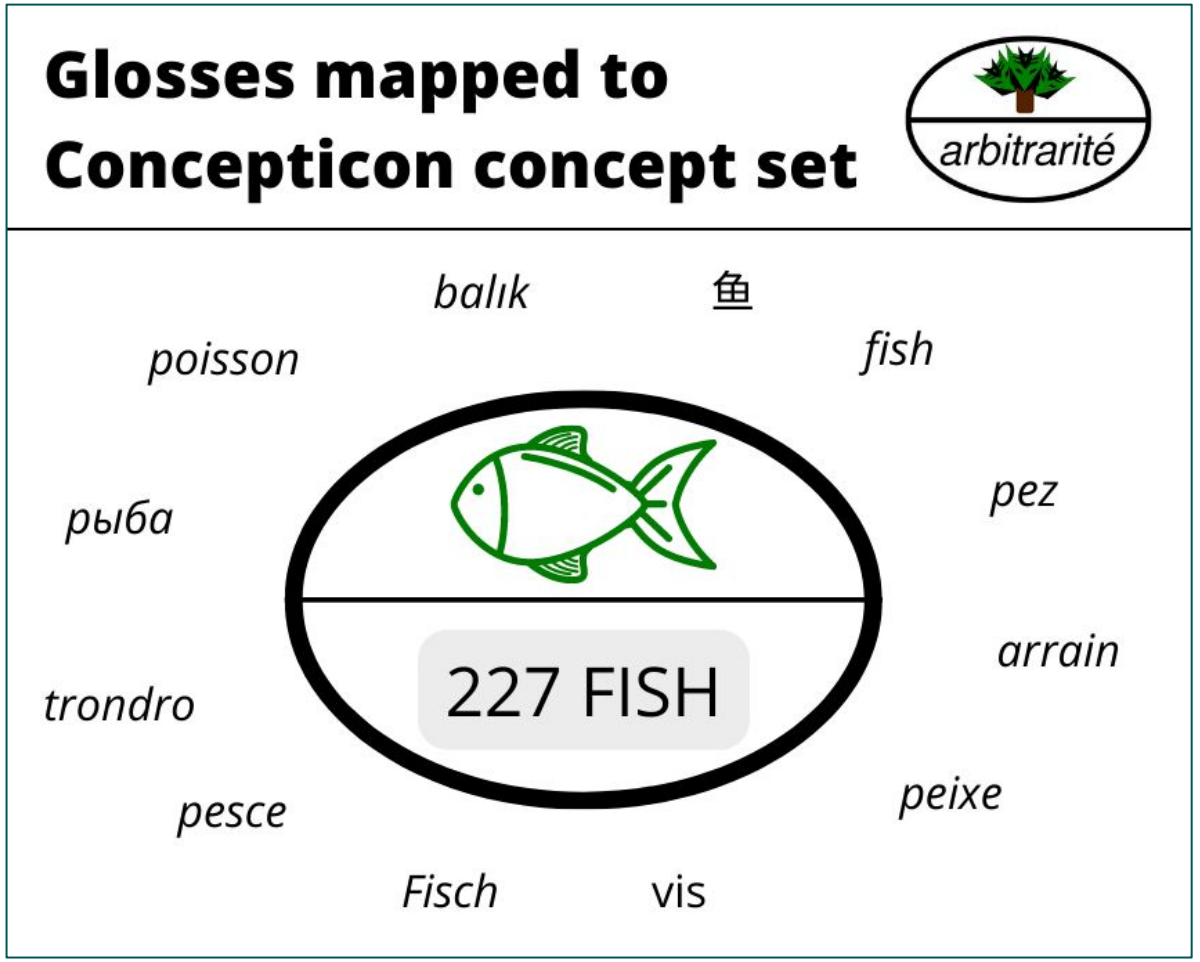


Concepticon



Concepticon Concept Sets

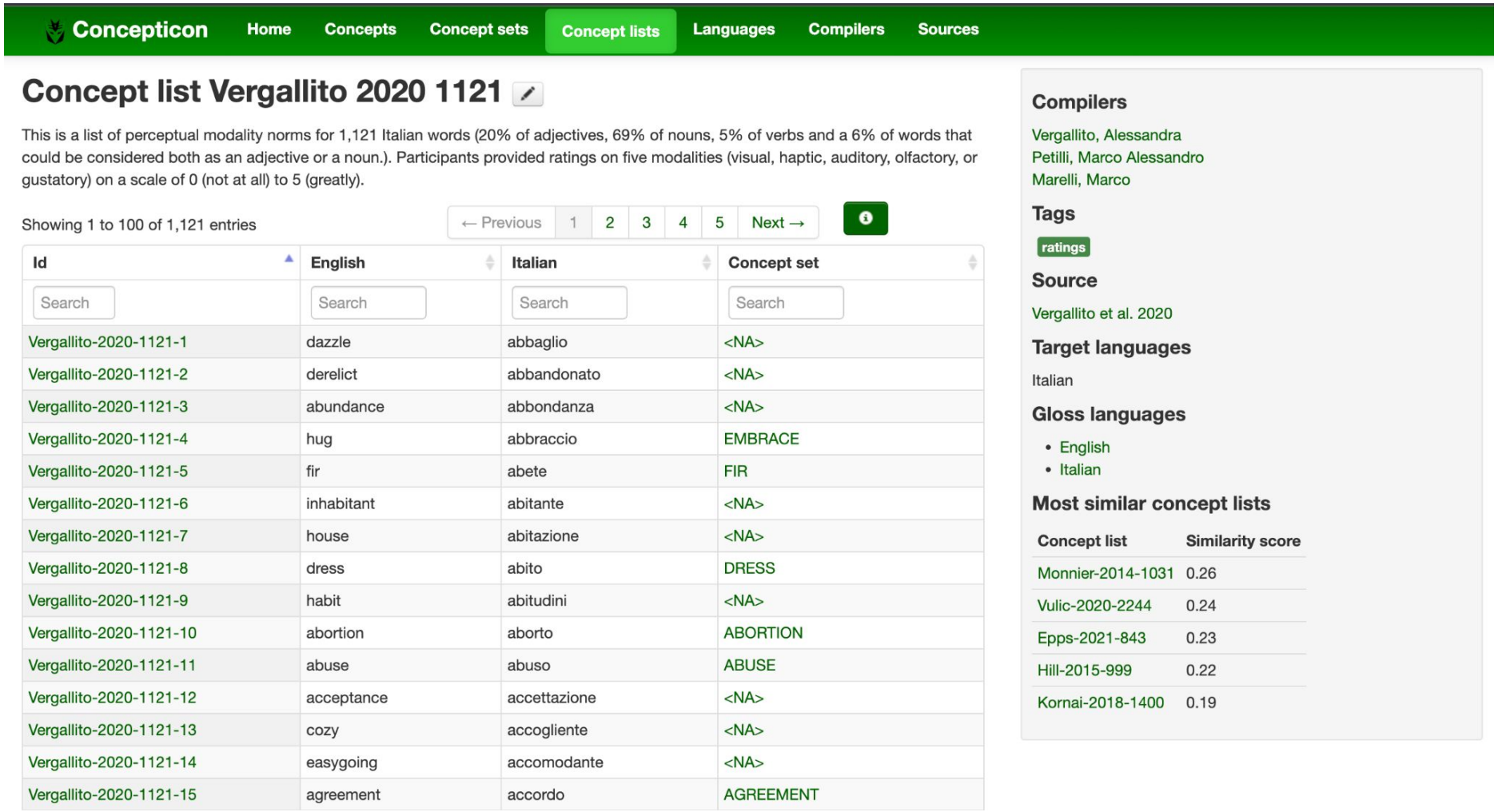
- Consist of a unique identifier, a label, a definition, a semantic field, and an ontological category.
- Concept identifiers (e.g., “227”) are connected to a unique label (e.g., “FISH”).
- Concepticon concept sets reflect concepts that are deemed interesting for comparison by linguists and occur frequently in concept lists (List et al. 2016).
- Elicitation glosses are established by linguists and are often based on already existing concept lists.



Concepticon Mapping

ID	NUMBER	ITALIAN	ENGLISH	CONCEPTICON_ID	CONCEPTICON_GLOSS	WORD_LENGHT	POS
Vergallito-2020-1121-1	1	abbaglio	dazzle			8	n
Vergallito-2020-1121-2	2	abbandonato	derelict			11	a
Vergallito-2020-1121-3	3	abbondanza	abundance			10	n
Vergallito-2020-1121-4	4	abbraccio	hug	928	EMBRACE	9	n
Vergallito-2020-1121-5	5	abete	fir	1915	FIR	5	n
Vergallito-2020-1121-6	6	abitante	inhabitant			8	n
Vergallito-2020-1121-7	7	abitazione	house	1252	HOUSE	10	n
Vergallito-2020-1121-8	8	abito	dress	474	DRESS	5	n

Concepticon



Challenge: Network-like Concept Lists

Mapping Urban (2011)

- Analysis of the asymmetry in overt marking to establish which of the two concepts is the target and source concept.

- Synchronic data from a balanced sample of 149 languages and 47 concept pairs across four categories

Table 1. Cross-linguistic asymmetries in overt marking

	Semantic association	Number of languages with polysemy Example	Number of languages with overt marking Example	Cross-linguistically unmarked member of the meaning pair
INTRA-DOMAIN ASSOCIATIONS: NATURE > NATURE				
1.	'cloud' ~ 'fog/mist'[a]	24 Bakueri *limbaki*	7 White Hmong *pos huab* 'moist cloud'	'cloud'
2.	'sun' ~ 'moon'	17 Macaguán *jomét* ~ *-omét*	3 Lake Miwok *káwul híi* 'night sun'	'sun'
3.	'grass' ~ 'straw/hay'	11 Itzaj *su'uk*	10 Yoruba *koríko gbigbẹ* 'grass dry'	'grass'
4.	'smoke' ~ 'fog/mist'	11 Miskito *kiasma*	3 Otomí *'bipa* /'bifi-pa/ 'smoke-heat'	'smoke'

Mapping Urban (2011)

ID	NUMBER	ENGLISH	CONCEPTICON_ID	CONCEPTICON_GLOSS	SEMANTIC_CLASS_ID	CATEGORY	TARGET_CONCEPTS
Urban-2011-160-1	1	animal	619	ANIMAL	I	Natural and topological concepts	[{"id": "Urban-2011-160-6", "name": "bird", "polysemy": 8, "overt_marking": 8, "shift_id": 5}]
Urban-2011-160-2	2	ashes	646	ASH	I	Natural and topological concepts	[{"id": "Urban-2011-160-19", "name": "embers", "polysemy": 6, "overt_marking": 2, "shift_id": 10}]
Urban-2011-160-3	3	bark	1204	BARK	I	Natural and topological concepts	[]
Urban-2011-160-4	4	bay	663	BAY	I	Natural and topological concepts	[]
Urban-2011-160-5	5	beak	73	BEAK	I	Natural and topological concepts	[]
Urban-2011-160-6	6	bird	937	BIRD	I	Natural and topological concepts	[{"id": "Urban-2011-160-73", "name": "airplane", "polysemy": 3, "overt_marking": 3, "shift_id": 45}]

Mapping Winter & Srinivasan (2022)

- Analysis of the influence of concreteness and word frequency in the emergence of metaphor and metonymy

- Extended set of Urban's concept pairs (*n* = 71)

Table 1. 71 concept pairs that exhibit cross-linguistic asymmetries in semantic change based on Urban (2011) and a subsequent extension of this list by Matthias Urban; sources are listed to the left of targets. Translation equivalents for a particular concept are separated by a "/" (e.g., "fog/mist").

cloud~fog/mist	breast~milk	tongue~flame	feather~pen
sun~moon	mouth~lip	road/street/way~Milky Way	feather~beard
grass~straw/hay	belly/stomach~womb	bed~nest	star~meteroid
smoke~fog/mist	heart~belly/stomach	egg~testicle	river~Milky Way
steam~fog/mist	milk~nipple	sun~clock	mountain~valley
animal~bird	liver~lungs	seed~testicle	puddle~swamp
lake~swamp	car~train	shadow~mirror	sky~rain
smoke~dust	heart~kidney	mouth~estuary	river~spring
smoke~cloud	mirror~glasses	bird~airplane	river~river bed
tree~branch	heart~lungs	foam~lungs	moon~star
ashes~embers	(molar) tooth~jaw	ball~testicle	pus~brain
tree~forest	belly/stomach~navel	day~clock	eyebrow~eyelash
day~dawn	cheek~buttocks	bone~needle	eyelid~eyelash
flower/blossom~bud	mouth~cheek	day~cloud	fingernail~finger
day~noon	skin~bark	tooth~beak	eyeball~pupil
sun~noon	mouth~beak	seed~bud	
honey~wax	saliva/spittle~foam	dew~fog	
bone~horn	house~nest	soil/earth~dust	
river/stream~flood	mouth~estuary	flame~embers	

Mapping Winter & Srinivasan (2022)

- Analysis of the influence of concreteness and word frequency in the emergence of metaphor and metonymy

- Extended set of Urban's concept pairs (*n* = 71)

Table 1. 71 concept pairs that exhibit cross-linguistic asymmetries in semantic change based on Urban (2011) and a subsequent extension of this list by Matthias Urban; sources are listed to the left of targets. Translation equivalents for a particular concept are separated by a "/" (e.g., "fog/mist").

cloud~fog/mist	breast~milk	tongue~flame	feather~pen
sun~moon	mouth~lip	road/street/way~Milky Way	feather~beard
grass~straw/hay	belly/stomach~womb	bed~nest	star~meteroid
smoke~fog/mist	heart~belly/stomach	egg~testicle	river~Milky Way
steam~fog/mist	milk~nipple	sun~clock	mountain~valley
animal~bird	liver~lungs	seed~testicle	puddle~swamp
lake~swamp	car~train	shadow~mirror	sky~rain
smoke~dust	heart~kidney	mouth~estuary	river~spring
smoke~cloud	mirror~glasses	bird~airplane	river~river bed
tree~branch	heart~lungs	foam~lungs	moon~star
ashes~embers	(molar) tooth~jaw	ball~testicle	pus~brain
tree~forest	belly/stomach~navel	day~clock	eyebrow~eyelash
day~dawn	cheek~buttocks	bone~needle	eyelid~eyelash
flower/blossom~bud	mouth~cheek	day~cloud	fingernail~finger
day~noon	skin~bark	tooth~beak	eyeball~pupil
sun~noon	mouth~beak	seed~bud	
honey~wax	saliva/spittle~foam	dew~fog	
bone~horn	house~nest	soil/earth~dust	
river/stream~flood	mouth~estuary	flame~embers	

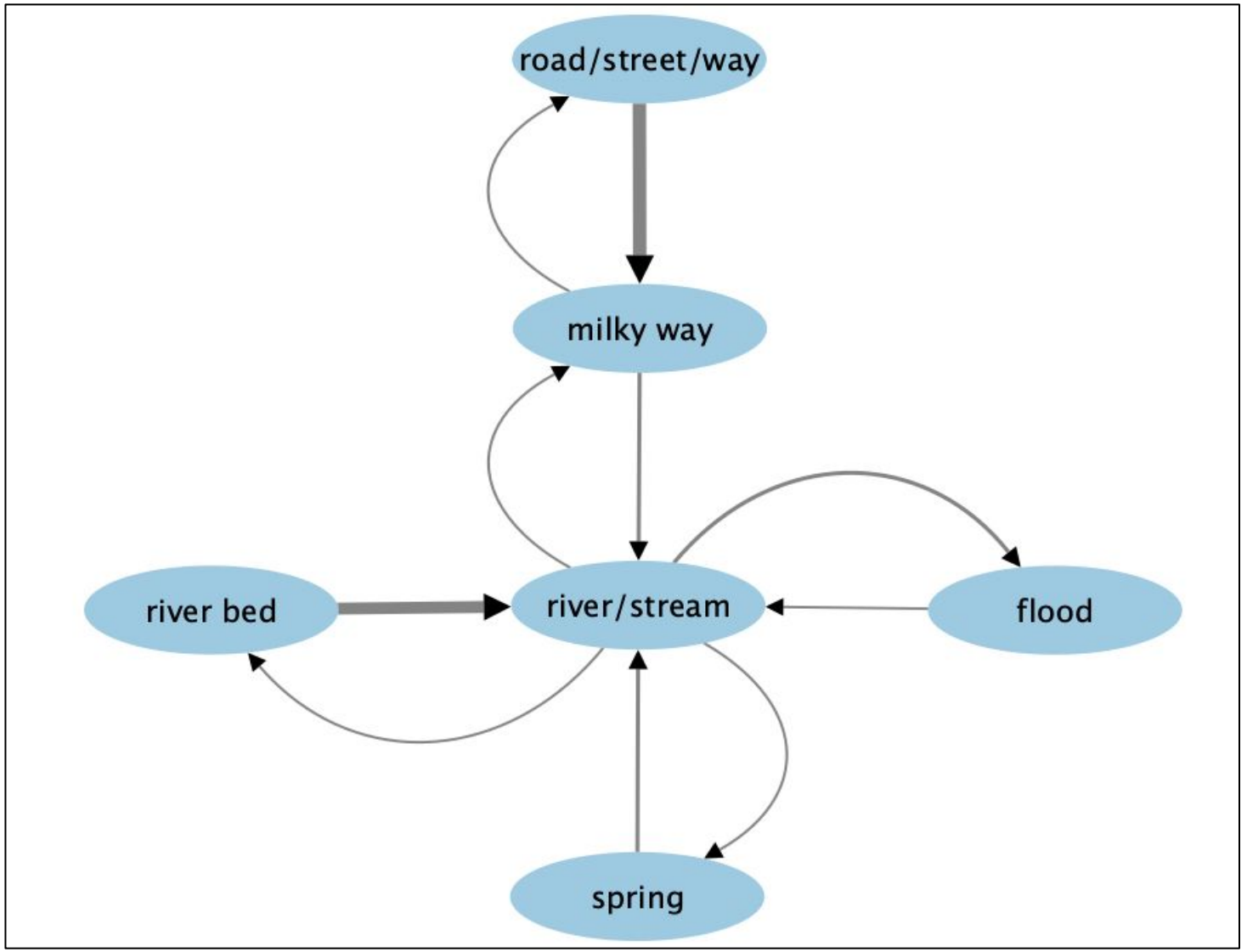
Mapping Winter & Srinivasan (2022)

ID	NUMBER	ENGLISH	CONCEPTICON_ID	CONCEPTICON_GLOSS	SOURCE_CONCEPTS	TARGET_CONCEPTS
Winter-2022-98-1	1	cloud	1489	CLOUD	[{"name": "fog/mist", "id": "Winter-2022-98-2", "polysemy": 24, "overt_marking": 0}, {"name": "smoke", "id": "Winter-2022-98-7", "polysemy": 7, "overt_marking": 2}, {"name": "day", "id": "Winter-2022-98-19", "polysemy": 2, "overt_marking": 0}, {"name": "sky", "id": "Winter-2022-98-82", "polysemy": 8, "overt_marking": 2}, {"name": "rain", "id": "Winter-2022-98-81", "polysemy": 4, "overt_marking": 2}]	[{"name": "fog/mist", "id": "Winter-2022-98-2", "polysemy": 24, "overt_marking": 7}, {"name": "smoke", "id": "Winter-2022-98-7", "polysemy": 7, "overt_marking": 0}, {"name": "day", "id": "Winter-2022-98-19", "polysemy": 2, "overt_marking": 3}, {"name": "sky", "id": "Winter-2022-98-82", "polysemy": 8, "overt_marking": 11}, {"name": "rain", "id": "Winter-2022-98-81", "polysemy": 4, "overt_marking": 2}]
Winter-2022-98-2	2	fog/mist	249	FOG	[{"name": "cloud", "id": "Winter-2022-98-1", "polysemy": 24, "overt_marking": 7}, {"name": "smoke", "id": "Winter-2022-98-7", "polysemy": 10, "overt_marking": 3}, {"name": "steam", "id": "Winter-2022-98-8", "polysemy": 9, "overt_marking": 2}, {"name": "dew", "id": "Winter-2022-98-71", "polysemy": 9, "overt_marking": 0}]	[{"name": "cloud", "id": "Winter-2022-98-1", "polysemy": 24, "overt_marking": 0}, {"name": "smoke", "id": "Winter-2022-98-7", "polysemy": 10, "overt_marking": 0}, {"name": "steam", "id": "Winter-2022-98-8", "polysemy": 9, "overt_marking": 0}, {"name": "dew", "id": "Winter-2022-98-71", "polysemy": 9, "overt_marking": 2}]

Edge List

SOURCE_ID	SOURCE_NAME	TARGET_ID	TARGET_NAME	OvertMarking
Winter-2022-98-1	cloud	Winter-2022-98-2	fog/mist	7
Winter-2022-98-1	cloud	Winter-2022-98-7	smoke	0
Winter-2022-98-1	cloud	Wnter-2022-98-19	day	3
Winter-2022-98-1	cloud	Winter-2022-98-82	sky	11
Winter-2022-98-1	cloud	Winter-2022-98-81	rain	2
Winter-2022-98-2	fog/mist	Winter-2022-98-1	cloud	0
Winter-2022-98-2	fog/mist	Winter-2022-98-7	smoke	0

Network Representation



Challenges

We face two major challenges when mapping lists that resemble network structures. First, concepts pairs are listed in one row. Second, the links between the concept pairs and their associated information, i.e., source/target, polysemy/overt marking, is not made explicit. We address these challenges by

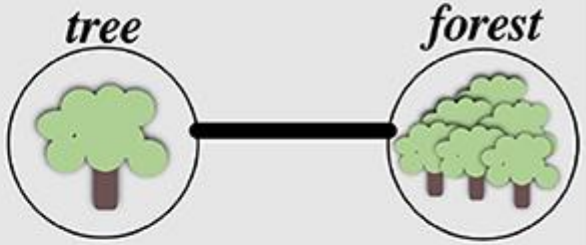
- Splitting the concept pairs and giving each concept a unique identifier,
- Including links that are machine-readable, and
- Aligning glosses with each other.

This process often reveals inconsistencies within the original dataset.

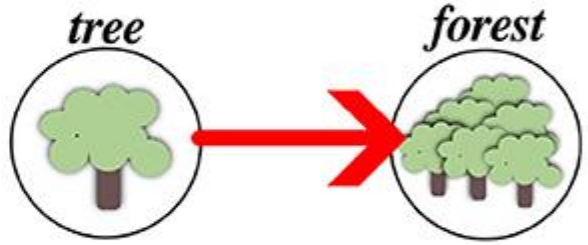
Best Case Example: Partial Colexifications (List 2023)

Partial Colexifications

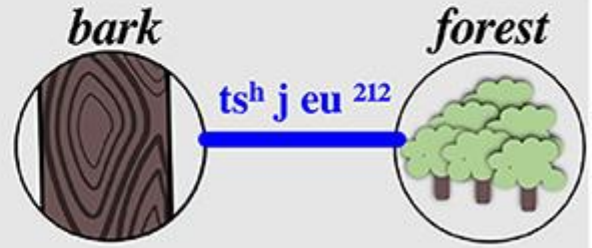
① Yaqui "tree": [dʒ u j a]
Yaqui "forest": [dʒ u j a]



② Guìlín "tree": [ɕ y ^{21}]
Guìlín "forest": [ɕ y ^{21} l i ŋ ^{22}]



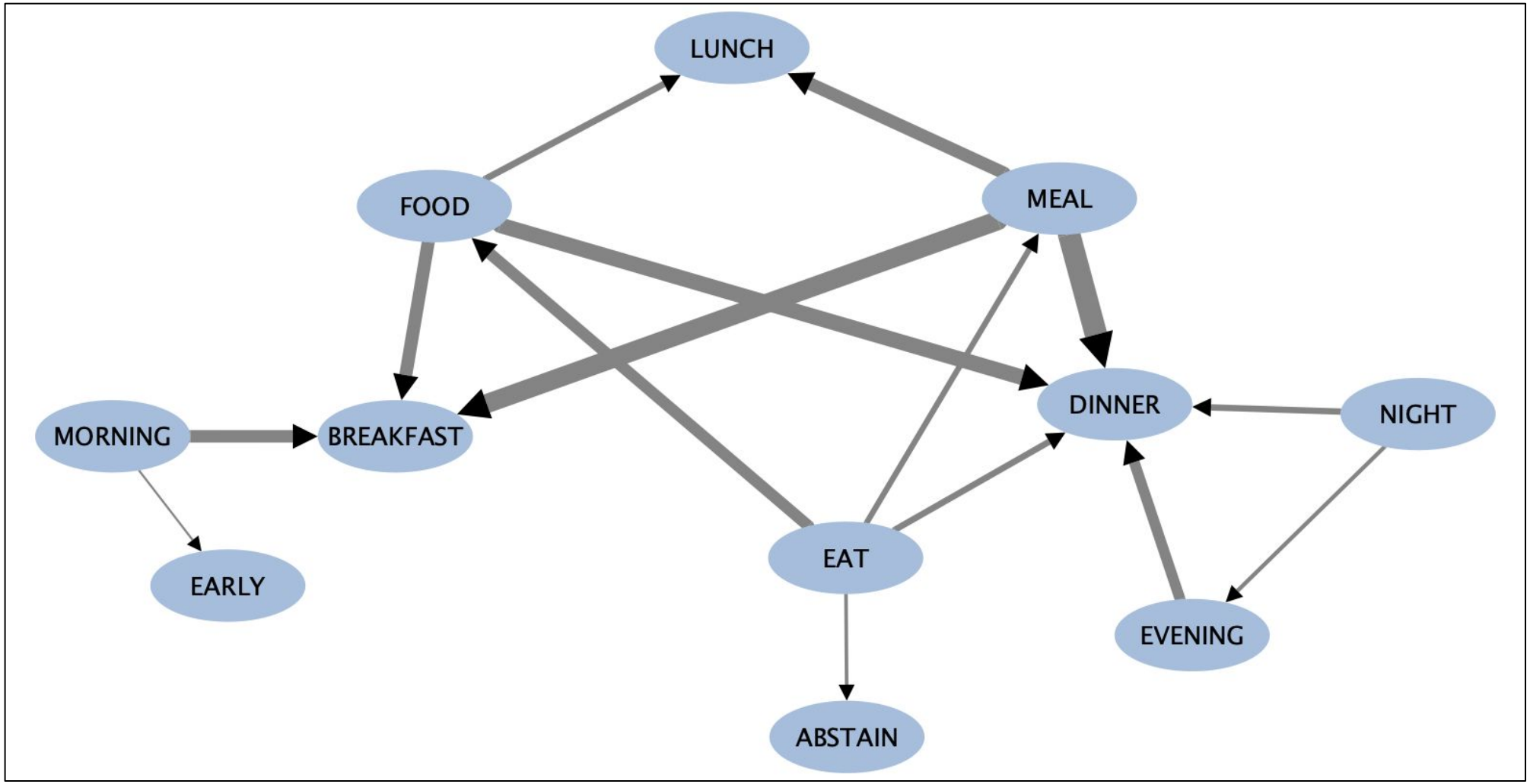
③ Fúzhōu "bark": [tsʰ j eu ^{212} pʰ w oi ^{5}]
Fúzhōu "forest": [tsʰ j eu ^{212} l i ŋ ^{53}]



Edge List of List (2023)

SOURCE_ID	SOURCE_NAME	TARGET_ID	TARGET_NAME	AffixVars	AffixLngs	AffixFams
List-2023-1308-1	WORLD	List-2023-1308-45	EARTHQUAKE	22	22	14
List-2023-1308-2	LAND	List-2023-1308-1	WORLD	19	19	13
List-2023-1308-2	LAND	List-2023-1308-1185	NATIVE COUNTRY	72	57	11
List-2023-1308-2	LAND	List-2023-1308-12	PLAIN	21	21	14
List-2023-1308-2	LAND	List-2023-1308-15	MAINLAND	41	41	12
List-2023-1308-2	LAND	List-2023-1308-240	WORM	11	11	10
List-2023-1308-2	LAND	List-2023-1308-4	DUST	14	14	11

Directed Network



Conclusion

Conclusion

By improving our workflows and adopting a consistent representation of network-like lists, we are able to add the lists, check the data for accuracy, and create network graphics conveniently.

Future tasks:

- Adding the edge lists for all CLICS versions
- Testing the predictions about semantic change made in Urban (2011)
- Comparing the network data with psycholinguistic measures in NoRaRe