

Designing a derivational resource for non-concatenative Morphology: the Hebrewnet database



shutterstock.com - 1016524270

Lior Laks

Bar-Ilan University

Israel

Fiammetta Namer

Université de Lorraine

France



shutterstock.com - 1011792952



DeriMo 2021, Nancy
Thursday, 9 september 2021



Chateaubriand
Fellowship Program

Outline

- Hebrew Morphology
- Overview of the Hebrewnet Database
- Inter-family (and inter-paradigm) relations
- Case study: doublet formation

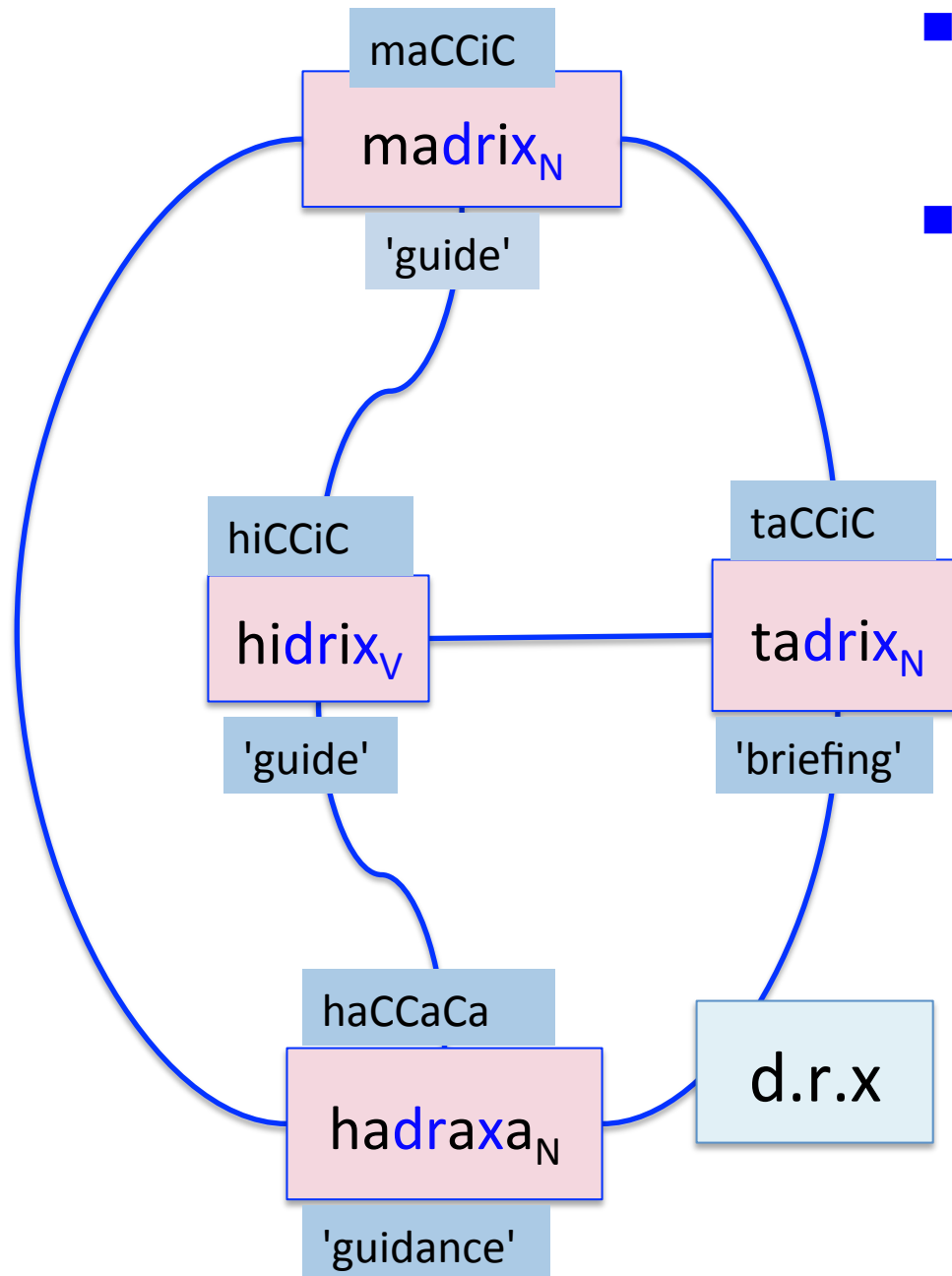
Non-concatenative morphology

- Semitic word formation relies highly on non-concatenative/ non-linear morphology
- Formation via **root and pattern** (Berman 1978, Bolozky 1978, Schwarzwald 1981, Ravid 1990, Aronoff 1994).

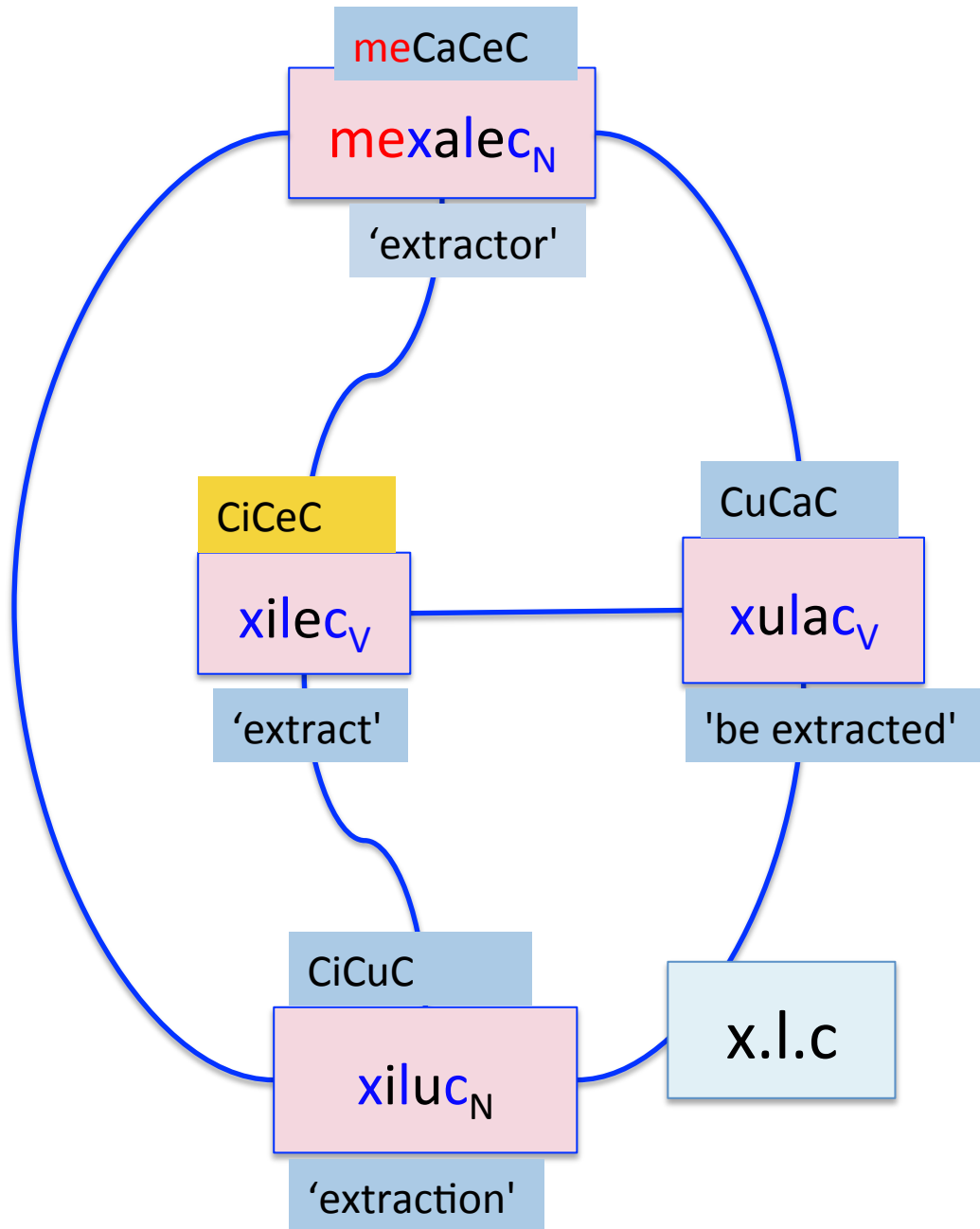
Non-concatenative morphology

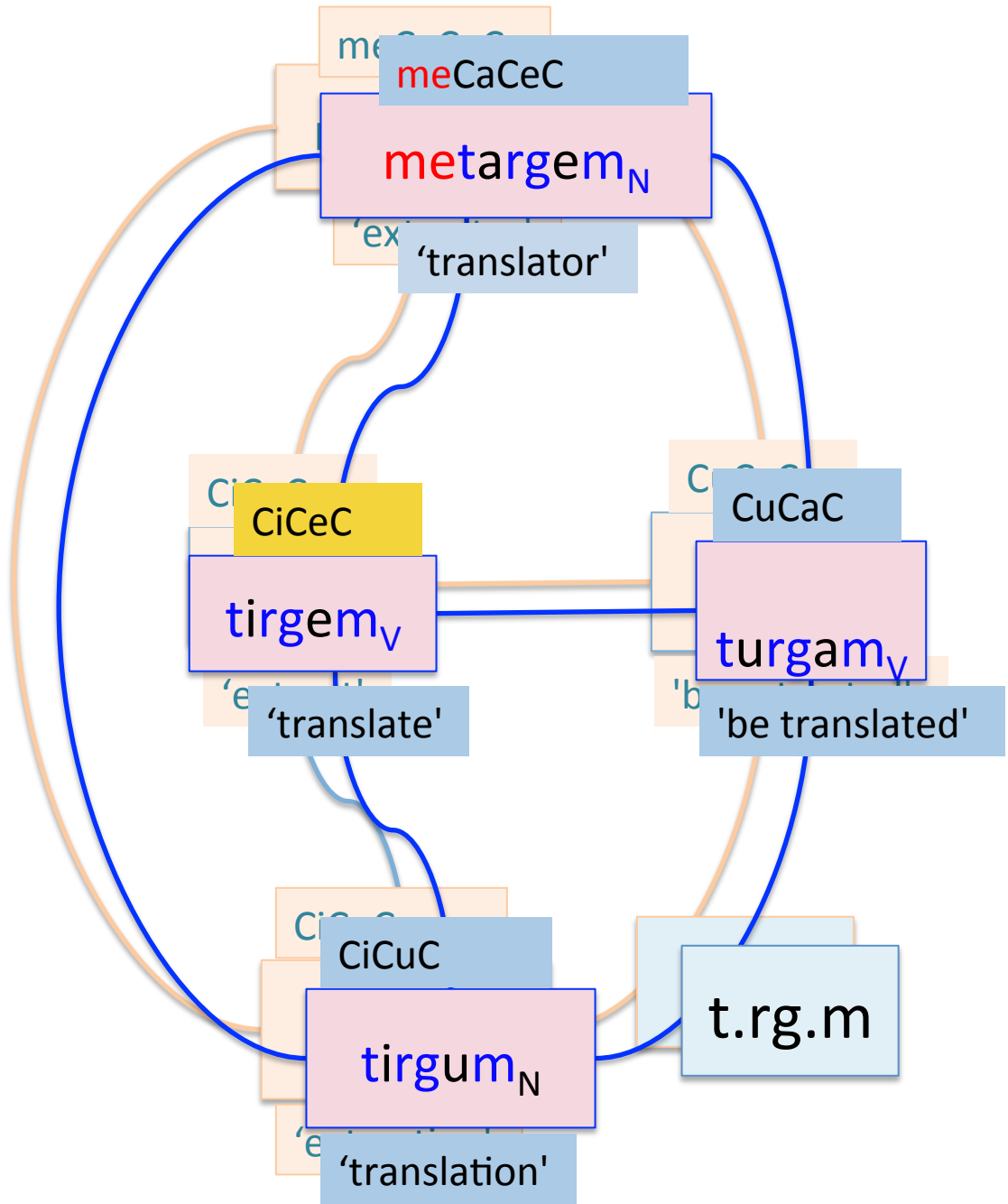
- Patterns consist of:
 - prosodic structure
 - vocalic melody
 - possible affixes

(Berman 1978, Bolozky 1978, Schwarzwald 1981, Ravid 1990, Aronoff 1994, Bat-El 1994, 2017, among others)



- Family members share the same root
- A relation connects two words with the same root





Non-concatenative morphology

- Hebrew verbs are formed only in patterns
- Nouns and adjectives can be formed in patterns as well as by other strategies.
- In this study we focus on non-concatenative formation.

The Hebrewnette DataBase

- Prototype developed following Demonette's guidelines
 - Each entry describes a derivational relationship between two lexemes.
 - Entries sharing similar properties form series.
 - Derivational families are graphs of pairs of words connected by derivational relations.
 - Derivational paradigms are graphs of pairs of word series.

The Hebrewnet Database

- Specific Features for Hebrew :
- On lexemes
 - (Root, Syllabic structures)
- On relations
 - (Formal orientation vs. Semantic orientation)

Description of šavar_v / šavir_A

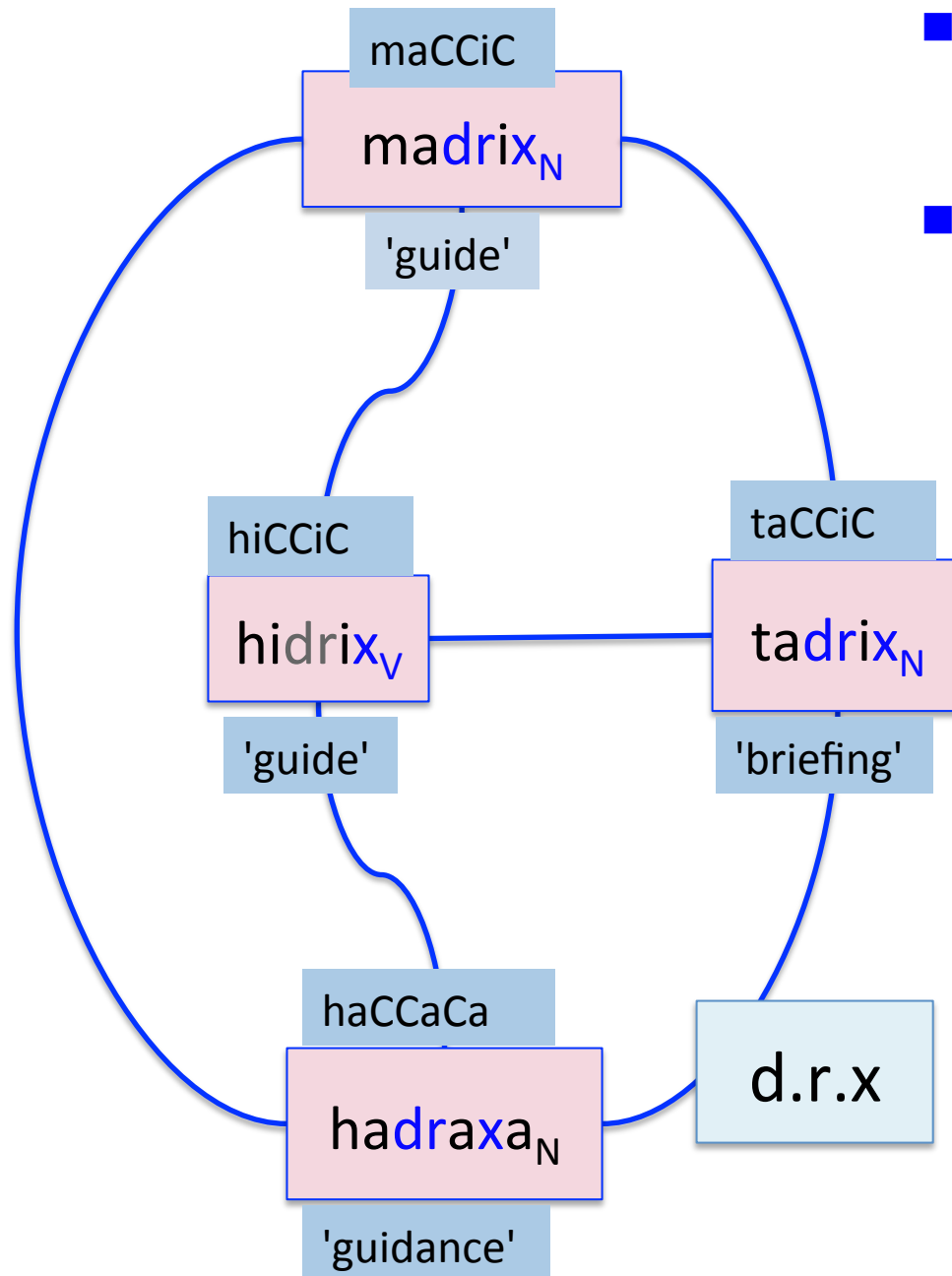
	Word1	Word2
Graphical form	שבר	שביר
Transcription	šavar	šavir
English Gloss	break	breakable
Part of speech	v	a
Pattern	CaCaC	CaCiC
Root	š.v.r	š.v.r
Morph. representation	aa	ai
Semantic type	dyn caus	prop
Argument Structure	(proto-)agent, patient	patient

Description of the relation between šavar_V and šavir_A

	šavar--šavir
Relation between patterns	CaCaC/CaCiC
Structural Complexity	simple
Structural Orientation	--
W1/W2 Phon alternation	--
Relation between roots	=
Semantic orientation	W1 → W2
Relation btw arg. struct.	XY / Y
W1/W2 cross-definition	" (Y) is šavir if (X) can šavar (Y)"

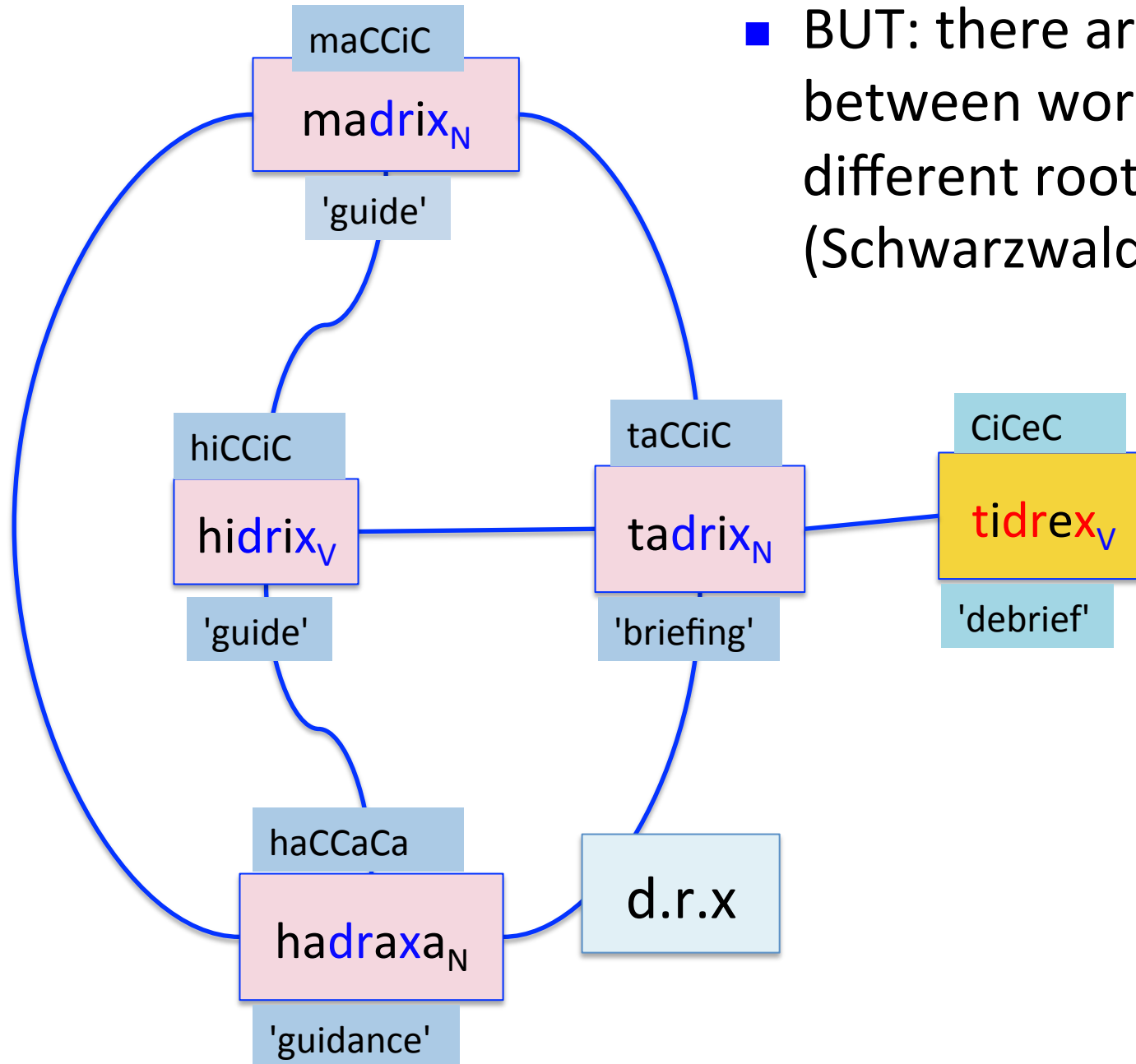
Roots and inter-family relations

- A feature describes the variation between the roots of the two related words.

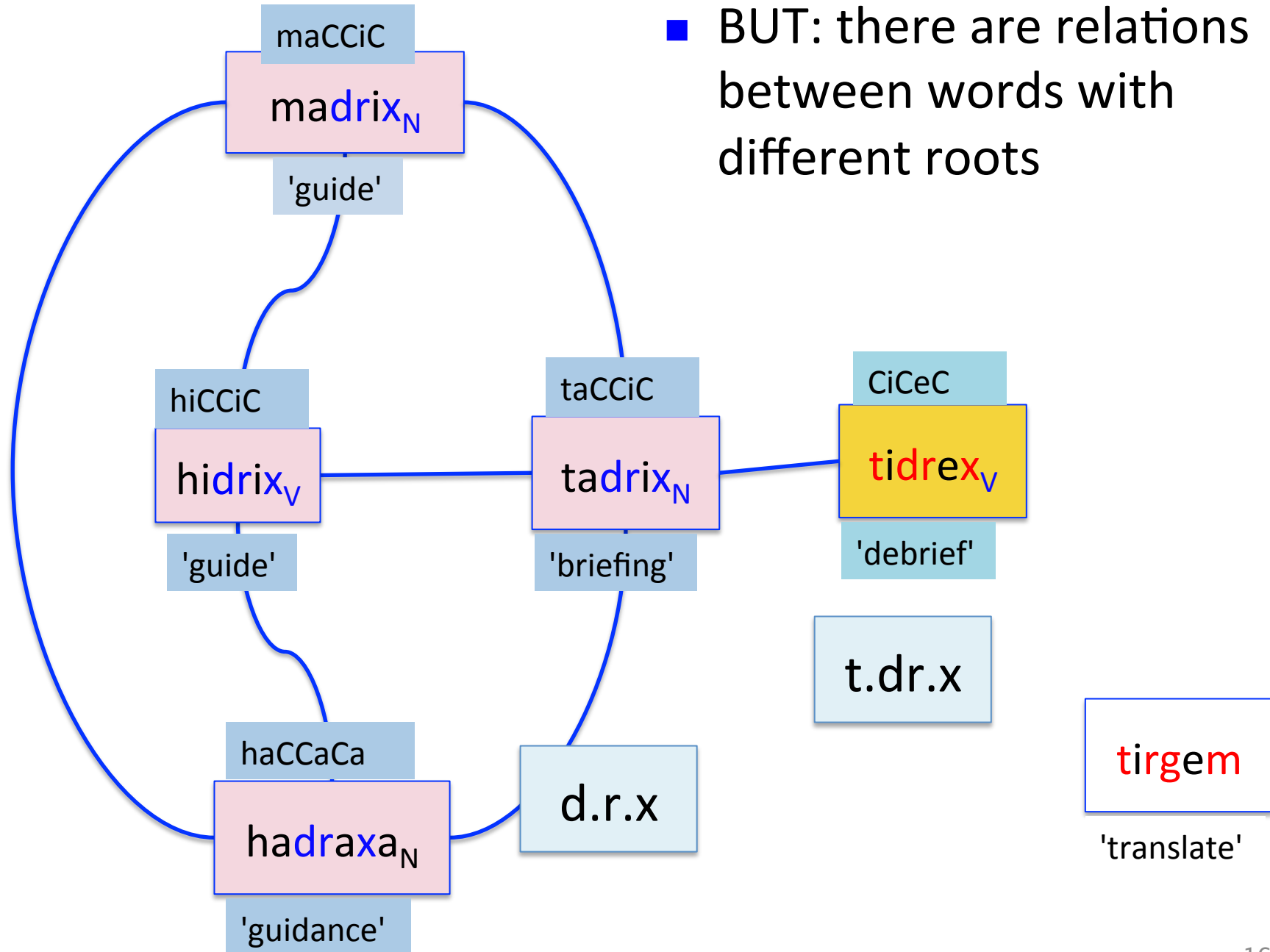


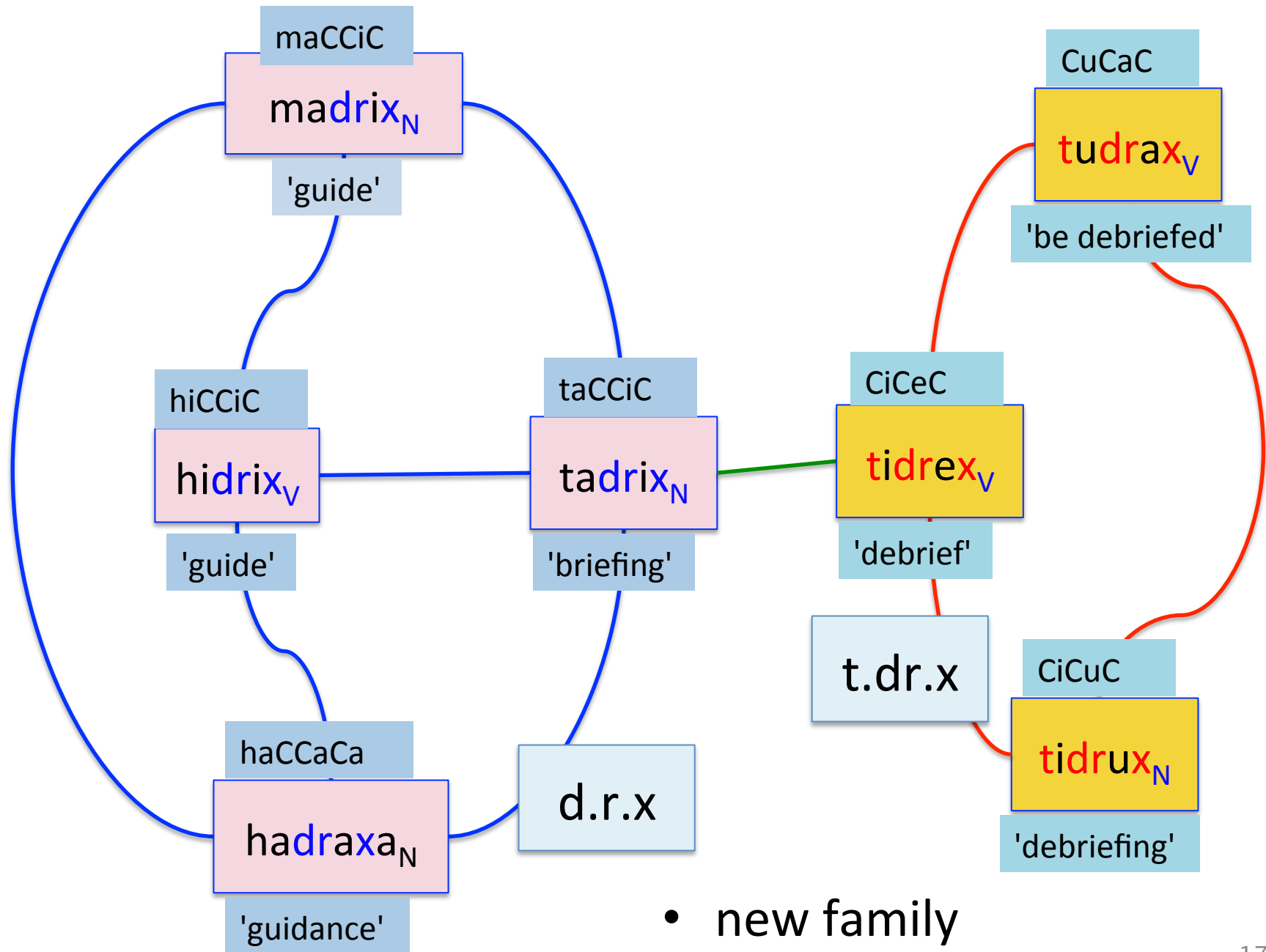
- Family members share the same root
- A relation connects two words with the same root

- BUT: there are relations between words with different roots (Schwarzwald 2018)



- BUT: there are relations between words with different roots





- new family

Case study: maCCuC doublet formation

- Some adjectives have doublets in the maCCuC pattern (Bolozky 1999, 2003, Laks, in press).
- This pattern is associated with negative meaning.

– lavašti jins **maxrid**

‘I wore an awful pair of jeans’

<https://bike.co.il/?p=2239>

– hi xorešet al oto jins **maxrud**

‘she wears the same awful jeans’

<http://tmi.maariv.co.il/style/Article-609396>

Case study: maCCuC doublet formation

- eyze **miskun**, ma hu asa la

“what a poor (guy), what did he do to her?”

https://www.tiktok.com/@einabl_253/video/6948081577649818881

- eyze **maskun**, kol paam ani yocet alexa

‘what a poor (guy), I lash out at you every time’

<https://www.inn.co.il/Forum/Forum.aspx/t851240>

Likelihood of maCCuC adjective doublets formation

W1	W2	Str. W1	Str. W2	String Operations	Distance
Existing maCCuC formations					
maxrid	maxrud	ma 0i	ma 0u	i > u	1
‘awful’					
muznax	maznux	mu 0a	ma 0u	u > a a > u	2
‘neglected’					

W1	W2	Str. W1	Str. W2	String Operations	Distance
Unfrequent or unattested maCCuC formations					
metoraf	matruf	me oa	ma 0u	e > a, a > u o > 0	6
'crazy'					
metunaf	*matnuf	me ua	ma 0u	e > a, a > u u > 0	6
'filthy'					
satum	*mastum	au	ma 0u	prefix ma- a > 0	6
'thickheaded'					

Current coverage

- The current coverage provides samples of different kinds of word formation processes, with different degrees of productivity.
 - contains relations among lexemes of 29 families of various size (between 3 and 16 members)
- Semi-automatically complemented with the families of 10 regular CiCeC verbs
- 250 entries, each described by 37 features.

Conclusions

- The design of Démonette's annotation system makes its features suitable for capturing both:
 - morphological relations
 - semantic relationsbetween Hebrew words, regardless of the type of morphology.

Conclusions

- Other morphological tools and resources for Hebrew and other Semitic languages (*e.g.* Wintner 2004, Itai & Wintner 2008, Singh & Habash 2012, Nir et al. 2013, Klimek et al. 2016) rely mostly on the consonantal root as and not on paradigmatic relations between words.

- Word-based approach

(Aronoff 1976, Blevins 2006)

- The central role of paradigms in derivational relations

(Bauer 1997, Pounder 2000, Beecher 2004, Booij & Lieber 2004, Štekauer 2014, Bonami & Strnadová 2019, Blevins 2016).

THANK YOU!

This work benefited from the support of the project DEMONEXT ANR-17-CE23-0005 of the French National Research Agency (ANR) and from the Chateaubriand Fellowship Program of the French Embassy in Israel.