



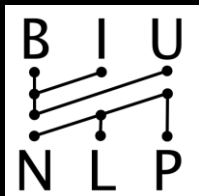
Hebrew Psychological Lexicons

לקסיקונים פסיכולוגיים בעברית

Natalie Shapira

Dana Atzil-Slonim, Daniel Juravski, Moran Baruch, Adar Paz, Dana Stolorowicz-Melman, Tal Alfi-Yogev, Roy Azoulay, Adi Singer, Maayan Revivo, Chen Dahbash, Limor Dayan, Tamar Naim, Lidar Gez, Boaz Yanai, Adva Maman, Adam Nadaf, Elinor Sarfati, Amna Baloum, Tal Naor, Ephraim Mosenkis, Matan Kenigsbuch, Badreya Sarsour, Yarden Elias, Liat Braun, Moria Rubin, Jany Gelfand Morgenshteyn, Noa Bergwerk, Noam Yosef, Sivan Peled, Coral Avigdor, Rahav Obercyger, Rachel Mann, Tomer Alper, Inbal Beka, Ori Shapira, Yoav Goldberg

June 2021



Psychotherapy
Research Lab



Computer Science and Psychology Departments

Supervisors →



Prof. Yoav Goldberg



Dr. Dana Atzil-Slonim

Leading →



Natalie Shapira

Interning
Therapists and →
Psychotherapy



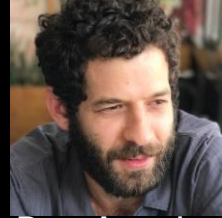
Adar Paz



Dana Stolowicz-Melman



Tal Alfi-Yogev



Roy Azoulay

Natural Language
Processing and →
Machine Learning



Ori Shapira



Moran Baruch



Daniel Juravski



Inbal Beka

Psychology
Research Practicum



Adi Singer Ruskin



Maayan Revivo



Chen Dahbash



Limor Dayan



Tamar Naim



Lidar Gez



Boaz Yanai



Adva Maman



Elinor Sarfati



Amna Baloum



Tal Naor



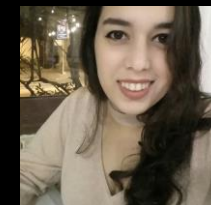
Ephraim Mosenkis



Matan Kenigsbuch



Badreya Sarsour



Yarden Elias



Liat Braun



Moria Rubin



Jany Gelfand



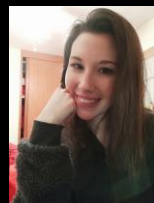
Morgenshteyn



Noa Bergwerk



Noam Yosef



Sivan Peled



Coral Avigdor



Rahav Obercyger



Rachel Mann



Tomer Alper



Adam Nadaf

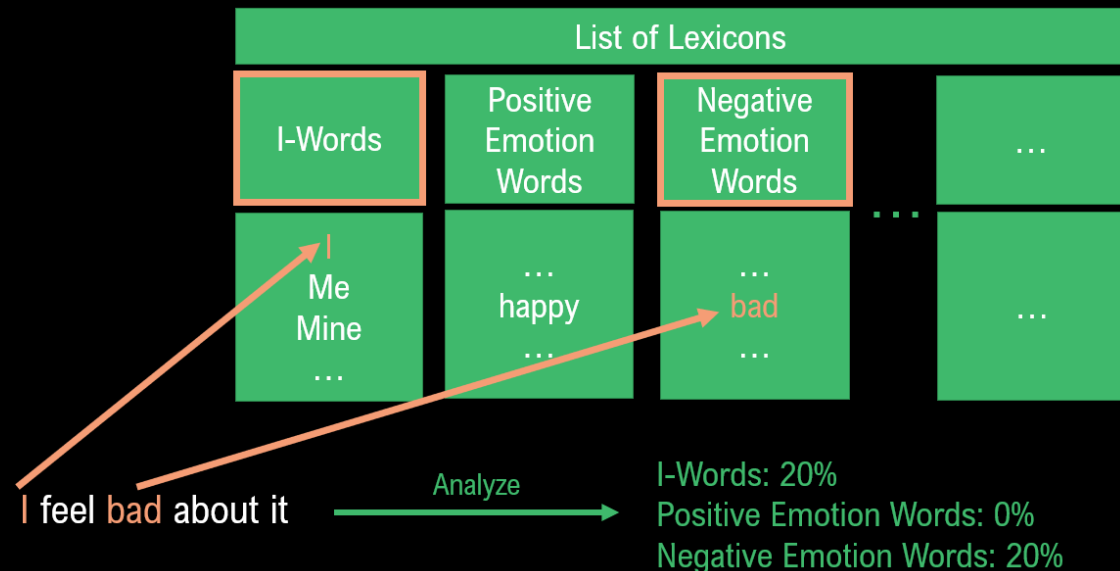
Outline

- Motivation
- Introduce large set of Hebrew lexicons
- Challenges in creating and validating lexicons
- The Emotional Variety lexicon construction process

Outline

- Motivation
- Introduce large set of Hebrew lexicons
- Challenges in creating and validating lexicons
- The Emotional Variety lexicon construction process

Why Do We Need Lexicons?



- Scarce data
 - Few samples are available in clinical trials
 - Confidentiality limits sharing of data
- Serve as clinical markers
- Interpretation of results
- Easy to use and improve performance

data-hungry models are not practical in such cases

Outline

- Motivation
- Introduce large set of Hebrew lexicons
- Challenges in creating and validating lexicons
- The Emotional Variety lexicon construction process

Collections

Lexicons & Word-Lists

Collection Name (# Lexicons or Lists, # Words)

Valence (2, 200)

Emotional Variety (42, 7313)

Paralinguistics (11, 154)

Depressive Characteristics (14, 194)

Well-Being (2, 40)

Conversation Topics (200, 4000)

Hebrew LIWC (under construction)

Extended Emotional Variety (under construction)

Construction Methods

Expert Knowledge Based Lexicons

Data-Driven Lists

- Supervised
- Unsupervised

Expert Knowledge + Automatic Methods

- Translation
- Expansion

Validity & Reliability

Coverage

Domain expert verification

Initial research use case

- Outcome Rating Scale (ORS; Miller et al., 2003)
- Profile of Mood States (POMS; McNair, 1992)
- Post-Session Questionnaire (PSQ; Muran et al., 2004)

Outline

- Motivation
- Introduce large set of Hebrew lexicons
- Challenges in creating and validating lexicons
- The Emotional Variety lexicon construction process

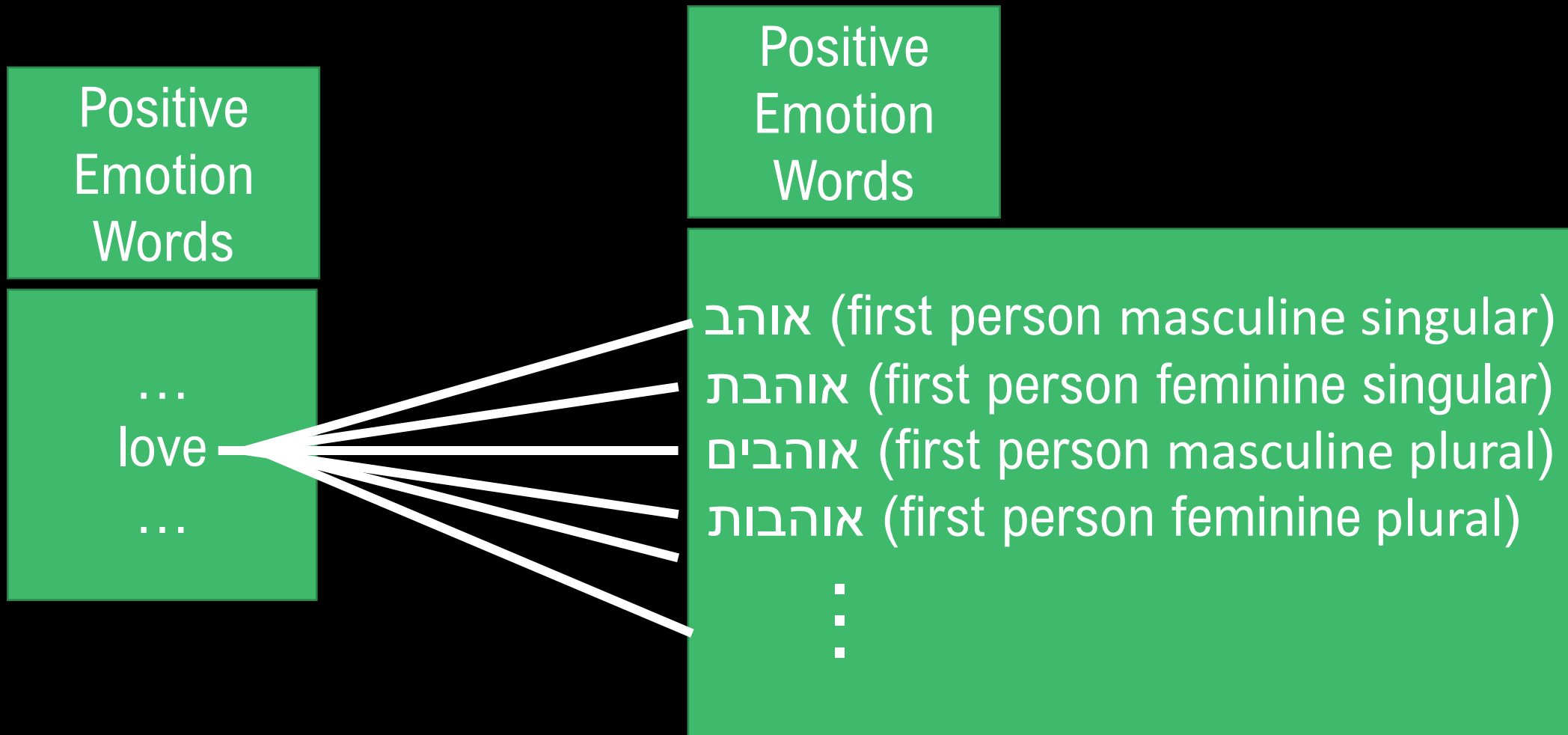
The Challenge - Hebrew

- No LIWC version in Hebrew
- Languages behave differently
- Morphology makes things harder



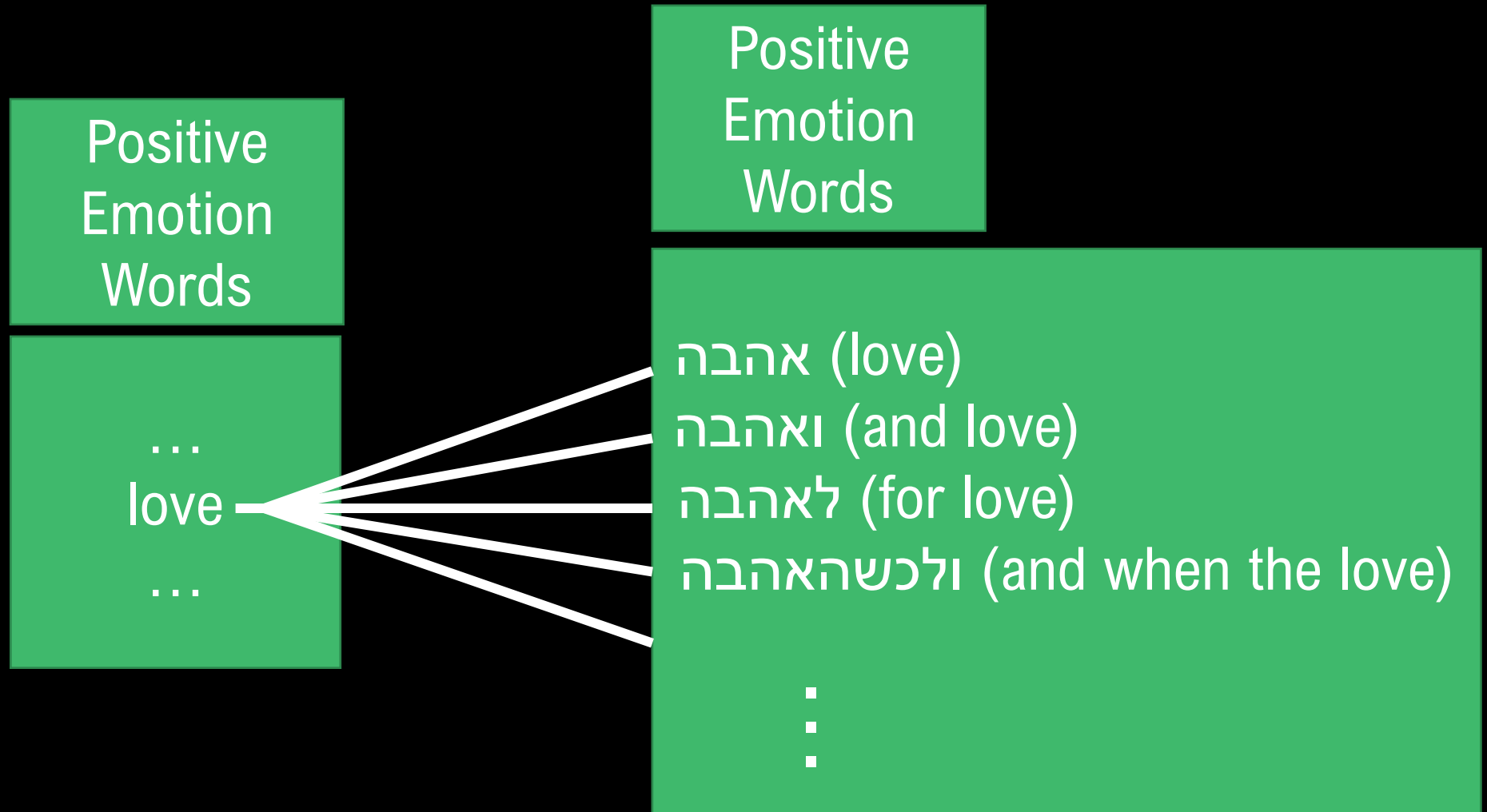
Morphology

(1) Verb inflections

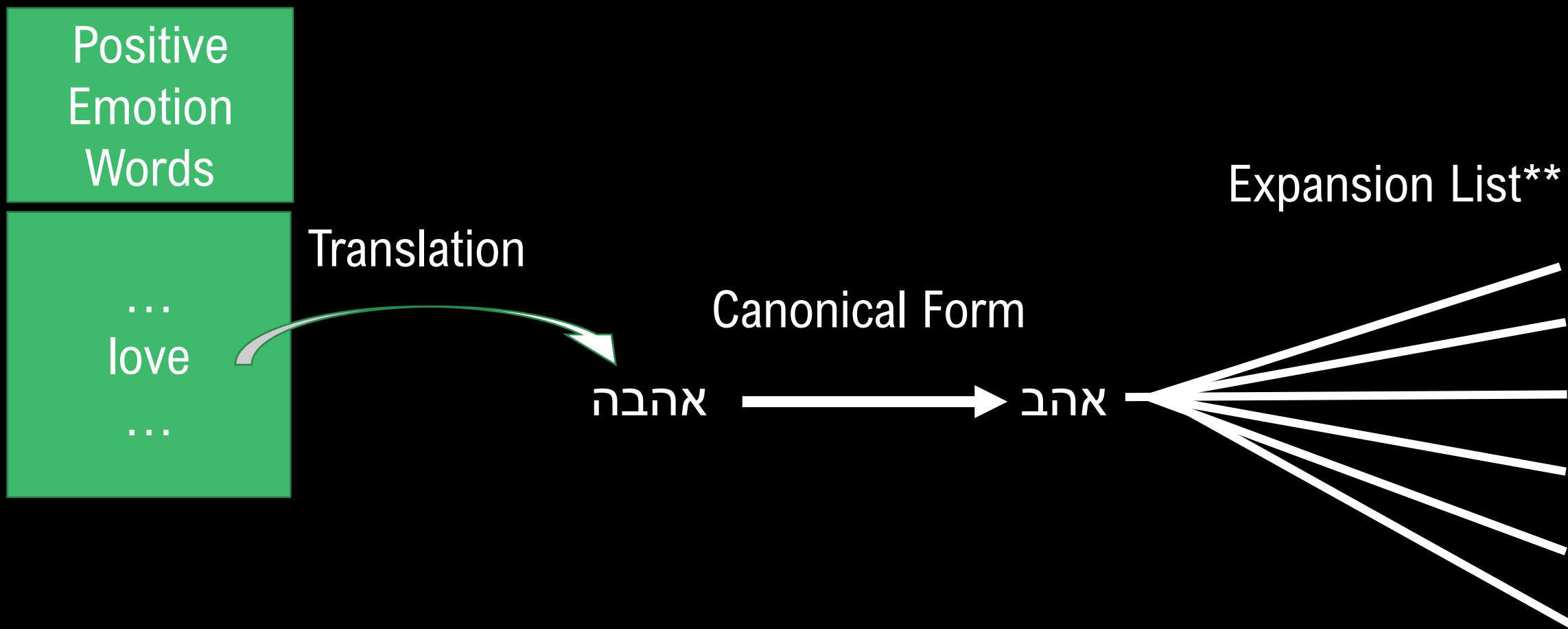


Morphology

(2) Clitics / Morphemes



Solution



** Such as the MILA lexicon (Itai and Wintner, 2008) or the BGU-version of the lexicon, which is bundled with the YAP Hebrew parser (More and Tsarfaty, 2016) as the file bgulex.utf8.hr.

Solution

Positive
Emotion
Words

...
Cool
...

Translation

Canonical Form

Expansion List

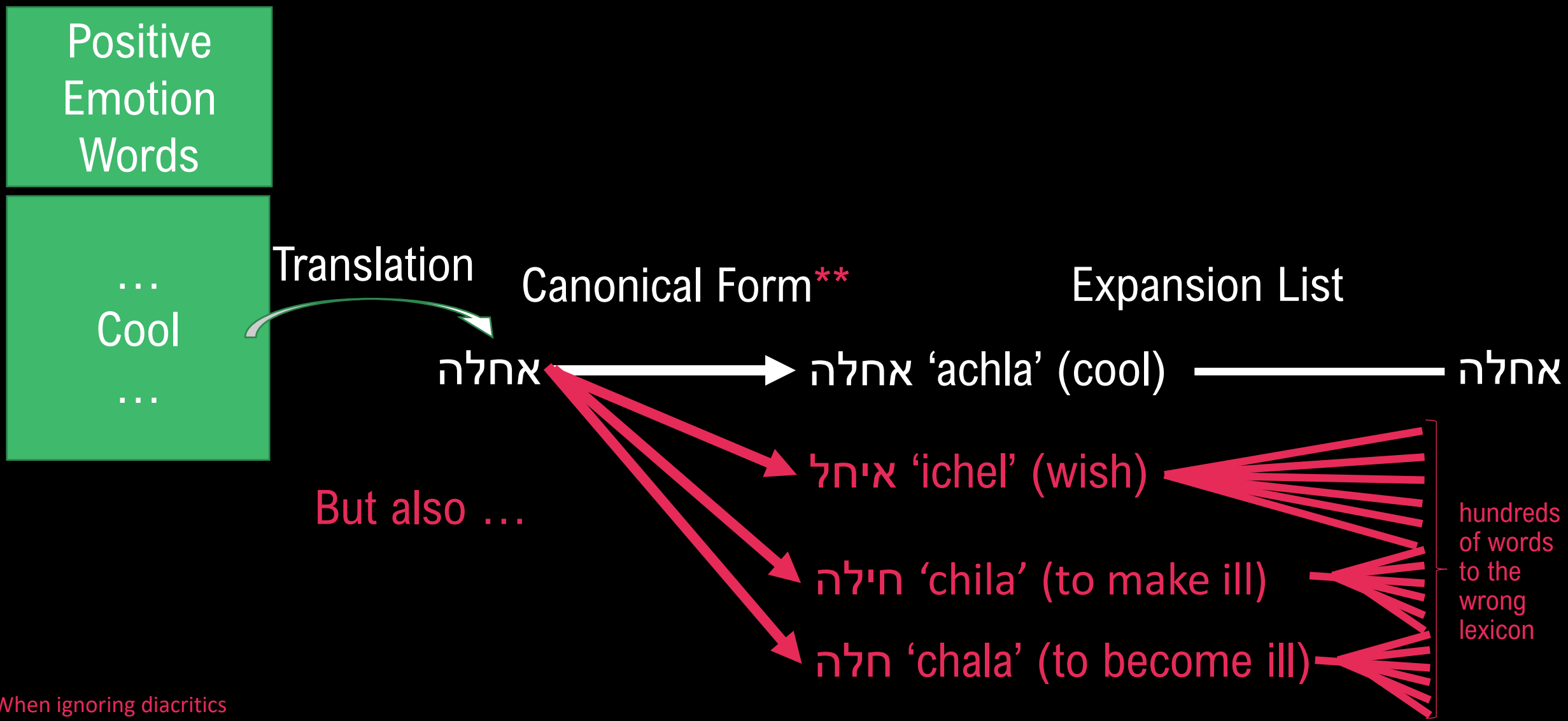
אחלה

אחלה 'achla' (cool)

אחלה



Solution...?

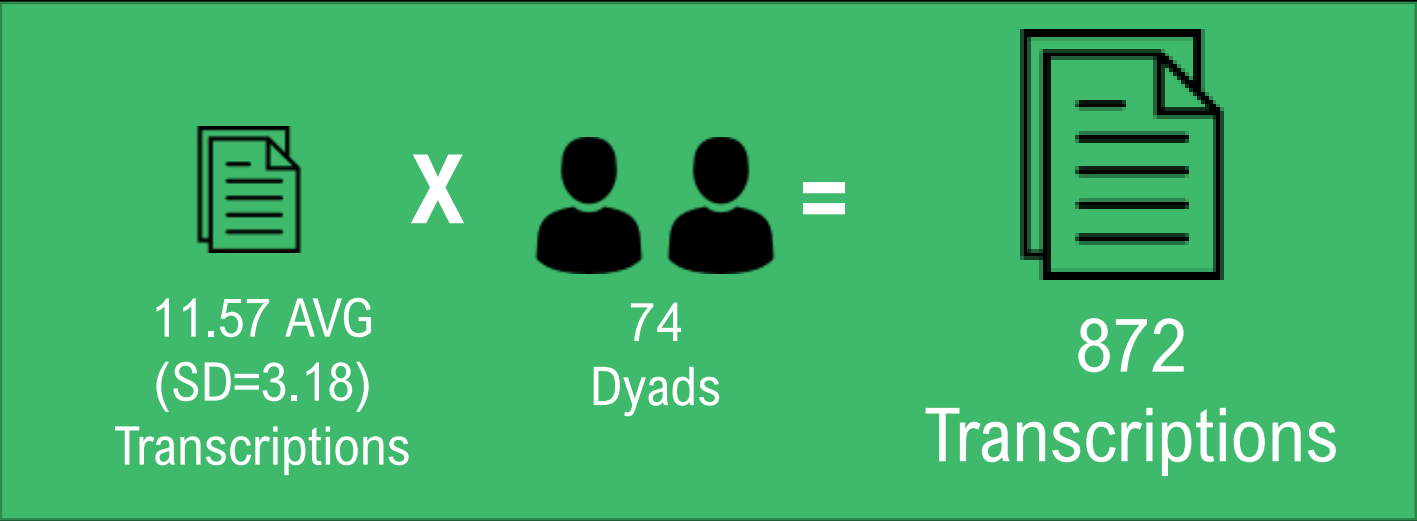


** When ignoring diacritics

Outline

- Motivation
- Introduce large set of Hebrew lexicons
- Challenges in creating and validating lexicons
- The Emotional Variety lexicon construction process

Base Dataset

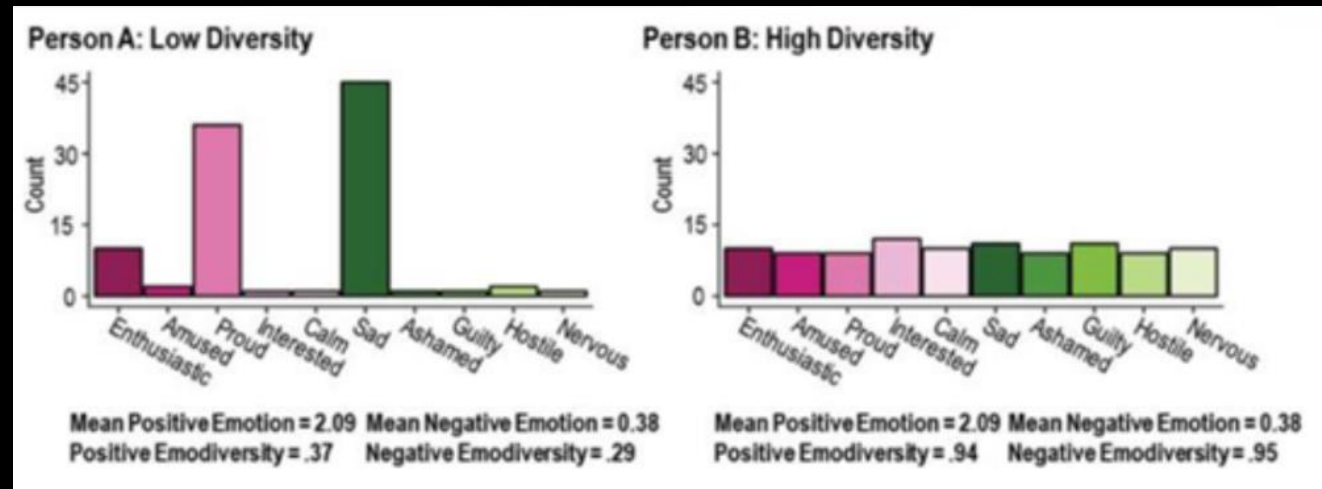


Statistics	Total	Per Patient	Per Session
Talk Turns	~150K	~2K	~200
Client Tokens	~4 Million	~50K	~4.5K
All Tokens	~5 Million	~70K	~6K

Client content is 80%
of the session

Emotional Variety

Motivation:



Ong et al. (2018)

Emotional Variety

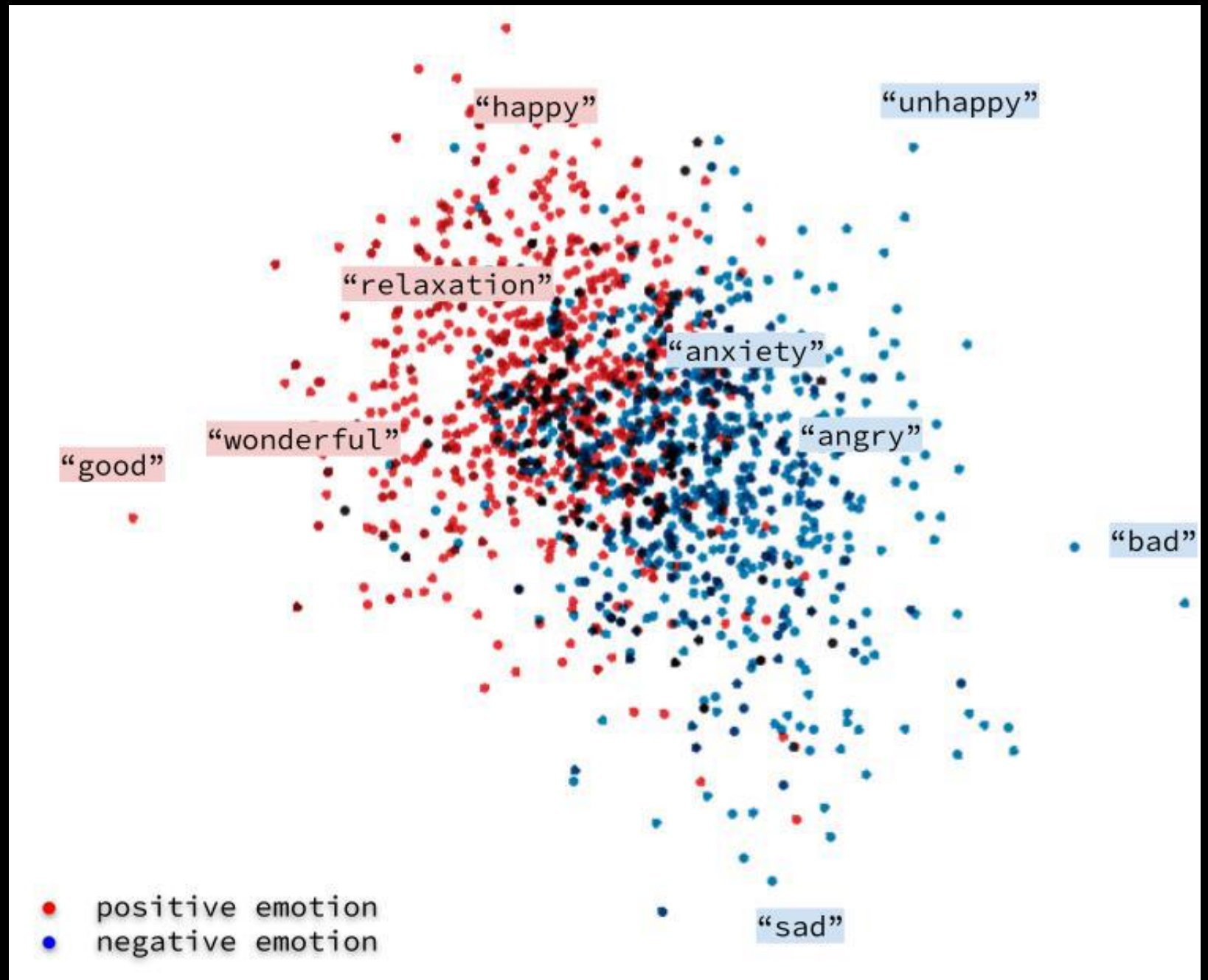
- Enthusiastic
- Amused
- Proud
- Interested
- Calm
- Sad
- Ashamed
- Guilty
- Hostile
- Nervous
- Anger
- Contentment
- Anxiety
- Vigor
- Joy
- Disgust
- Surprise
- Trust
- Anticipation
- Confusion
- Fatigue

Emotional Variety

Complementing-emotion

- | | | | | | |
|----------------|--------------------|---------------|-------------------|----------------|--------------------|
| • Enthusiastic | • Not Enthusiastic | • Guilty | • Not Guilty | • Joy | • Not Joy |
| • Amused | • Not Amused | • Hostile | • Not Hostile | • Disgust | • Not Disgust |
| • Proud | • Not Proud | • Nervous | • Not Nervous | • Surprise | • Not Surprise |
| • Interested | • Not Interested | • Anger | • Not Anger | • Trust | • Not Trust |
| • Calm | • Not Calm | • Contentment | • Not Contentment | • Anticipation | • Not Anticipation |
| • Sad | • Not Sad | • Anxiety | • Not Anxiety | • Confusion | • Not Confusion |
| • Ashamed | • Not Ashamed | • Vigor | • Not Vigor | • Fatigue | • Not Fatigue |

2D-Projection of emotion word embeddings



Emotional Variety

- Freely-suggested words by 19 advanced undergraduate psychology students
- 5000 most frequent words, covering 90% of all tokens in all transcripts



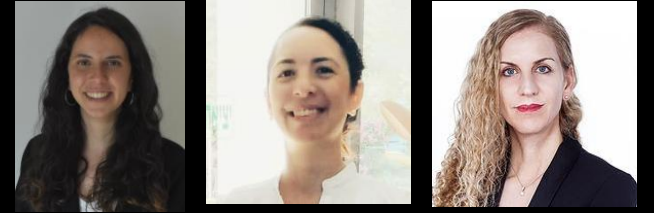
Merged by majority of three judges

Emotional Variety

- Freely-suggested words by 19 advanced undergraduate psychology students
- 5000 most frequent words, covering 90% of all tokens in all transcripts
- Automatic seed expansion (under construction)



Merged by majority of three judges



Emotional Variety

- Freely-suggested words by 19 advanced undergraduate psychology students
- 5000 most frequent words, covering 90% of all tokens in all transcripts
- Automatic seed expansion (under construction)

Correlations with
Profile of Mood States
(POMS; McNair, 1992)
& research use case
– work in progress

Merged by majority of three judges

Conclusion

- Lexicons are important for psychotherapy research, they enable interpretable prediction abilities and serve as clinical markers
- Creating and validating lexicons is challenging because languages behave differently, and morphology makes things harder
- Introduce with Hebrew Psychological Lexicons
 - 8 collections of lexicons and word lists
 - The Emotional Variety lexicon construction process

We hope that this work will inspire the creation of additional lexicons in Hebrew and in other languages and will enable high quality and cross lingual research

Thank You!

8 Lexicons and Word List Collections

	Expert Knowledge Based Lexicons				Data-Driven Lists		Expert Knowledge + Automatic Methods	
					Supervised	Unsupervised	Translation	Expansion
Collection name	Valence (Positive-Negative)	Emotional Variety	Paralinguistics	Depressive Characteristics	Well-Being	Conversation Topics	Hebrew LIWC	Extended Emotional Variety
Number of lexicons/lists	2	42	11	14	2	200	~40 out of 125	44
Total number of words	200	7313	154	194	40	4000	under construction	under construction
Coverage	2000 most frequent word types in dataset	5000 most frequent word types in dataset	31,067 tokens 1022 word types	several hundred most important word types	139 non-clinical sessions 38 clinical sessions	the whole dataset ~5 million tokens	-	-
Verified by at least three domain experts	yes	yes	yes	yes	-	-	yes	under construction
Initial research use case	yes	work in progress	yes	yes	-	yes data-dependent	-	-
Freely available	yes	yes	yes	yes	yes	yes	internal use only	will be released

Questions – nd1234@gmail.com