Compounding patterns in Hebrew writing: A developmental study

Avital Braun, Dorit Ravid & Elitzur Dattner

Hebrew Noun–Noun Compounds

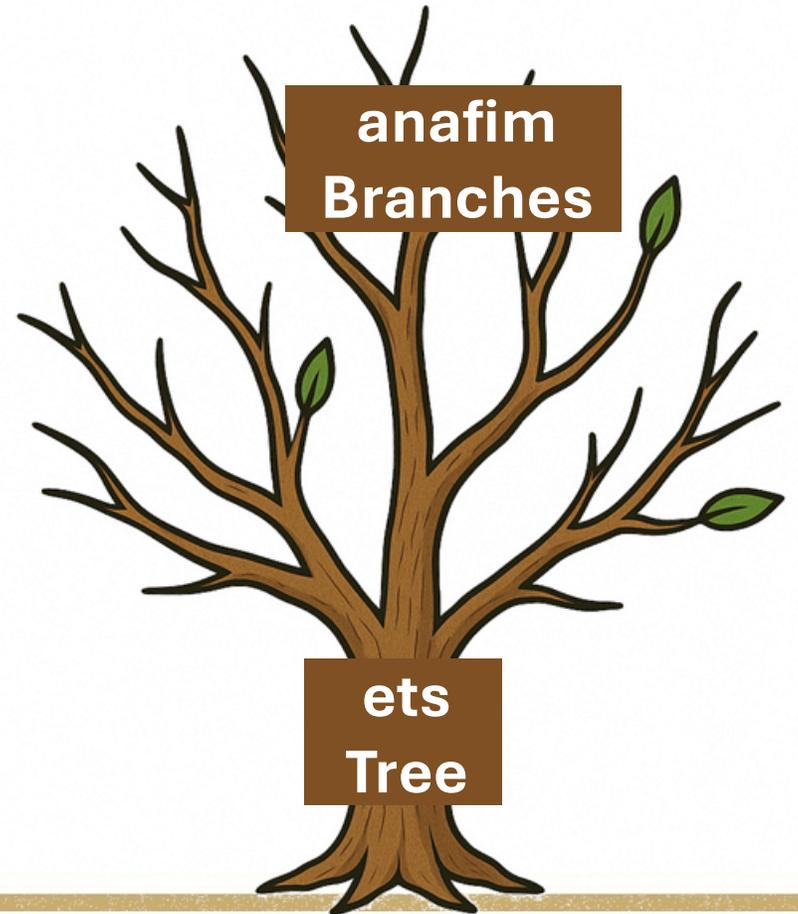## 1. Bound compound

2. Free compound

3. Double compound

Hebrew Noun–Noun Compounds

1. **Bound compound**

2. **Free compound**

3. **Double compound**

anafim
Branches

ets
Tree

Hebrew Noun–Noun Compounds

1. **Bound compound**

anf<mark>ey</mark>^ha-ets

branch.PL CONSTR DEF-tree

'the tree branches'

anafim
Branches

ets
Tree

Hebrew Noun–Noun Compounds

1. **Bound compound**

anfey^ha-ets

branch.PL CONSTR DEF-tree

'the tree branches'

2. **Free compound**

ha-anafim shel ha-ets

DEF-branch.PL of DEF-tree

'the branches of the tree'

anafim
Branches

ets
Tree

Hebrew Noun–Noun Compounds

1. **Bound compound**

anfey^ha-ets

branch.PL CONSTR DEF-tree

'the tree branches'

2. **Free compound**

ha-anafim shel ha-ets

DEF-branch.PL of DEF-tree

'the branches of the tree'

3. **Double compound**

anafav^shel ha-ets

3MSg-his of the-tree

the branches of the tree

6
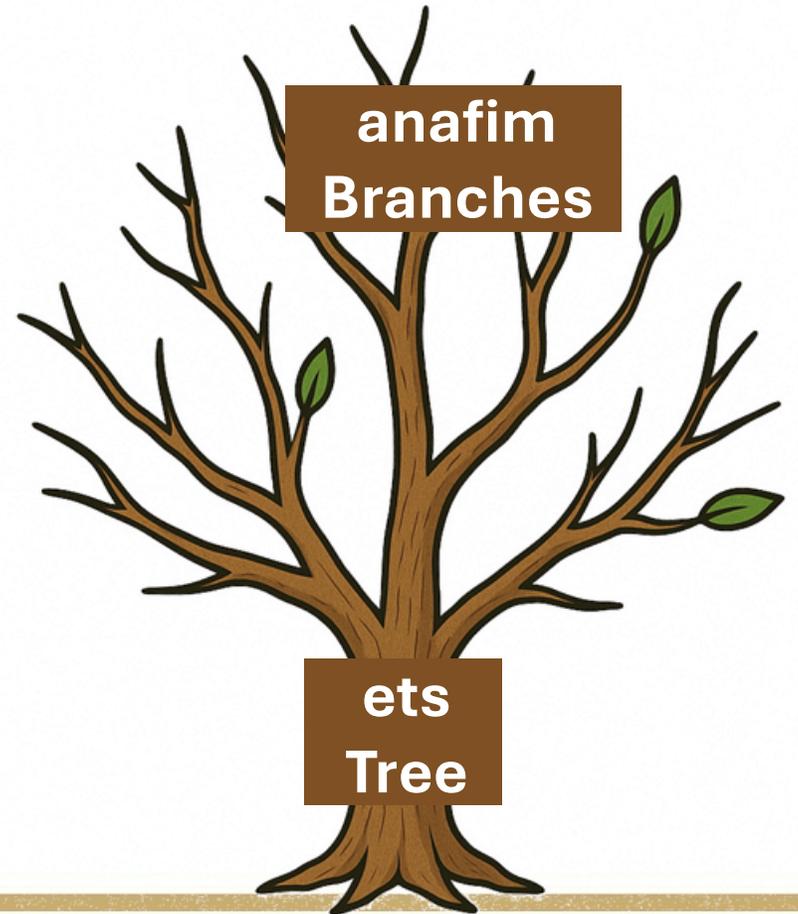
**anafim
Branches**

**ets
Tree**

Characteristics of Hebrew Compounds

A highly frequent structure:

- About 10% of all words in expository texts Berman & Nir-Sagiv, 2004

- About 40% of the nouns within a single text   Ravid & Zilberbuch, 2001; Schlesinger & Ravid, 1998

- In our study, nearly every third clause contained a compound

Part of later language development

Characteristics of Hebrew Compounds

A critical interface between the lexicon, ==morphology==, and syntax

1. Bound compound

anf==ey==^ha-ets

branch.PL CONSTR DEF-tree

'tree branches'

2. Free compound
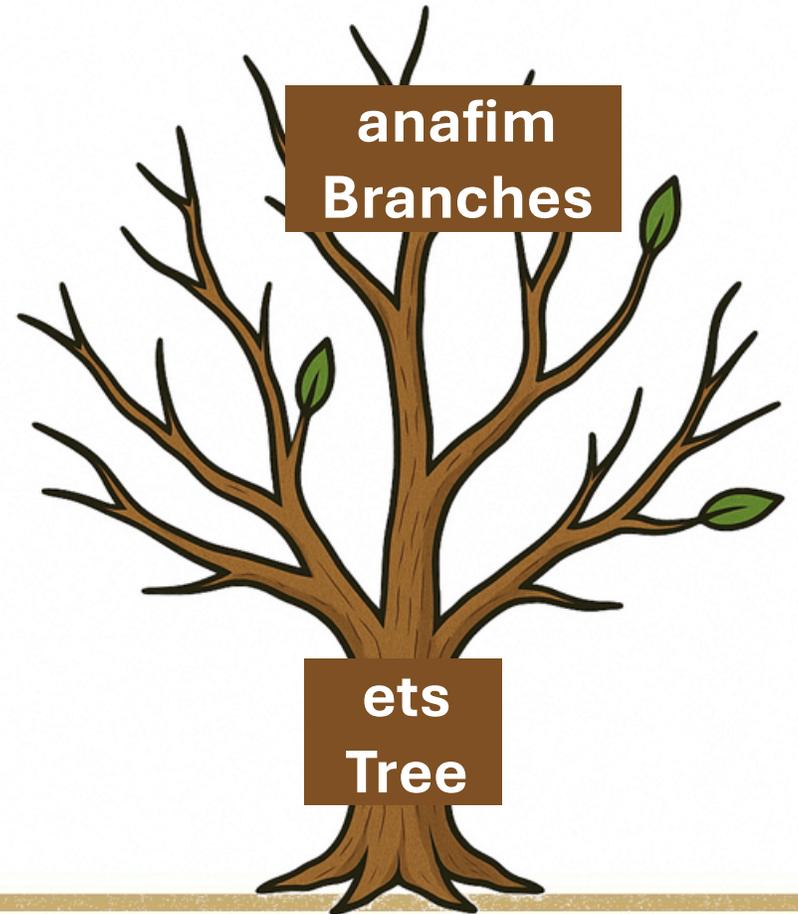
ha-anafim shel ha-ets

DEF-branch.PL of DEF-tree

'the branches of the tree'
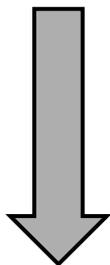
3. Double compound

anaf==av==^shel ha-ets

3MSg-his of the-tree

the branches of the tree

Berman, 1988; Shlezinger, 1991; Shlezinger, Ravid & Sharel, 1996; Ravid & Zilberbuch, 2003; Ravid & Bruckman-Braun, 2012; Ravid & Assulin-Tzabar, 2017
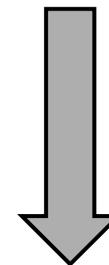
Characteristics of Hebrew Compounds

A critical junction between the lexicon, morphology, and syntax

The branches of the tree

הענפים של העץ

The wide and impressive branches of the ancient tree that I planted in my yard

הענפים הרחבים והמרשימים של העץ העתיק ששתלתי בחצר שלי

Berman, 1988; Shlezinger, 1991; Shlezinger, Ravid & Sharel, 1996; Ravid & Zilberbuch, 2003; Ravid & Bruckman-Braun, 2012; Ravid & Assulin-Tzabar, 2017

Characteristics of Hebrew Compounds

- **A highly frequent structure**
- **Part of later language development**
- **A critical junction between the lexicon, morphology, and syntax**
- **Theoretical frameworks underlying the study of Hebrew compounds:**
  - Usage-based approaches Arnon, 2015; Bybee, 2005; Elman, 2001, 2005; Wray, 2002
  - Construction Grammar framework Goldberg, 2003

- **key psycholinguistic aspects**

key psycholinguistic aspects

- **Cognitive Representation of Compounds: Holistic vs. Analytic?**

- **Psycholinguistic factors influencing stability and mental representation**
  - **Frequency & Productivity:** Compound frequency; Productivity of the first constituent; Morphological family size
  - **Semantic Factors:** Semantic transparency; Frequency of semantic relations between constituents; Acceptability of the compound
  - **Cognitive & Developmental Factors:** Age of acquisition (AoA); Imageability of the compound
  - **Formal Properties:** Length; Morpho-phonological alternations; Graphic cue

Bertram, Hyönä & Laine, 2000; Fiorentino & Poeppel, 2007; Frisson, Niswander-Klement & Pollatsek, 2008; Gagné & Spalding, 2009; Gagné, Spalding, Figueredo & Mullaly, 2009; Günther & Marelli, 2020; Hyönä & Olkoniemi, 2020; Hyönä, Bertram & Pollatsek, 2000; Libben, 1998; Libben, Gallant & Dressler, 2021; Niswander-Klement & Pollatsek, 2006; Pollatsek, Hyönä & Bertram, 2000; Rayner, Andrews & Miller, 2004; Wang & Xu, 2025

Hebrew Compounds

- A highly frequent structure

- A critical interface between the lexicon, morphology, and syntax

- Part of later language development

- Grounded in major theoretical frameworks of language acquisition

- Key psycholinguistic aspects

No previous study has examined:
- ✓ All three compound structures together
- ✓ In natural written monologic discourse
- ✓ Across two genres – narrative and informative
- ✓ Developmentally
- ✓ From a psycholinguistic perspective

The Present Study: Written Texts

Corpus

500 texts
39,708 words
8,438 clauses
2,565 compounds

2 Genres

Narrative texts
Informative texts

5 Age Groups

- Grade 4
- Grade 7
- Grade 11
- young adults (M=20)
- Adults (M=30)

Range of Analyses:

lexical, semantic, morphological, syntactic, and discourse analyses

The Present Study: A Range of Analyses of Compound Structure profiles

Morpho–Syntactic

- Structure of each constituent
- Inflections (obligatory & optional)
- Derived vs. root-compound
- Definiteness

Structural–Semantic

- Compound type
- Semantic relation between constituents
- Level of abstractness
- Level of expectancy *

Discourse Profile

- Reiteration (full / partial)
- Position in the text

Syntactic Profile

- Syntactic function (subject vs. other roles)
- Number of expansions per constituent

*Part of my doctoral dissertation, submitted to TAU last week

The Present Study: A Range of Analyses of Compound Structure profiles

Morpho–Syntactic
- Structure of each constituent
- Inflections (obligatory & optional)
- Derived vs. root-compound
- Definiteness

Structural–Semantic Compound type
- Semantic relation between constituents
- Level of abstractness
- Level of predictably *

Discourse Profile
- Reiteration (full / partial)
- Position in the text

Syntactic Profile
- Syntactic function)subject vs. other roles)
- Number of expansions per constituent

Predictability Level: Judgment Questionnaire

- Identification of all bound and double compounds in the corpus

- **Initial aim:** code each compound as Lexicalized / Collocational / Novel

- **"Crowd wisdom":** large-scale judgment questionnaires
  - Given the first constituent ____, how likely is ____ to appear next?
    ○ Very low   ○ Low   ○ Medium   ○ High   ○ Very high

  - Given the first constituent <u>anfei</u>, how likely is <u>ha-ets</u> to appear next?

Branches anafim

Tree ets

Predictability Level: Judgment Questionnaire

- **"crowd Wisdom":** large-scale judgment questionnaires
  - Given the first constituent _____, how likely is _____ to appear next?
 ○ Very low   ○ Low   ○ Medium   ○ High   ○ Very high

Data Collection:

- Approximately 2,000 questionnaires, completed mostly by SLPs

- Each questionnaire included 50 items, with about 100 responses per item on average
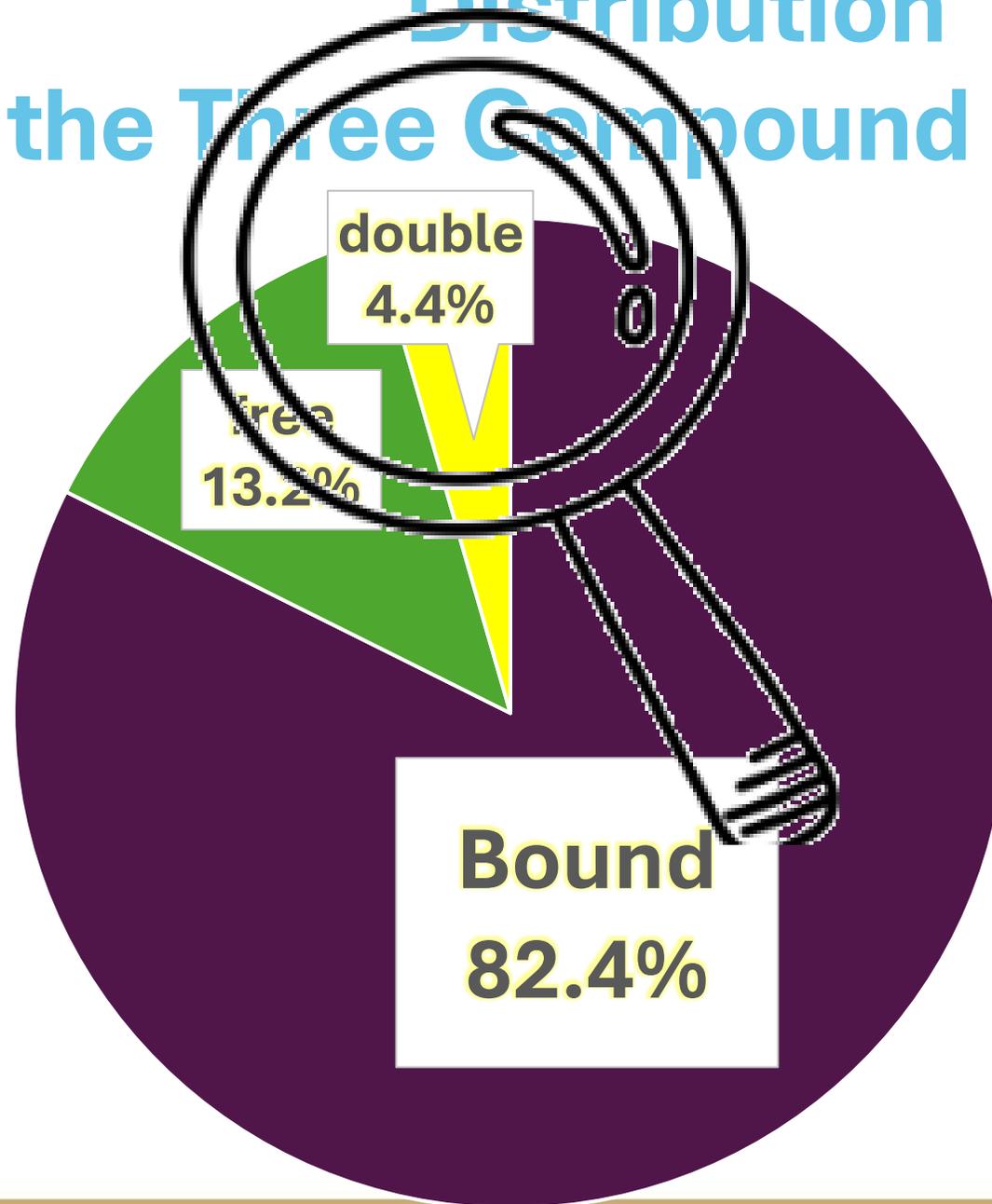
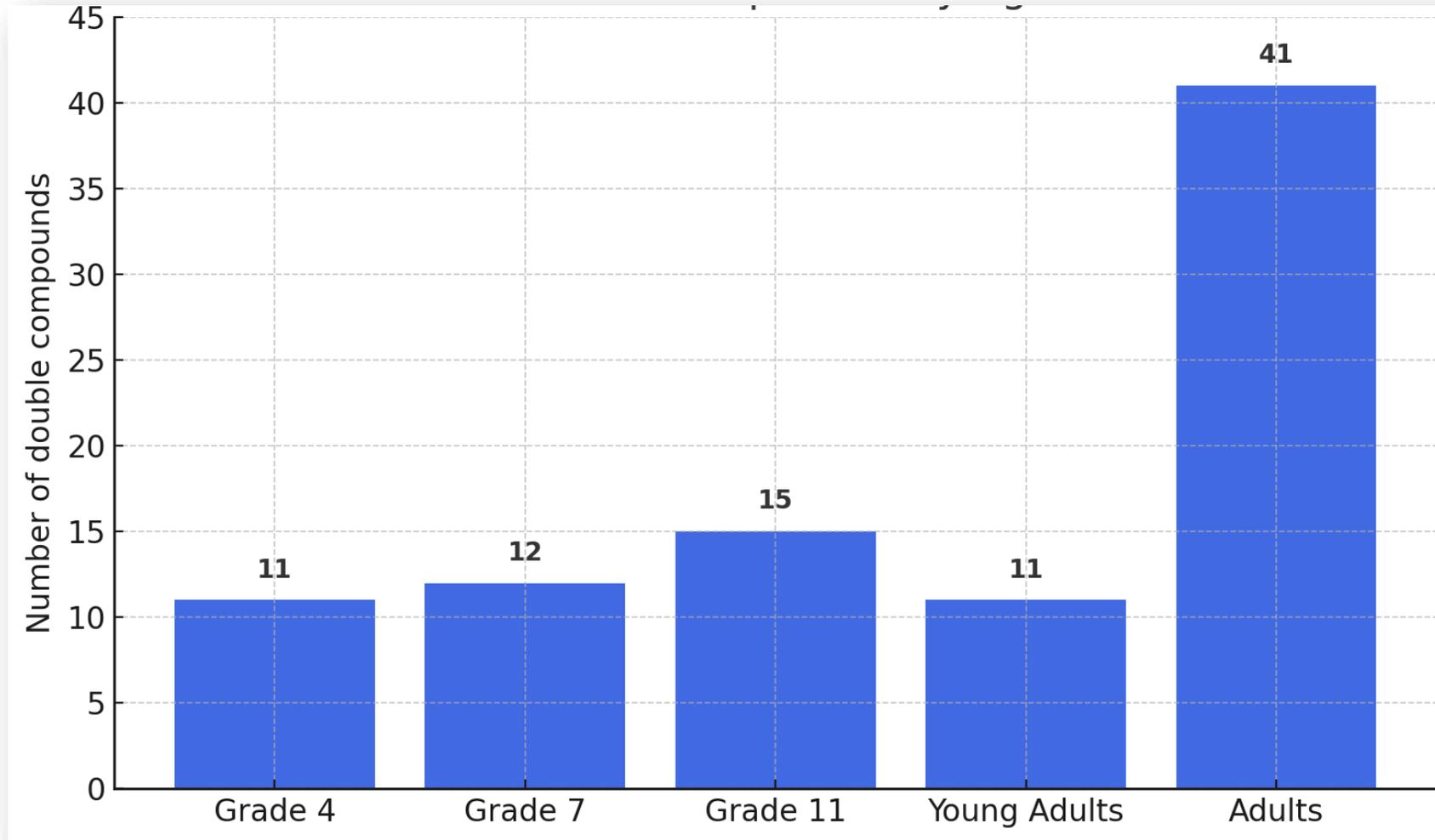# Results

Distribution
of the Three Compound Structures

✓ No significant differences by age/genre
- From Grade 4 through adulthood
- In both narrative and informative texts

✓ A foundational distribution pattern in Hebrew

double 4.4%

free 13.2%

Bound 82.4%
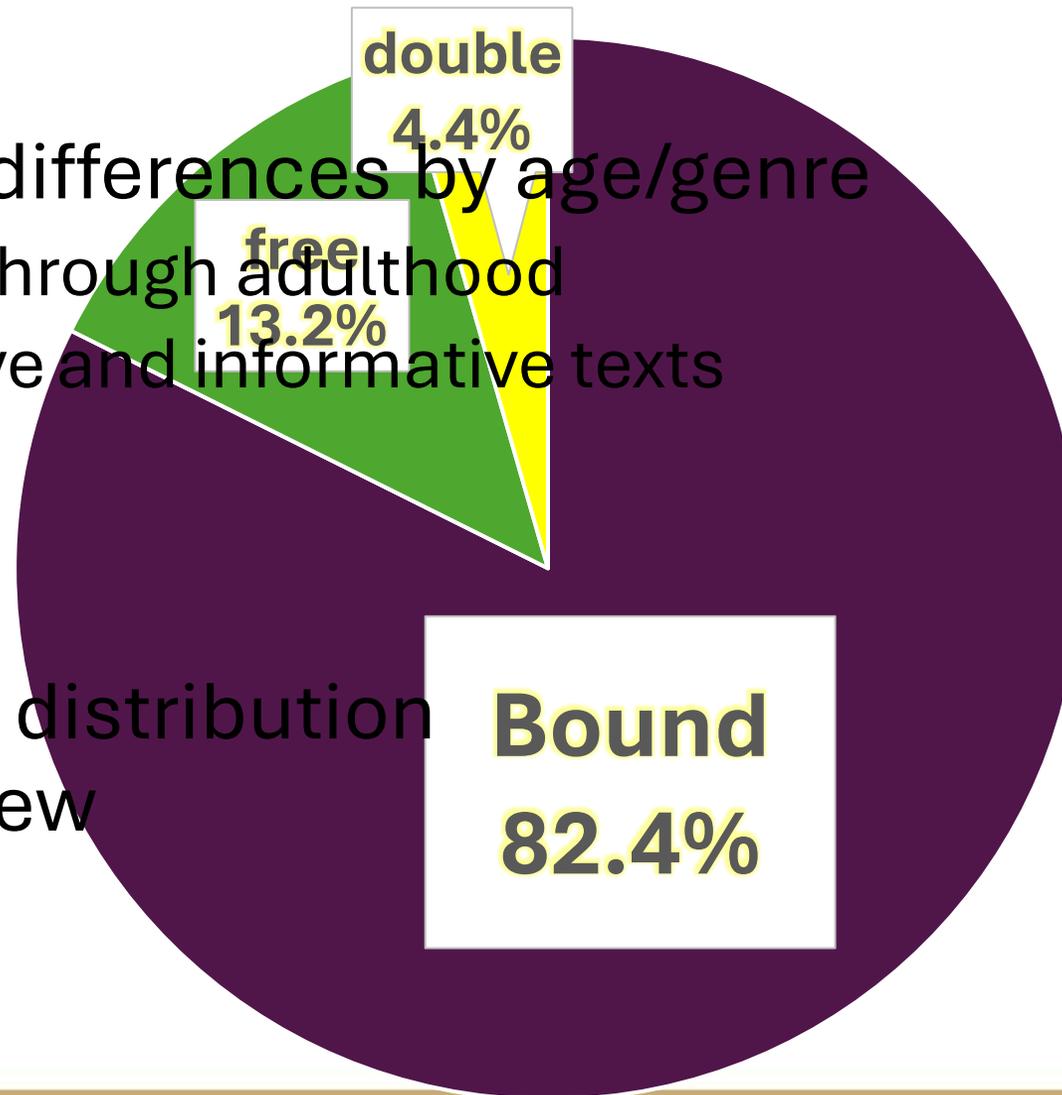
Distribution
of the Three Compound Structures

double
4.4%

free
13.2%

Bound
82.4%

Double compounds by age

Type–Token Ratio (TTR) by Compound Structure

✓ **Bound compounds**: Higher level of repetition, indicating a tendency toward collocationality

  • Bound compounds ar

Lower TTR =

more repetition

0.78
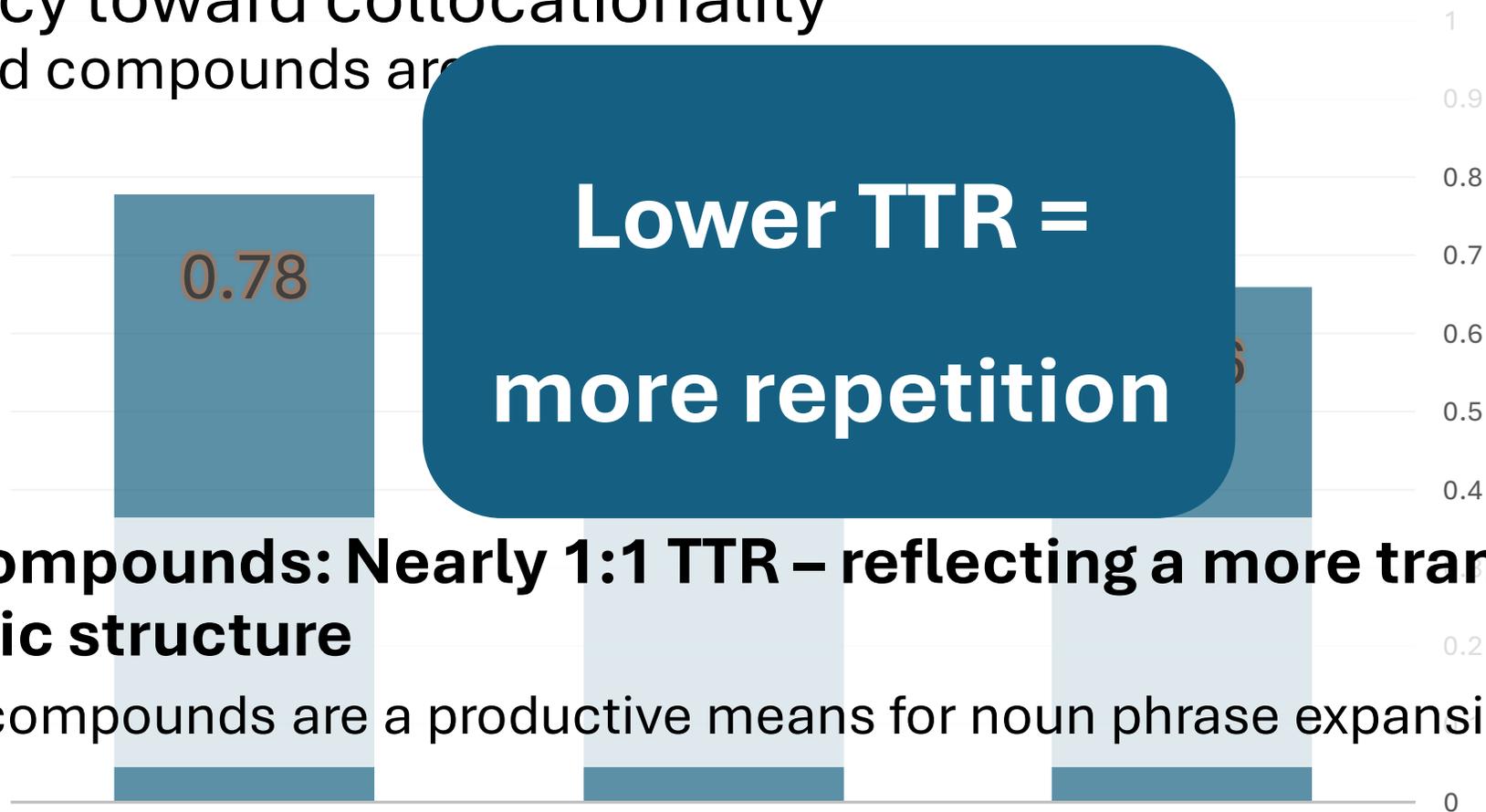
✓ **Free compounds: Nearly 1:1 TTR – reflecting a more transparent syntactic structure**

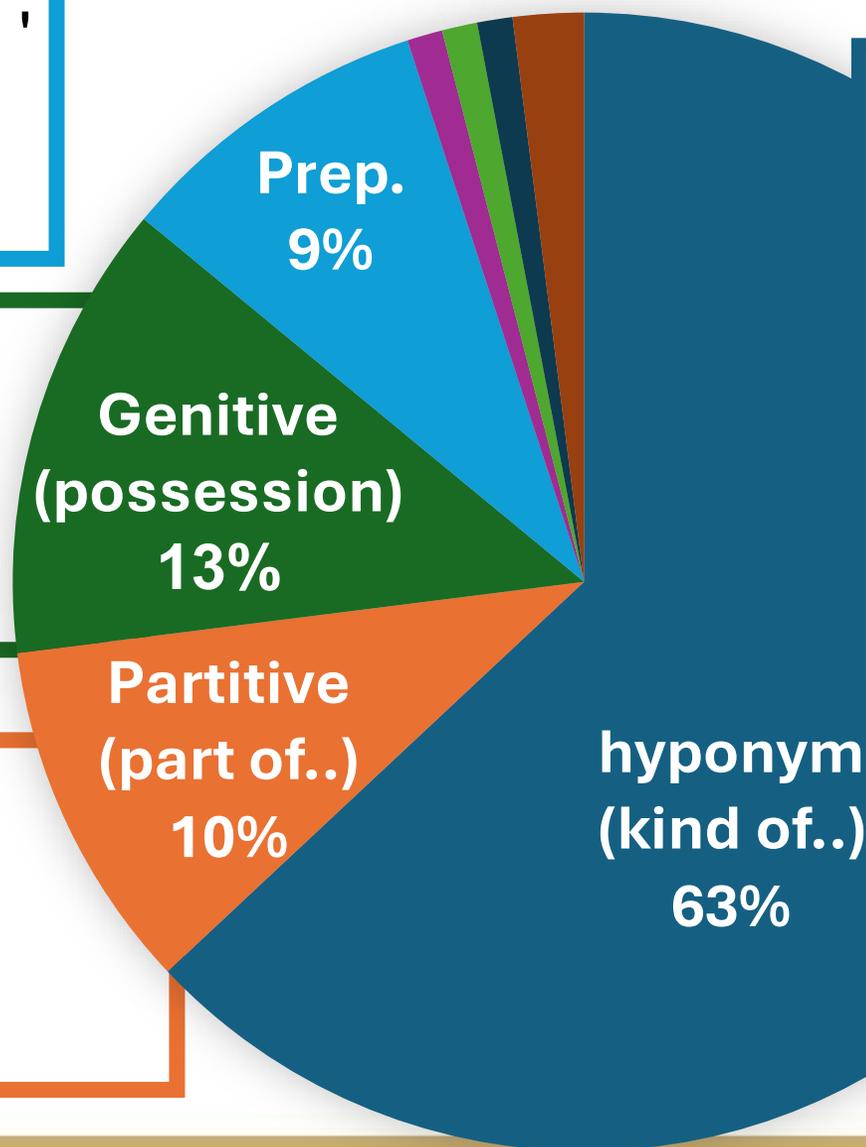 • Free compounds are a productive means for noun phrase expansion

Double   free   bound

semantic relations between constituents

semantic relations

•*be-tsad ha-kvish*
'on the side of the road '
•*mi-ta'am ha-bank*
' from the bank '

•*hora'ot ha-more*
'the teacher's instructions'
•*mishpaxto shel Yossi*
'Yossi's family '

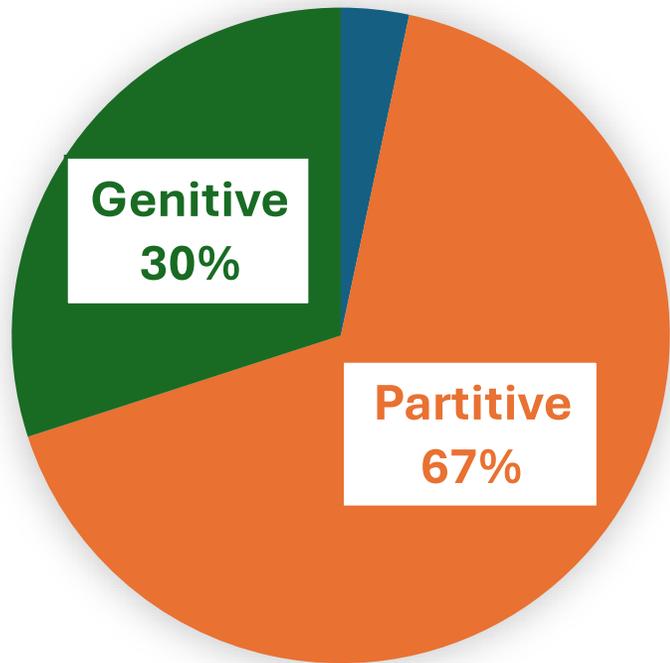•*masax ha-makhshir*
'the device's screen'
•*rosha shel ha-isha*
'the woman's head'

•*shnat ha-limudim*
'academic year'
•*shnat sherut*
'service year'
•*kvutza shel anashim*
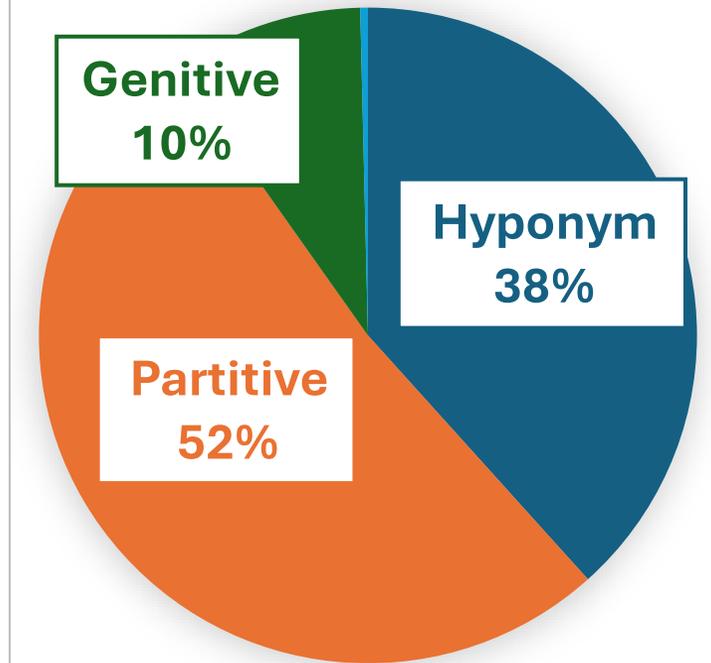'group of people'
•*kvutza shel rakdaniyot*
'group of dancers'

Prep.
9%

Genitive
(possession)
13%

Partitive
(part of..)
10%

hyponym
(kind of..)
63%

24
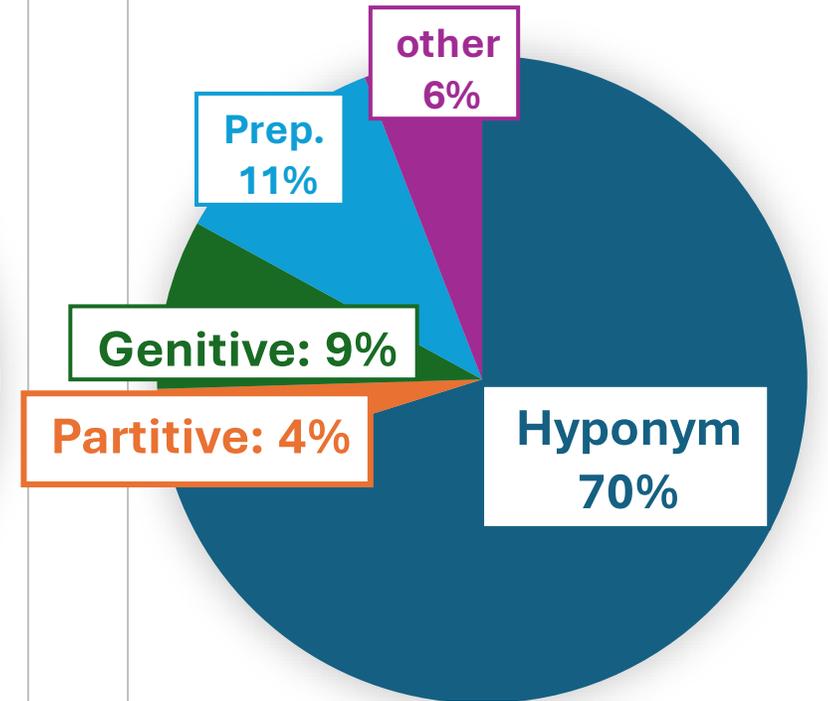
Morpho-syntactic Profile

Morpho-syntactic Profile by Compound Type

- More **syntactic expansions** in free and double compounds than in bound compounds

- **complex constituents** appeared mainly in free compounds

- **Optional possessive inflection**: more frequent in free than in bound compounds
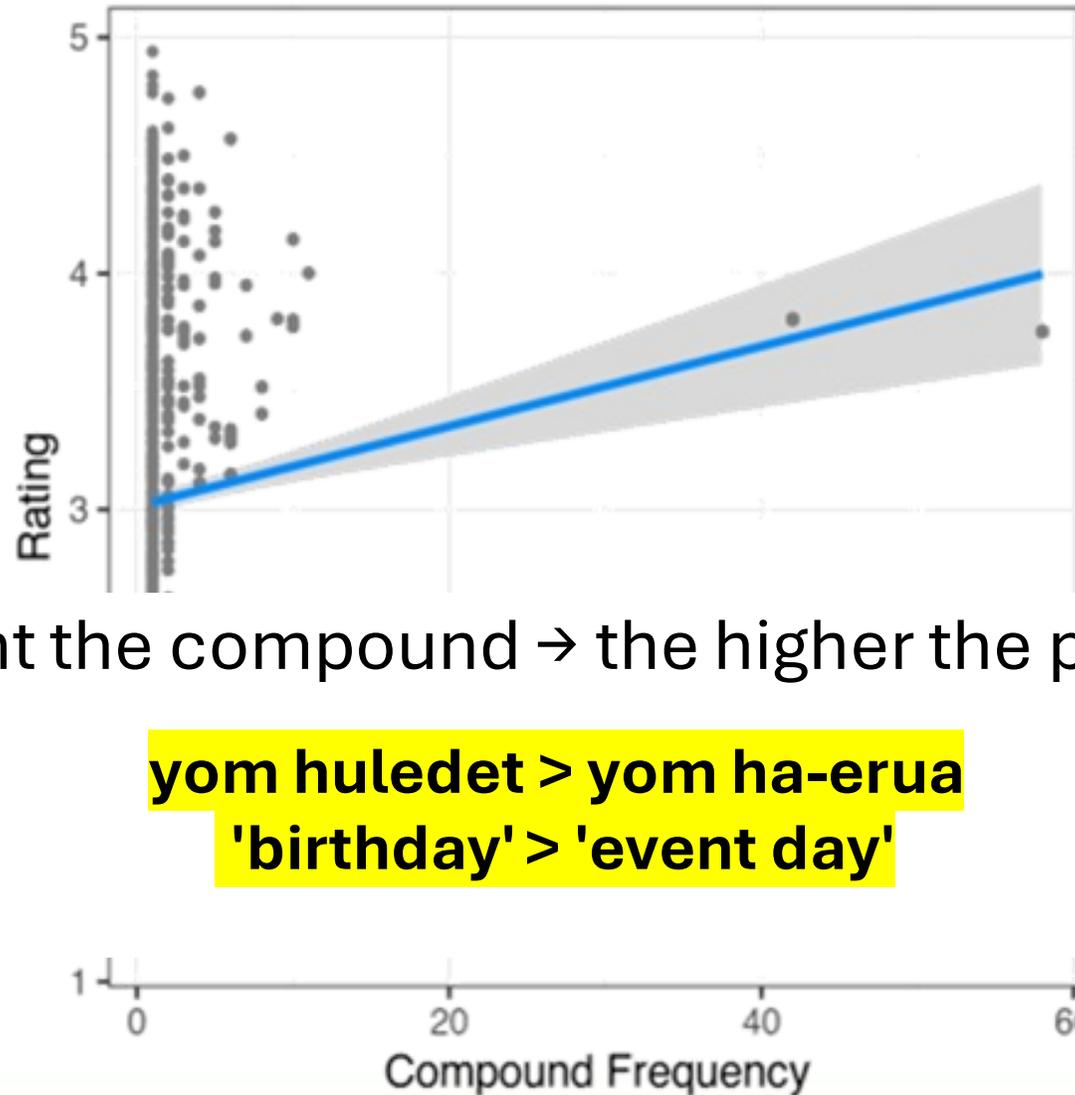
Predictability Level

judgment questionnaires: inter-rater agreement

✓ High inter-rater agreement was found, indicating the reliability and validity of the developed tool

What Influences the Predictability Rate?

1: Compound Frequency

The more frequent the compound → the higher the predictability rating

yom huledet > yom ha-erua
'birthday' > 'event day'
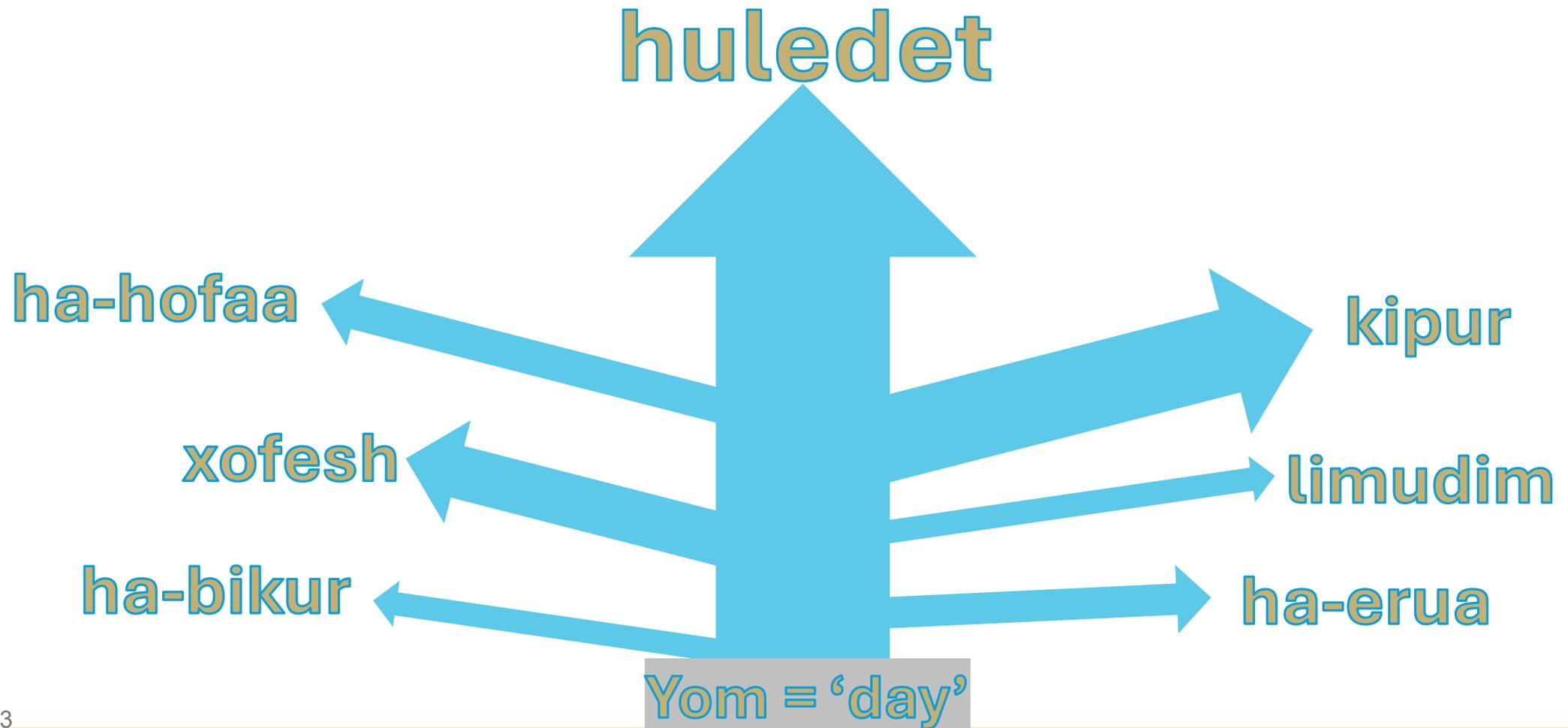
2: Productivity of the First Constituent

1. *yom ha-imun* = training day
2. *yom ha-erua* yad tneve =
3. *yom ha-bikur* = visit day
4. *yom huledet* yadhtrib =
5. *yom ha-hofa'a* yad ecnamrofrep =
6. *yom xofesh* ffo yad =
7. *yom yerushalayim* yad melasureJ =
8. *yom kipur* ruppiK moY =
9. *yom ha-limudim* yad yduts =
10. *yom ha-mivxan* yad noitidua =
11. *yom ha-masa* yad yenruoj =
12. *yom ha-kontsert* = concert day

The more productive the first constituent → the lower the predictability rating

==**yom huledet < zeuf panim**==
==**'birthday' < 'sour-faced'**==

huledet

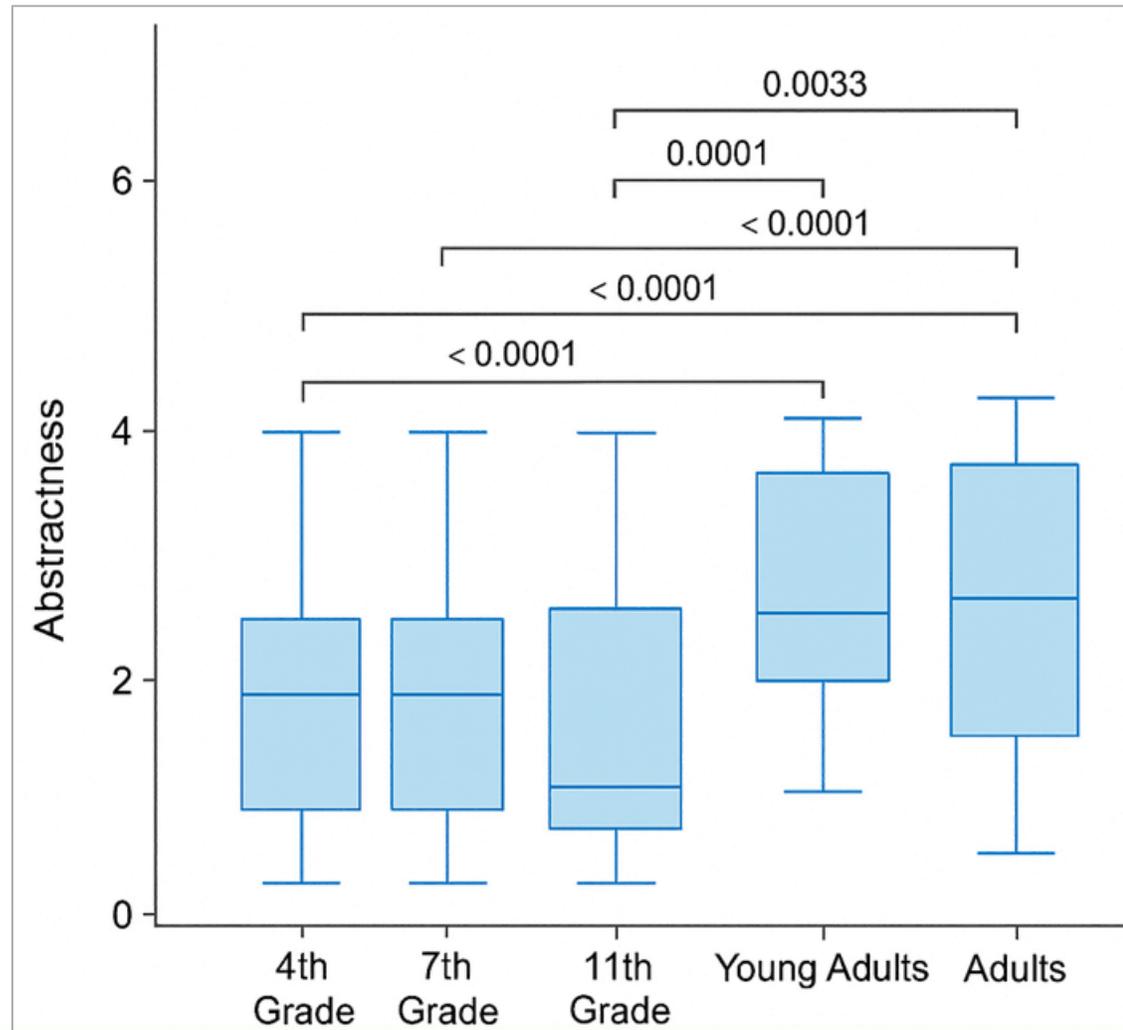ha-hofaa

kipur

xofesh

limudim

ha-bikur

ha-erua

Yom = 'day'

What affects the predictability level of compounds?

× No age effect

× No genre effect

↖ A positive effect of frequency

↙ A negative effect of first constituent productivity

Constituent Abstractness Level

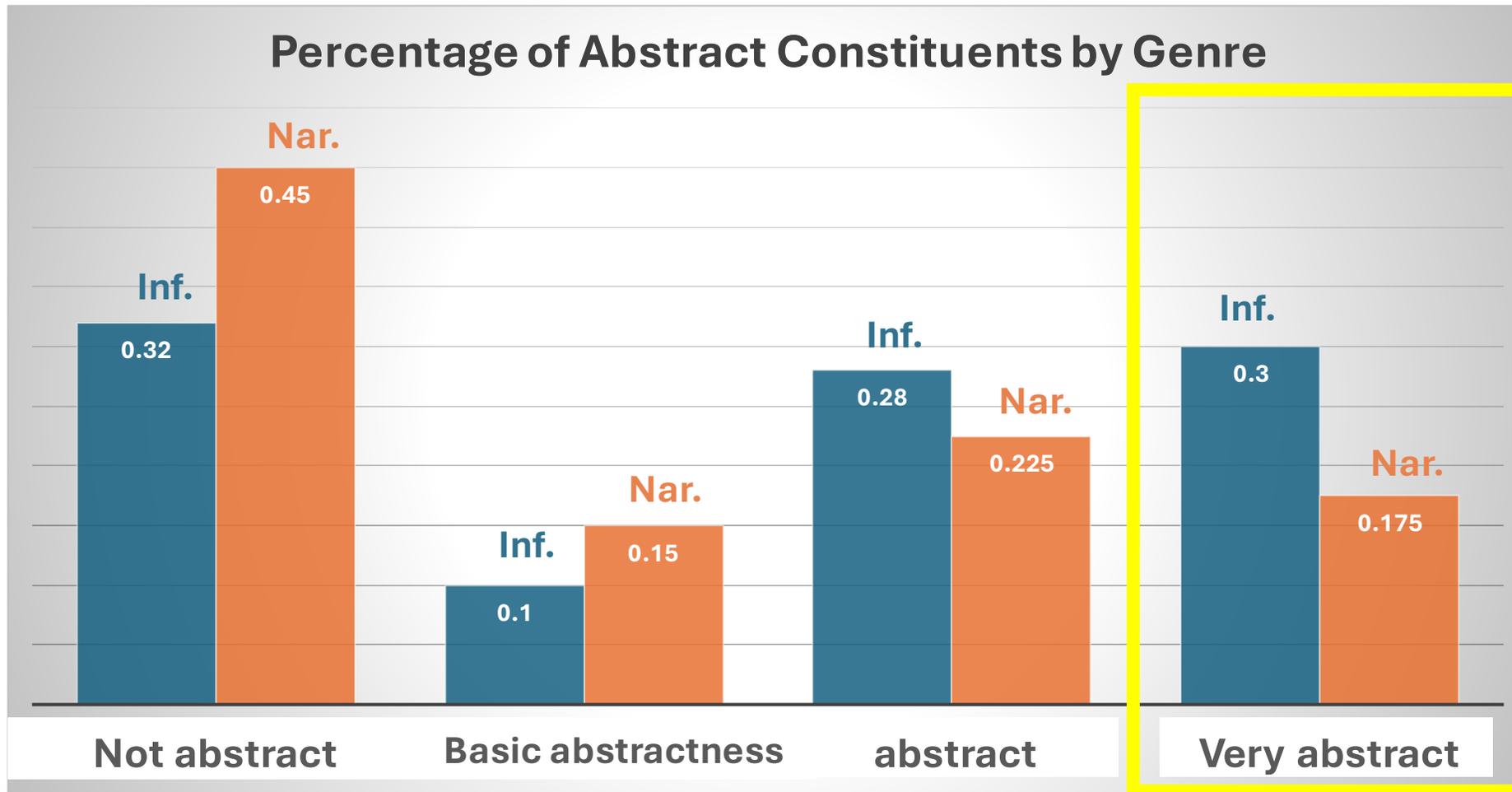# **Abstractness by age**

- Percentage of abstractness increases with age

Abstractness by genre

- Compounds in informative texts are more abstract

Percentage of Abstract Constituents by Genre

	Not abstract	Basic abstractness	abstract	Very abstract
Inf.	0.32	0.1	0.28	0.3
Nar.	0.45	0.15	0.225	0.175

SUMMARY & DISCUSSION

Structural Pattern

- Most compounds are bound
  - In Biblical Hebrew, only bound compounds
  - In Mishnaic Hebrew, the word "shel" (of) was added

- TTR:
  1. Bound compounds show the most repetition
  2. Free compounds show almost no repetition

- Morpho-syntactic profile: more complex in free than in bound compounds

STRUCTURE & MEANING

- The most common compounds in Hebrew are hyponyms
    - Birthday / performance day
    - Applications for players / applications for notices

- Different structures express different meanings:
    - Bound compounds: mainly hyponyms + diversity across all semantic categories
    - Free compounds: genitive + also hyponyms and partitive
    - Double compounds: genitive + partitive

41

Predictability level

- A new way to address a complex aspect — predictability

- We designed the first questionnaires to measure it!

- A reliable tool for mapping predictability:
  highly predictable → → → → unpredictable second constituent

- Predictability rating increases as:
  1. Compound frequency increases
  2. First constituent morphological family size decreases

Abstractness Level

- Abstractness level increases with age

    - A clear leap between school-age groups and young adults

        4th + 7th + 11th < young adults + adults

- Informative texts are more abstract than narrative texts

Thank you

Avitalbru@gmail.com