The productivity of morphological and syntactic evaluative constructions in Italian:
exploring networks and daughters

Francesca Masini & Flavio Pisciotta

University of Bologna & University of Salerno

UNRAVELING LINGUISTIC PRODUCTIVITY • GHENT, MAY 21–23, 2024

Productivity: what else can we possibly say?

- Productivity has been assessed by both quantitative and qualitative factors, such as:
  - construction-internal constraints on 'open slot' (e.g., phonological, semantic, morphological)
  - level of schematicity and/or level of entrenchment for constructions (e.g., Barðdal 2008; Perek 2020)
  - productivity measures like type/token ratio, hapax/token ratio, etc. (e.g., Baayen 1991, Zeldes 2001)
- Productivity measures, however, sometimes give unexpected results
- Evaluative morphology is a case in point
  - Some low-frequency evaluative affixes are recognized by quantitative measures to be more productive than more frequent evaluative and derivational affixes (Gaeta & Ricca 2003; Albair 2010)

> "Does it make sense, linguistically, that such low-frequency items exhibit a top value in productivity?" [Gaeta & Ricca 2003: 109]

  - This poses questions on how much these productivity measures can account for entrenchment

☞ since productivity is generally seen as a correlate of schema entrenchment (Lieven 2010; Stefanowitsch & Flach 2015, a.o.), should we assume that unfrequent evaluative schemas are more entrenched than regularly exploited ones?

Evaluative morphology *vs.* canonical derivation

- Now, one basic fact about evaluative morphology is that evaluative items are **normally nonce formations**, created on the spur of the moment (1), contrary to canonical derivation
  - Although they may be stored if they are frequent enough (2) or if they acquire an idiosyncratic meaning (3)
- Italian

(1) *managerino*	manager.DIM	'little manager'
(2) *cucchiaino*	spoon.DIM	'teaspoon'
(3) *bocconcino*	bite.DIM	'delicacy'

Evaluative morphology

- Mostly nonce-formations
- Occasionally stored

Canonical word-formation

- Mostly stored items
- Occasionally nonce-formations

Evaluative morphology = syntax?

"[**evaluative] items do not fully behave as derivational items, but rather border on syntax**, and therefore their productivity cannot be straightforwardly compared with the one displayed by core instances of bound derivational processes."
[Gaeta & Ricca 2003: 109-110; emphasis added]

- This alleged similarity between evaluative morphology and syntactic constructions, hypothesized by Gaeta & Ricca (2003), has been left **unexplored**
- In the meantime, a wealth of studies on productivity appeared, proposing new measures and methods, and we also started to talk about morphology in constructionist terms (Booij 2010)

Research questions for today

- Today we address this unexplored question by looking at a set of Italian **morphological** and **syntactic constructions expressing approximation** (Masini & Micheli 2020; Masini, Norde & Van Goethem 2023)
- We adopt a constructionist perspective, which assumes no principled morphology-syntax division
  - Interestingly, in CxG, the intuition that evaluatives tend to behave more similarly to syntactic constructions (than to canonical derivation) can be re-interpreted in terms of **networks and daughters**
  - Evaluative morphological schemas tend to be instantiated more by **constructs** (like syntax) than by lexically specified **daughter constructions** (like derivation)
- Keeping this in mind, our aim is to:

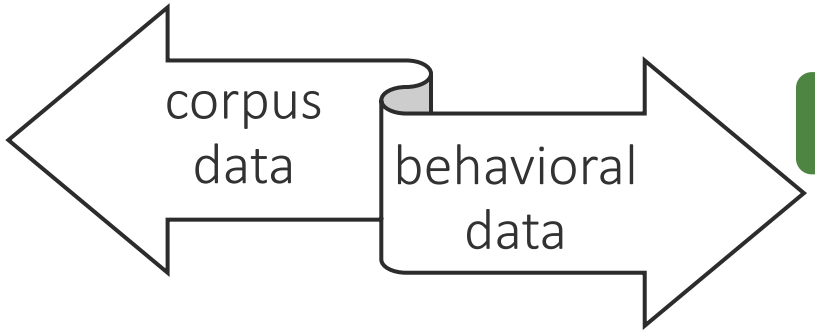
RQ1) verify if evaluative morphological schemas actually pattern with syntactic schemas in terms of productivity

RQ2) understand the relationship between productivity, entrenchment and familiarity in different types of cxns

RQ3) clarify how all this can be modelled in CxG, focusing on the role of networks and daughters

Methods: construction selection

- We selected **7 constructions** (cxns), including:
  - 2 established syntactic constructions (1-2)
  - 4 morphological constructions (3-6)
    - Some of which are more 'emergent' (5-6), falling in-between morphology and syntax
  - 1 frequent and productive derivational prefix (Gaeta & Ricca 2003) as a benchmark



SYNTACTIC_EVALUATIVE

1) ***specie di* X** 'kind/sort of X'
   - *una sorta di guerriglia contro tutte le istituzioni* 'a sort of guerrilla warfare against all institutions'
2) ***sorta di* X** 'kind/sort of X'
   - *costretto a fare una specie di corsa a ostacoli* '[said of commuter] forced to do a kind of obstacle race'

MORPHOLOGICAL_EVALUATIVE

3) ***pseudo-*X** 'pseudo-X'
   - *uno pseudo satanista che veste nero per vendere dischi* 'a pseudo satanist who dresses in black to sell records'
4) ***semi-*X** 'semi-X'
   - *Il palazzo lasciato semi-deserto dai vacanzieri*. 'The building left semi-deserted by vacationers'
5) ***mezzo-*X** 'half-X'
   - *Se i deputati sapessero, ci sarebbe una mezza rivolta.* 'If deputies knew, there would be half a riot'
6) ***non-*X** 'non-X'

DERIVATIONAL_MORPHOLOGY

7) ***in-*X** 'un/in-X'
   - *insufficiente* 'insufficient', *incapace* 'unable', *irregolare* 'irregular'

Methods I: corpus data & productivity

- We extracted a **sample of 500 occurrences** for each of the 7 selected cxns (except for *non*-) from the CORIS2021 corpus of contemporary written Italian (165Mw, https://corpora.ficlit.unibo.it/TCORIS/)
  - We decided not to control for the lexical category of the base, but we included only examples that convey approximation (not other possible meanings)
- We compared the 7 cxns by computing **various measures**, including both traditional productivity measures and newly developed ones to evaluate different aspects of the type and token distributions (e.g., the conventionalization of specific items)
  - We used the R package *zipfR* (Evert & Baroni 2007) to calculate the productivity measures

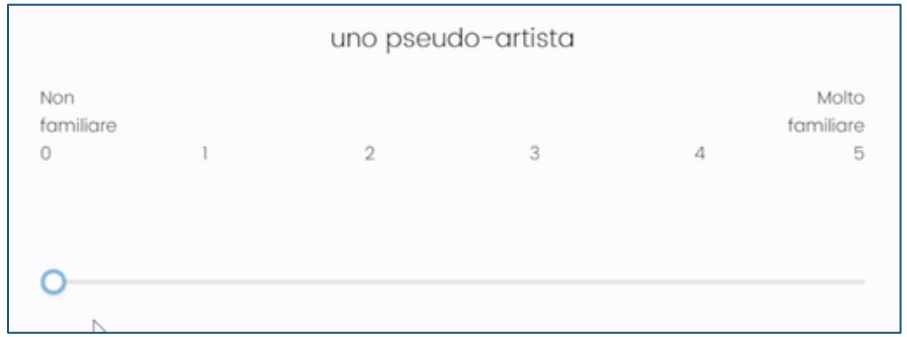
Productivity measures	**Conventionalization/entrenchment measures** (Van den Heede & Lauwers 2023)
V = number of types	**alpha** = approximation of the α coefficient of Zipf's law
TTR = type/token ratio	**FrTop1** = frequency of the top-ranked type
P = hapax/token ratio (Baayen 1991)	**FrTop3** = average frequency of the three top-ranked types
Hapax = hapax/type ratio (Baayen 2009)	**StDevTop3** = std. deviation of the frequency of three top-ranked types
Shannon's Entropy (Pankratz et al. 2022)	

Methods II: behavioral data & familiarity

- Then we decided to explore another dimension, namely the **familiarity** of these cxns by means of a **judgement test**, using Qualtrics for the survey and Prolific for recruiting participants
  - Familiarity is intended as a correlate of the strength of the mental representation of an item (Caldwell-Harris et al. 2012; Verhagen 2020), hence it should be a correlate of its frequency and entrenchment
- We designed the test to include **24 stimuli**
  - We included **6 constructions**
    - We included only *sorta di* X and not *specie di* X (because they behave very similarly in the corpus data)
  - For each of the 6 constructions we picked **3** (**real**) **examples** from our dataset
    - 1 top-ranked, 1 median, 1 hapax (randomly chosen)
    - All bases (Xs) were checked for use
      - They are all included in the *Basic Vocabulary* for Italian (as FO=fundamental or AU=high usage)
  - We added **6 fillers** of various types to check for the quality of responses
    - 1 familiar & 1 unfamiliar syntactic expression of the 'N of N' type
    - 1 familiar & 1 unfamiliar derived word
    - 1 familiar & 1 unfamiliar simple word

Methods II: behavioral data & familiarity

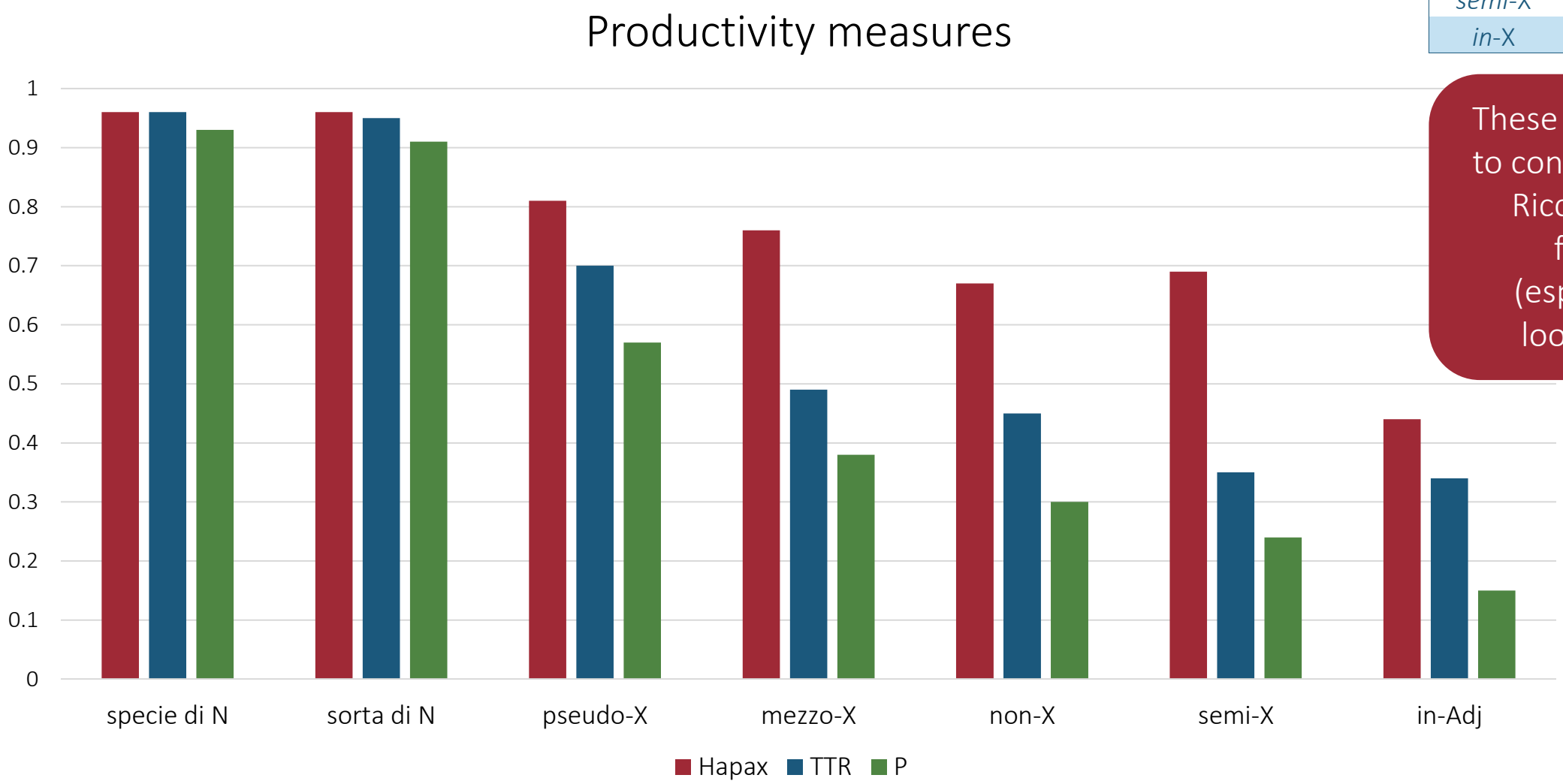
- 200 participants were asked to rate for familiarity the 24 stimuli on a **six-point Likert scale** (0-5)
  - Age range: 20-67 (Average 35, StDev 11.04)
  - Sex: female 50%, male 50%
- The participants were asked to rate by dragging a **sliding bar** going from 0 (not familiar, never heard of) to 5 (very familiar)



ID	Stimuli	Glosses	Type of stimulus
Q1	**una sorta di museo**	a sort of museum	top-ranked
Q2	**una sorta di scatola**	a sort of box	hapax
Q3	**una sorta di villaggio**	a sort of village	median
Q4	**un mezzo-sorriso**	a half-smile	top-ranked
Q5	**una mezza-notizia**	a half-news	hapax
Q6	**mezzo-spento**	half-switched_off	median
Q7	**semi-nudo**	semi-naked	top-ranked
Q8	**semi-magico**	semi-magical	hapax
Q9	**semi-professionale**	semi-professional	median
Q10	**pseudo-religioso**	pseudo-religious	top-ranked
Q11	**una pseudo-informazione**	a pseudo-information	hapax
Q12	**uno pseudo-artista**	a pseudo-artist	median
Q13	**una non-vita**	a non-life	top-ranked
Q14	**una non-regola**	a non-rule	hapax
Q15	**un non-lavoro**	a non-job	median
Q16	**inutile**	useless	top-ranked
Q17	**irreale**	unreal	hapax
Q18	**ingiusto**	unjust	median
Q19	**un episodio di violenza**	an episode of violence	filler_syn_fam
Q20	**un nottolino di arresto**	a door_latch of stopping	filler_syn_nofam
Q21	**ricominciare**	restart	filler_mor_fam
Q22	**antemarcia**	premarch	filler_mor_nofam
Q23	**una canzone**	a song	filler_simple_fam
Q24	**una brenna**	a nag (horse)	filler_simple_nofam

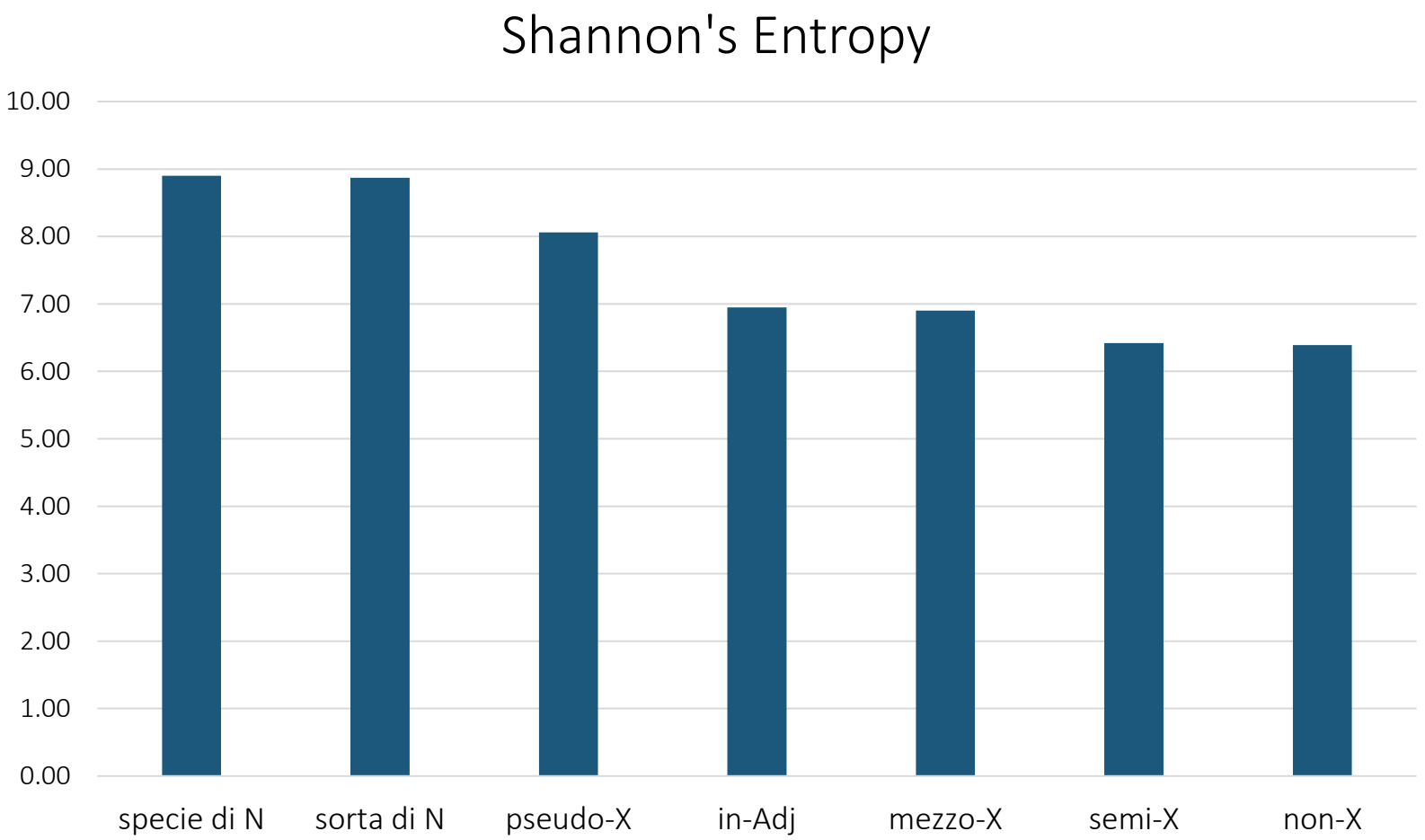
Results: corpus data & productivity

Cxn	Tokens	Types (V)
specie di X	500	483
sorta di X	500	477
pseudo-X	500	351
mezzo-X	500	246
non-X	361	161
semi-X	500	176
in-X	500	169



These results seem to confirm Gaeta & Ricca's (2003) findings (especially by looking at *P*)

Results: corpus data & productivity



Shannon's Entropy = uncertainty of encountering a specific type in a type frequency distribution

$$\mathrm{H}(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

Advantages (Pankratz et al. 2022):

- no privileged status of hapaxes
- comparability of samples of different sizes

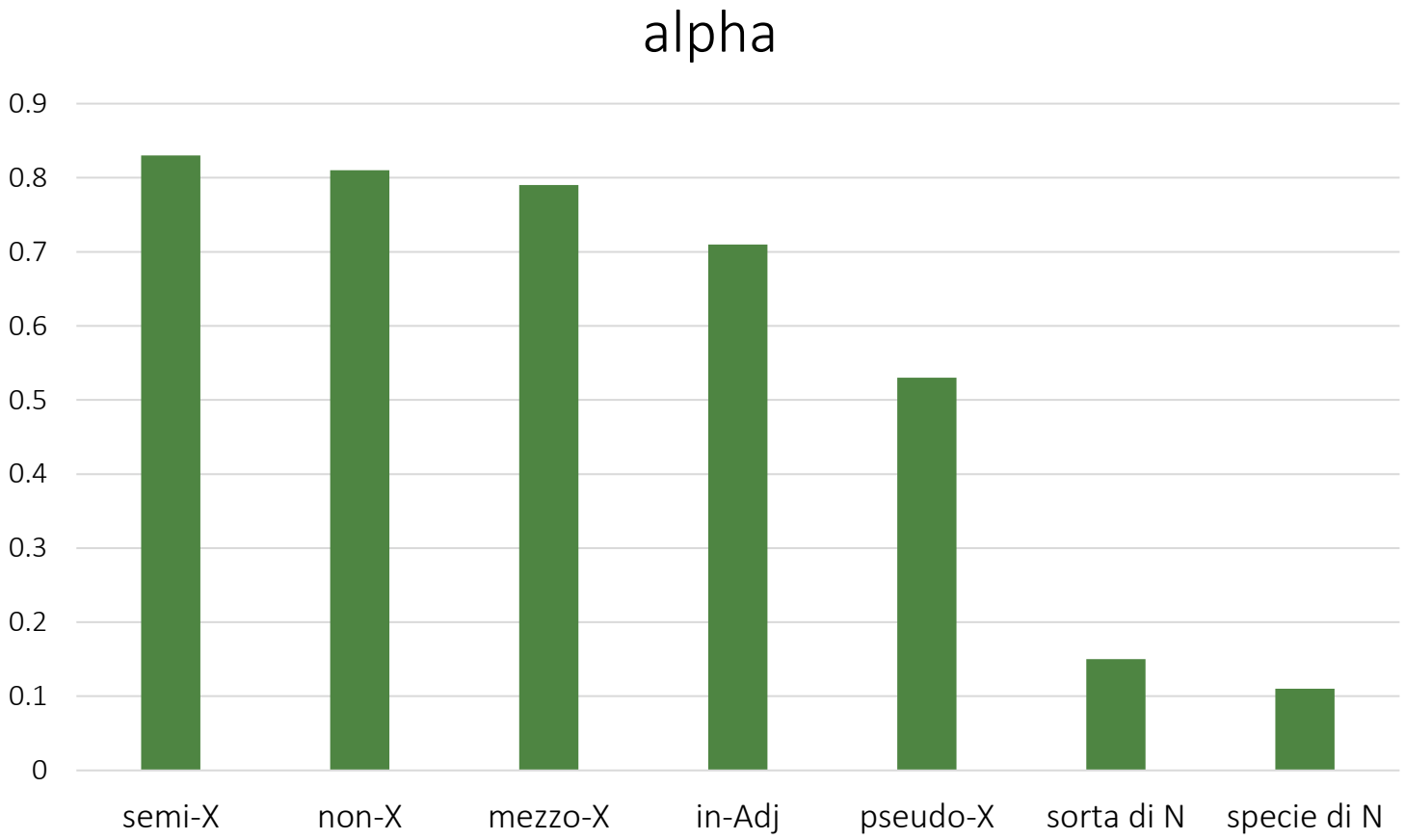
Notable difference (wrt TTR, P, HTR):

- *in-* seems to be **as productive as evaluative morphological cxns** (and even slightly more productive than some of them)

Results: corpus data & productivity

- **alpha** = *anti-productivity measure* (van Egmond 2013; Van den Heede & Lauwers 2023)
  - Absolute value of the slope of the linear relation between the log of the ranks and the log of the frequencies ☞ **rate of decline in frequency** from top-ranking items to less frequent ones

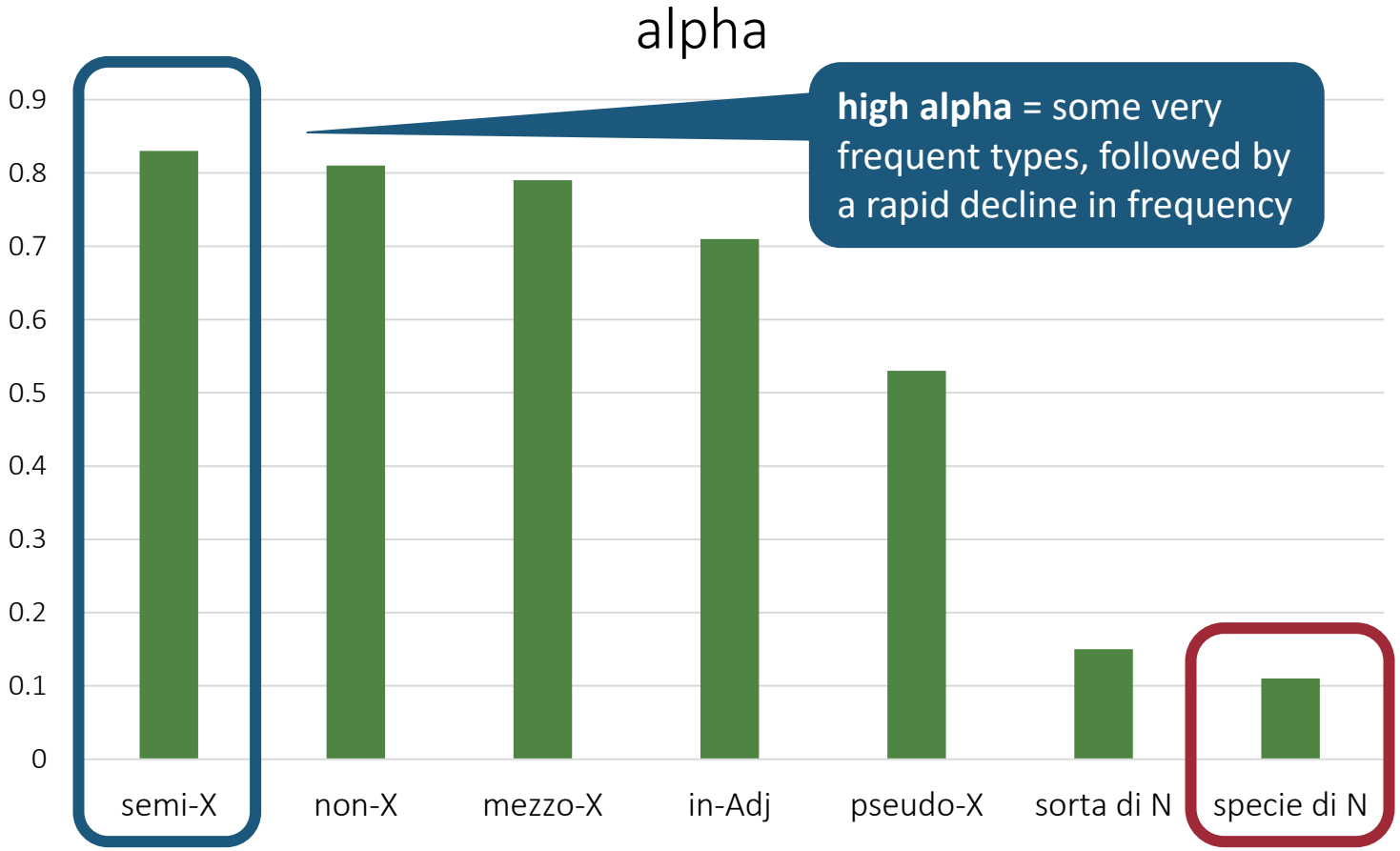
Cxn	FrTop1	FrTop3	StDevTop3
mezzo-X	63	35,3	20,49
non-X	57	27	21,27
semi-X	30	28,3	1,25
in-Adj	27	22,67	3,78
pseudo-X	25	15,67	7,04
sorta di N	3	3	0
specie di N	2	2	0

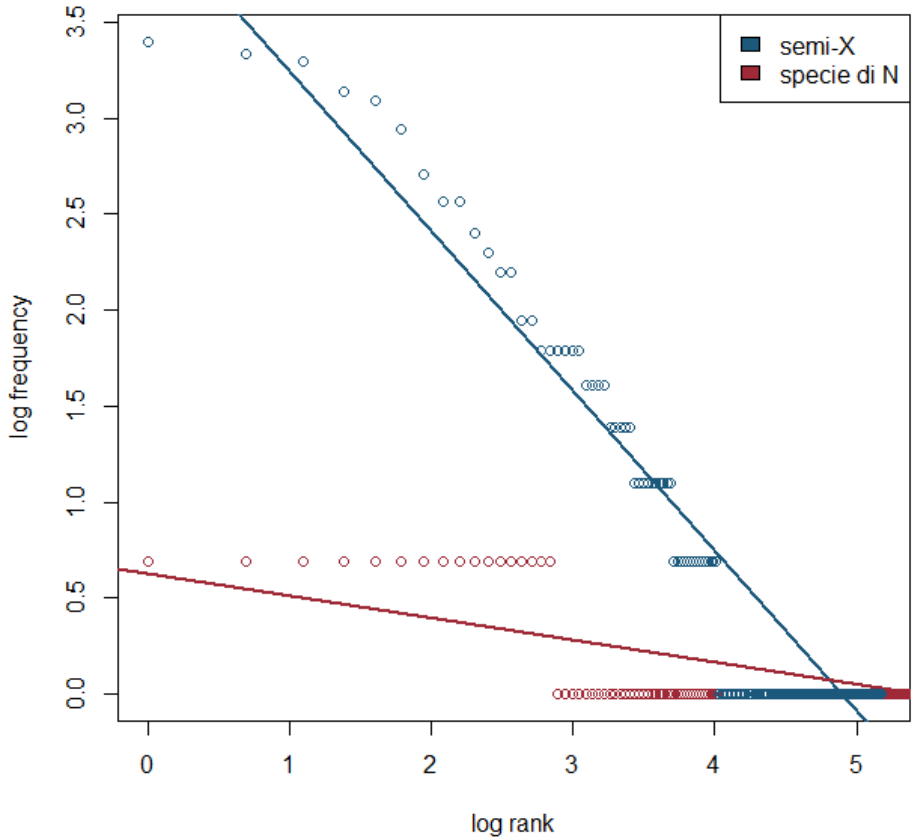


Results: corpus data & productivity

- **alpha** = *anti-productivity measure* (van Egmond 2013; Van den Heede & Lauwers 2023)
  - Absolute value of the slope of the linear relation between the log of the ranks and the log of the frequencies ☞ **rate of decline in frequency** from top-ranking items to less frequent ones

Cxn	FrTop1	FrTop3	StDevTop3
mezzo-X	63	35,3	20,49
non-X	57	27	21,27
semi-X	30	28,3	1,25
in-Adj	27	22,67	3,78
pseudo-X	25	15,67	7,04
sorta di N	3	3	0
specie di N	2	2	0





Corpus data & productivity: sum up

Productivity measures

- Evaluative **syntactic constructions** are very productive according to all the measures
- Canonical **word-formation** (*in-*) are much less productive (☞ sample too small?)
- Evaluative **morphology**:
  - *mezzo-*, *semi-* & *non-* always pattern alike, lying in the middle between syntax and morphology
  - *pseudo-* seems closer to syntactic constructions

Conventionalization measures

- Evaluative **syntactic constructions** show no peaks in frequency (maxFr = 3, alpha very low)
- ***pseudo-*** stands in between, but this time is **close to *in-*** by all measures
- *mezzo-*, *semi-* & *non-* show **high conventionalization** scores

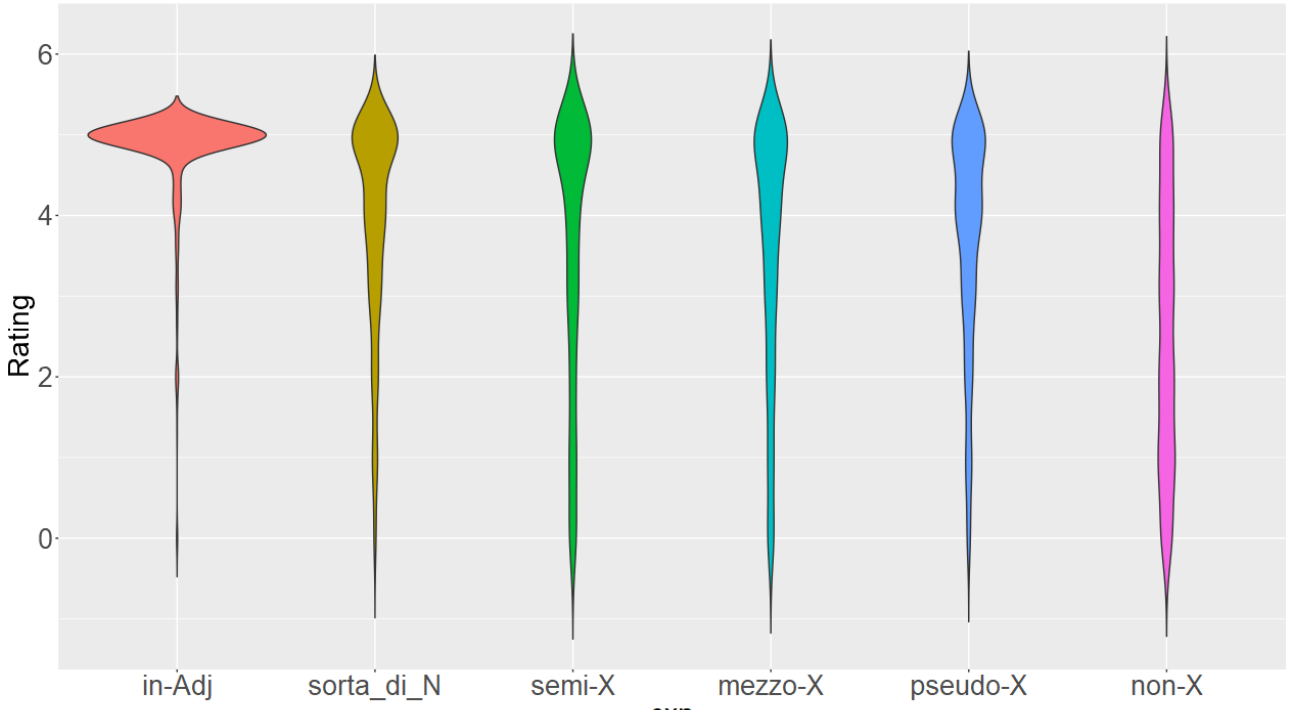
  ☞ Top-ranking items are much more frequent than the rest of the distribution

Productivity ~ Conventionalization

- As found by previous studies (Barðdal 2008; Van den Heede & Lauwers 2023) there seems to be a **negative correlation between productivity and conventionalization** of top items; esp. **entropy and alpha** show the best fit (-0.97, p = 0.001)

Results: behavioral data & familiarity

Average familiarity rating by cxn

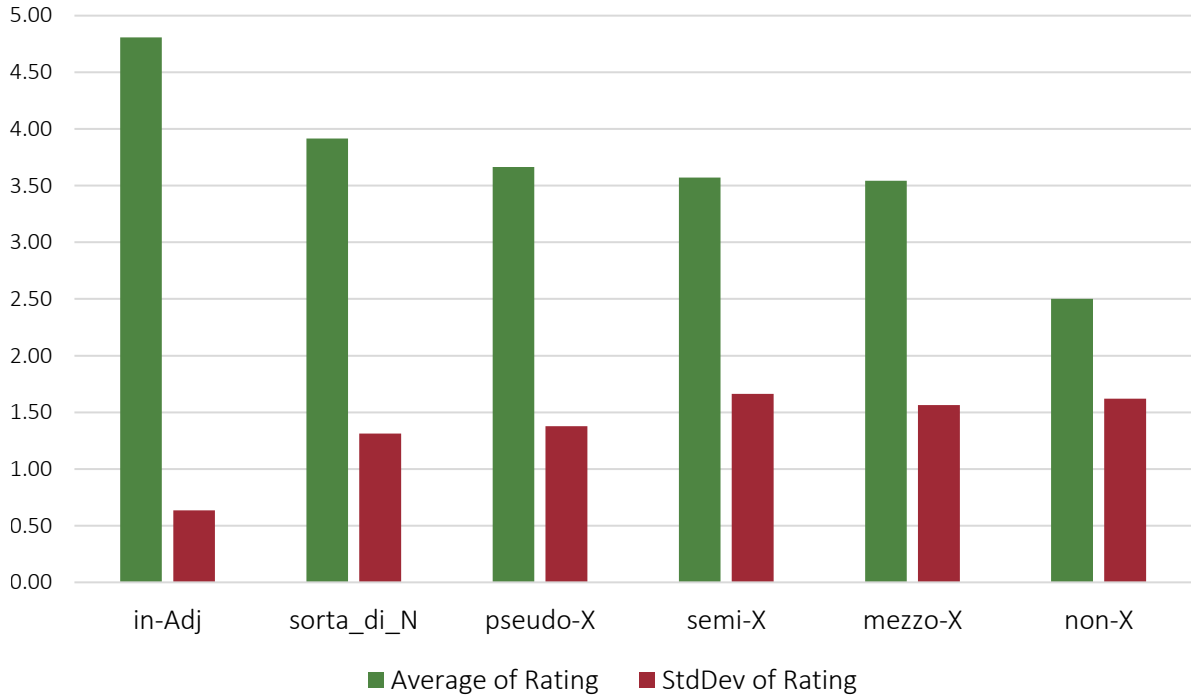


Kruskal-Wallis: $p < 0.0001$

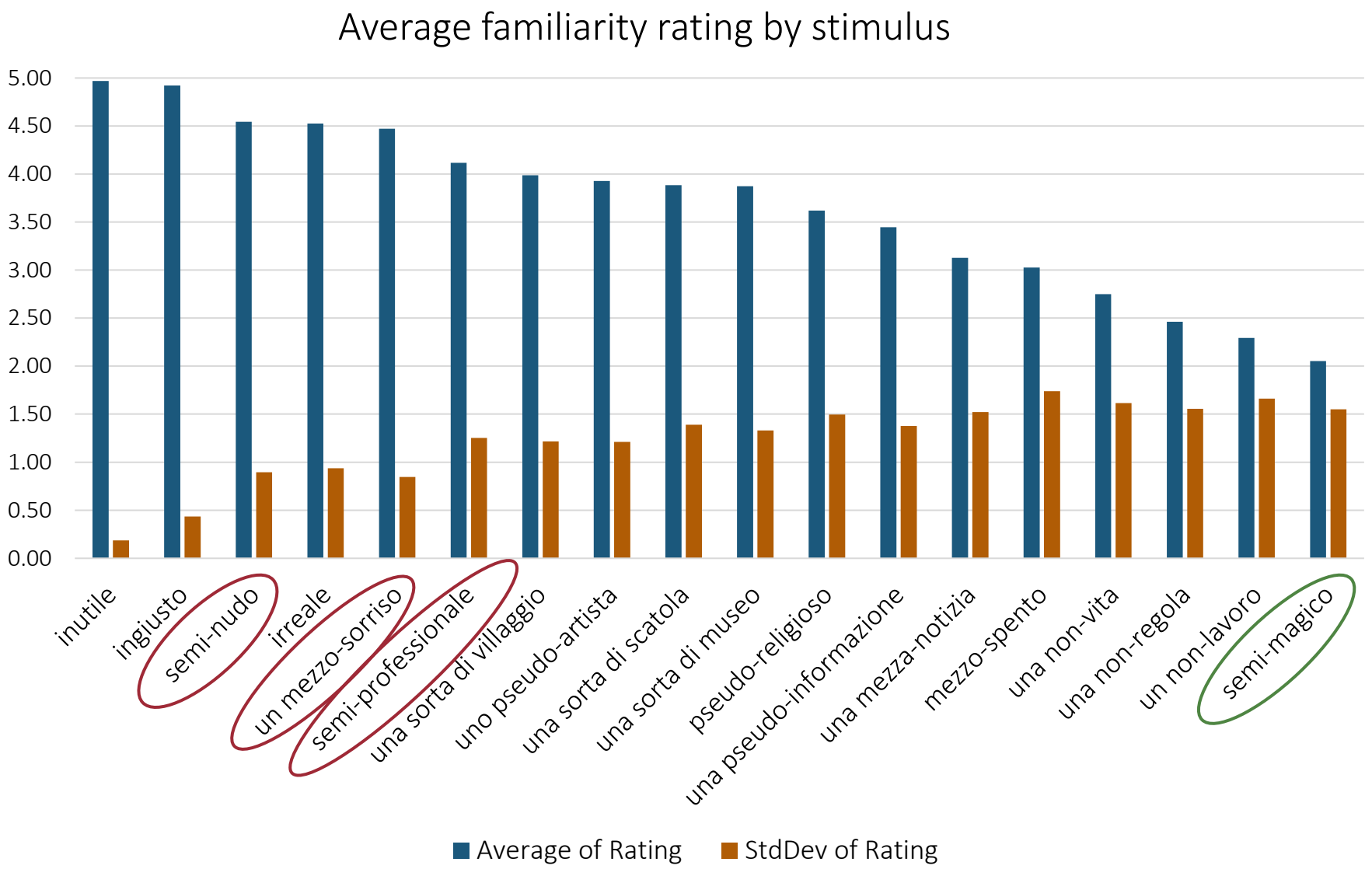
Significant differences (Dunn test w/ Holm-Bonferroni correction):

- *in-* vs the other cxns
- *sorta di* vs the other cxns
- *non-* vs the other cxns
- <u>No significant difference</u> between *pseudo-*, *mezzo-* and *semi-*

Interestingly, we found that the lower the **familiarity rating**, the higher the **standard deviation**

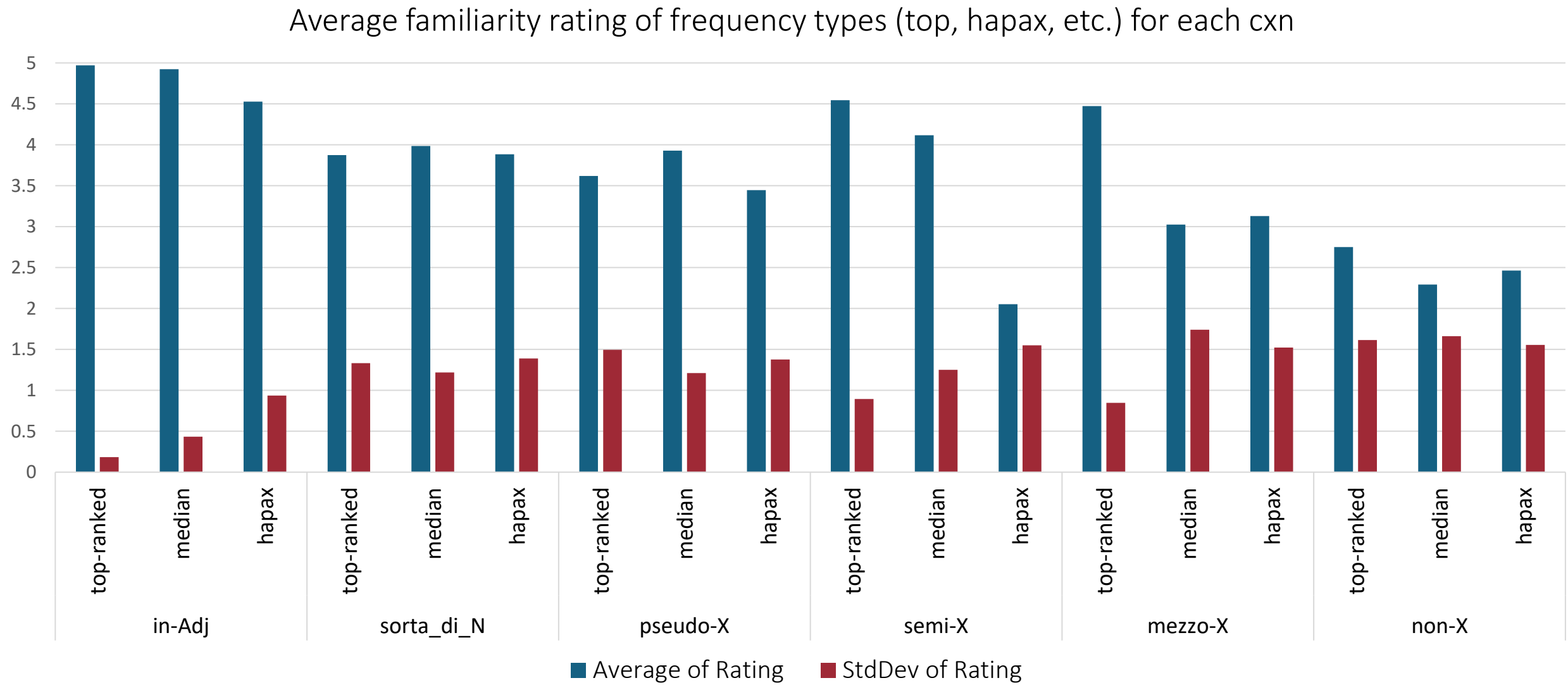


Results: behavioral data & familiarity



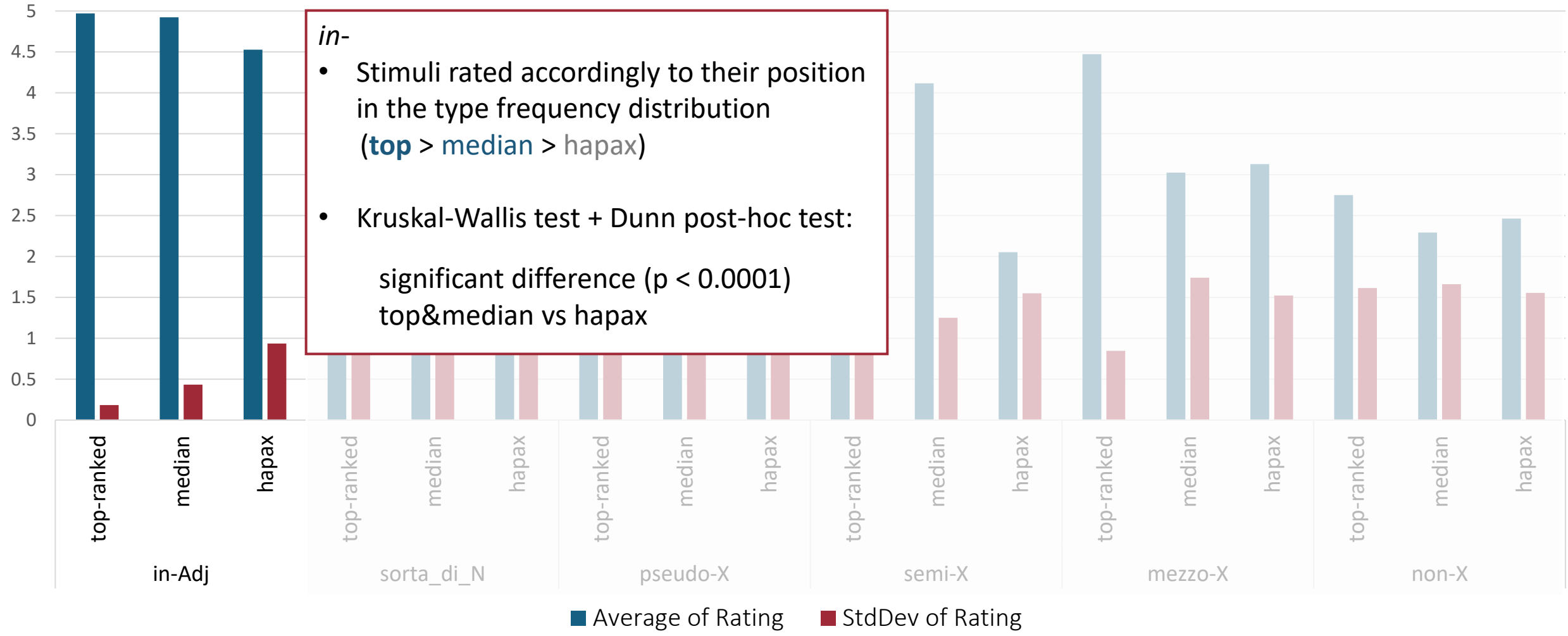
- Unexpectedly, the **top-ranking types of *semi-* and *mezzo-*** got rated almost as high as *in-*, and higher than *sorta di*
- Their less frequent types got quite low ratings
  - ☞ ***semi-magico*** (lit. 'half-magic') is even less familiar than *non-* types

Results: behavioral data & familiarity



Results: behavioral data & familiarity

Average familiarity rating of frequency types (top, hapax, etc.) for each cxn



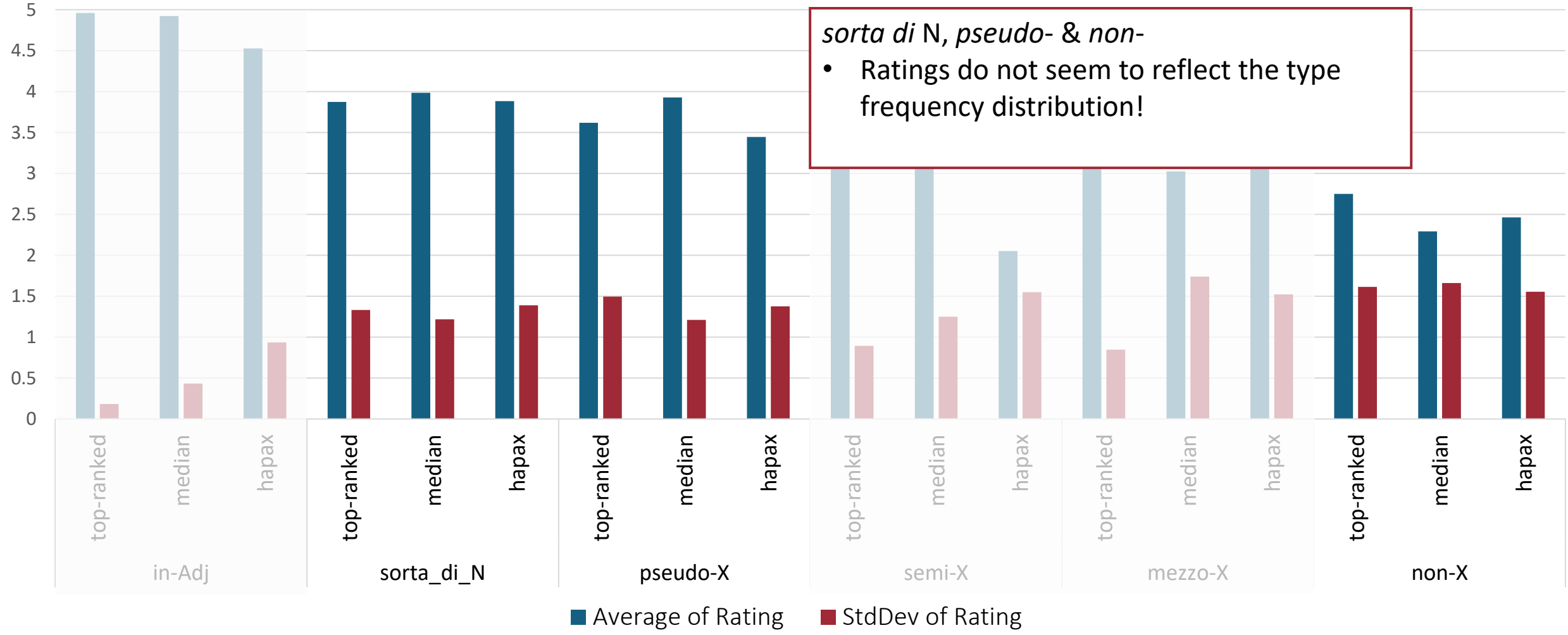
in-

- Stimuli rated accordingly to their position in the type frequency distribution (**top** > median > hapax)
- Kruskal-Wallis test + Dunn post-hoc test:

  significant difference ($p < 0.0001$)
  top&median vs hapax

Results: behavioral data & familiarity

Average familiarity rating of frequency types (top, hapax, etc.) for each cxn

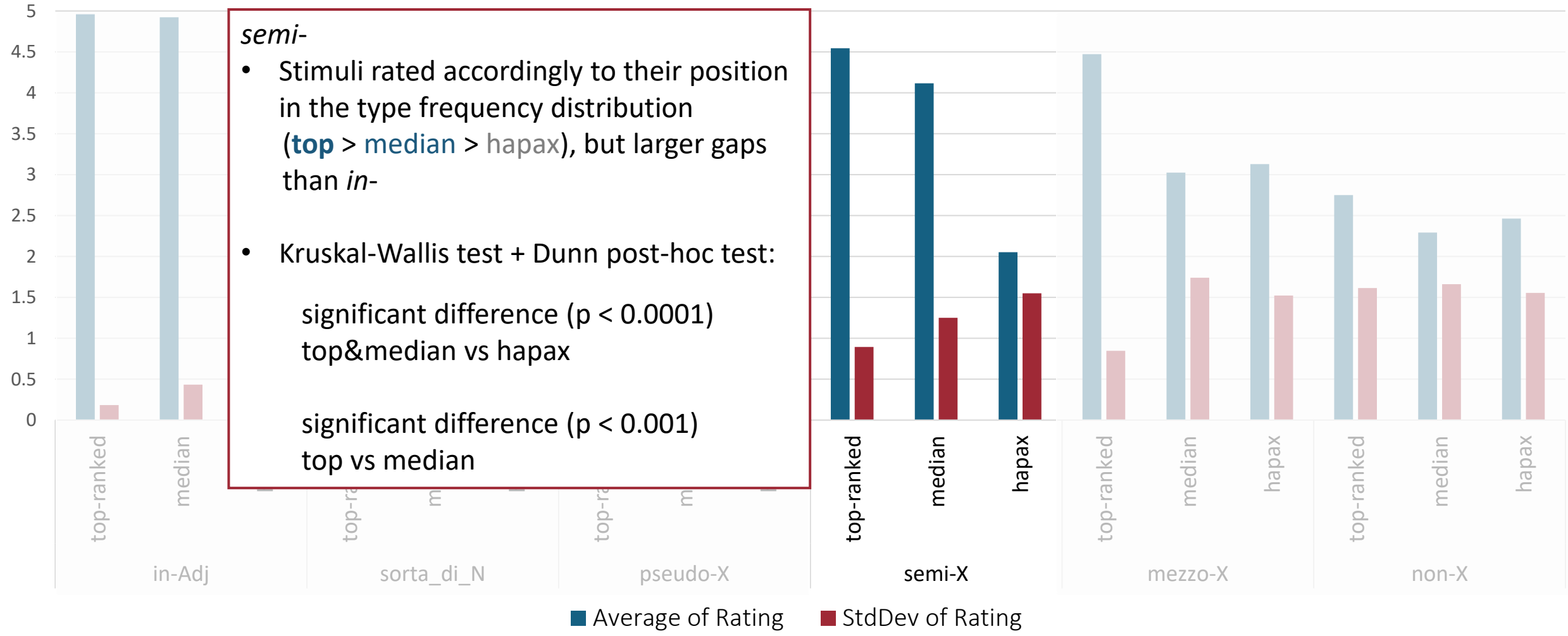


sorta di N, *pseudo-* & *non-*

- Ratings do not seem to reflect the type frequency distribution!

Results: behavioral data & familiarity

Average familiarity rating of frequency types (top, hapax, etc.) for each cxn



semi-

- Stimuli rated accordingly to their position in the type frequency distribution (**top** > median > hapax), but larger gaps than *in-*

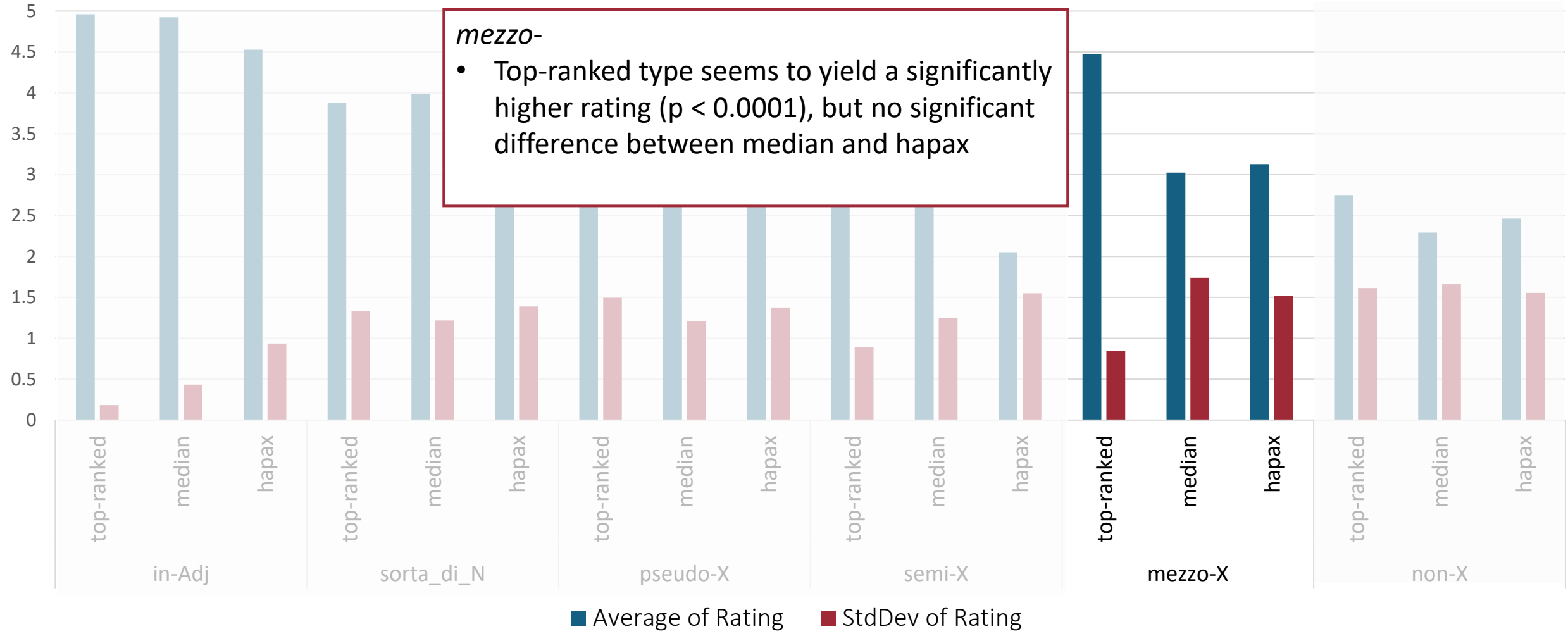
- Kruskal-Wallis test + Dunn post-hoc test:

  significant difference ($p < 0.0001$)
  top&median vs hapax

  significant difference ($p < 0.001$)
  top vs median

Results: behavioral data & familiarity

Average familiarity rating of frequency types (top, hapax, etc.) for each cxn



mezzo-

- Top-ranked type seems to yield a significantly higher rating ($p < 0.0001$), but no significant difference between median and hapax

Behavioral data & familiarity: sum up

Constructions

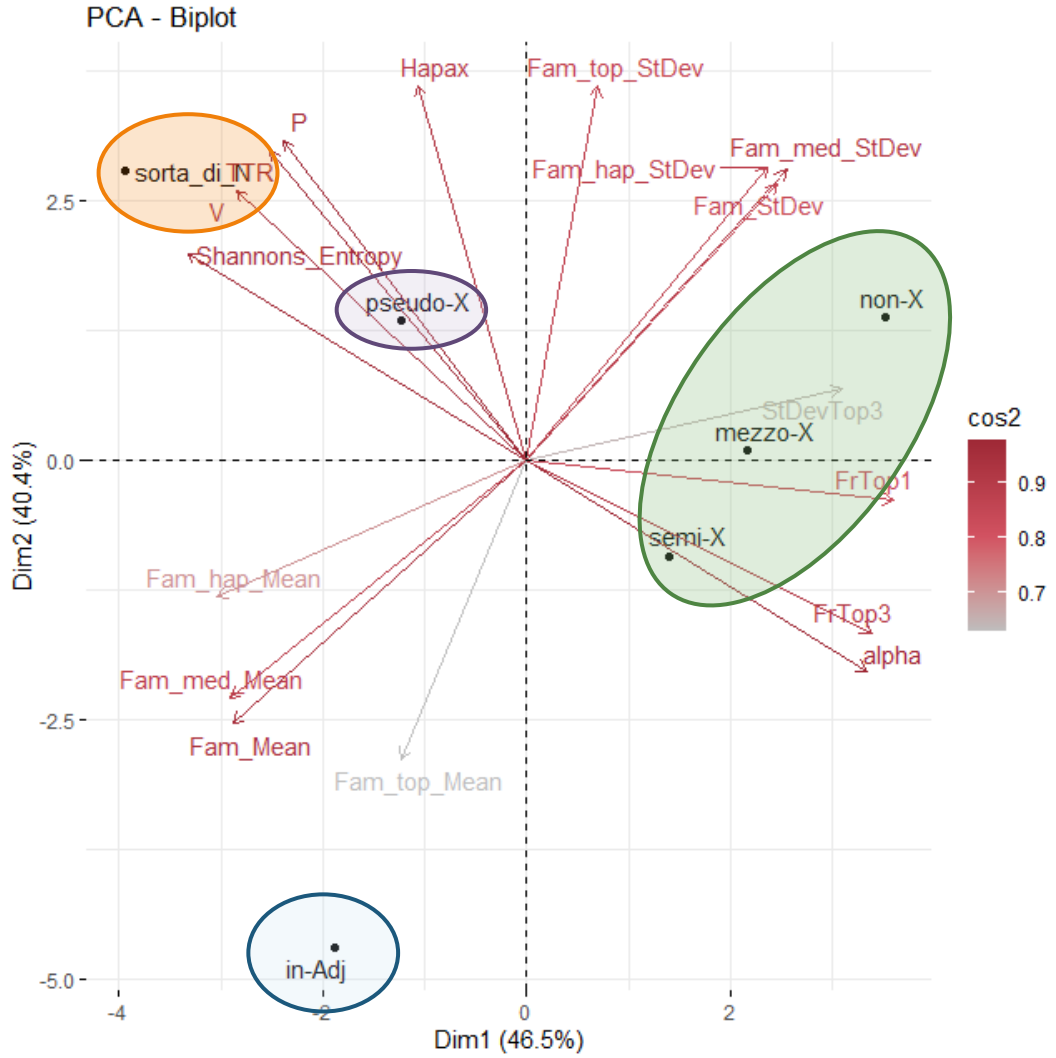
- *in-* is clearly the most familiar cxn
  - All three *in-* stimuli are likely stored items, althought only two exceed Divjak's (2019: 150) threshold for entrenchment (6pmw) (checked in a much larger corpus: itTenTen20)
    - *inutile* (top) 56.85pmw; *ingiusto* (median) 8.39pmw; *irreale* (hapax) 2.64pmw
- *non-* is clearly the least familiar cxn, showing the highest interspeaker variability (see the high StDev)

  ☞ The most 'emergent' one
- The other cxns are in between (*sorta di* N included), with no significant difference in rating among the three evaluative morphological construction

Stimuli types (top-ranked, median, hapax)

- *in-* & *semi-* familiarity reflects their frequency
- This is not so clear for the other cases! (except for *mezzo-*)

Putting things together: PCA in R



***sorta di* X** is rather isolated from everything else

***in*-X** represents another type, also quite distant from other cnxs

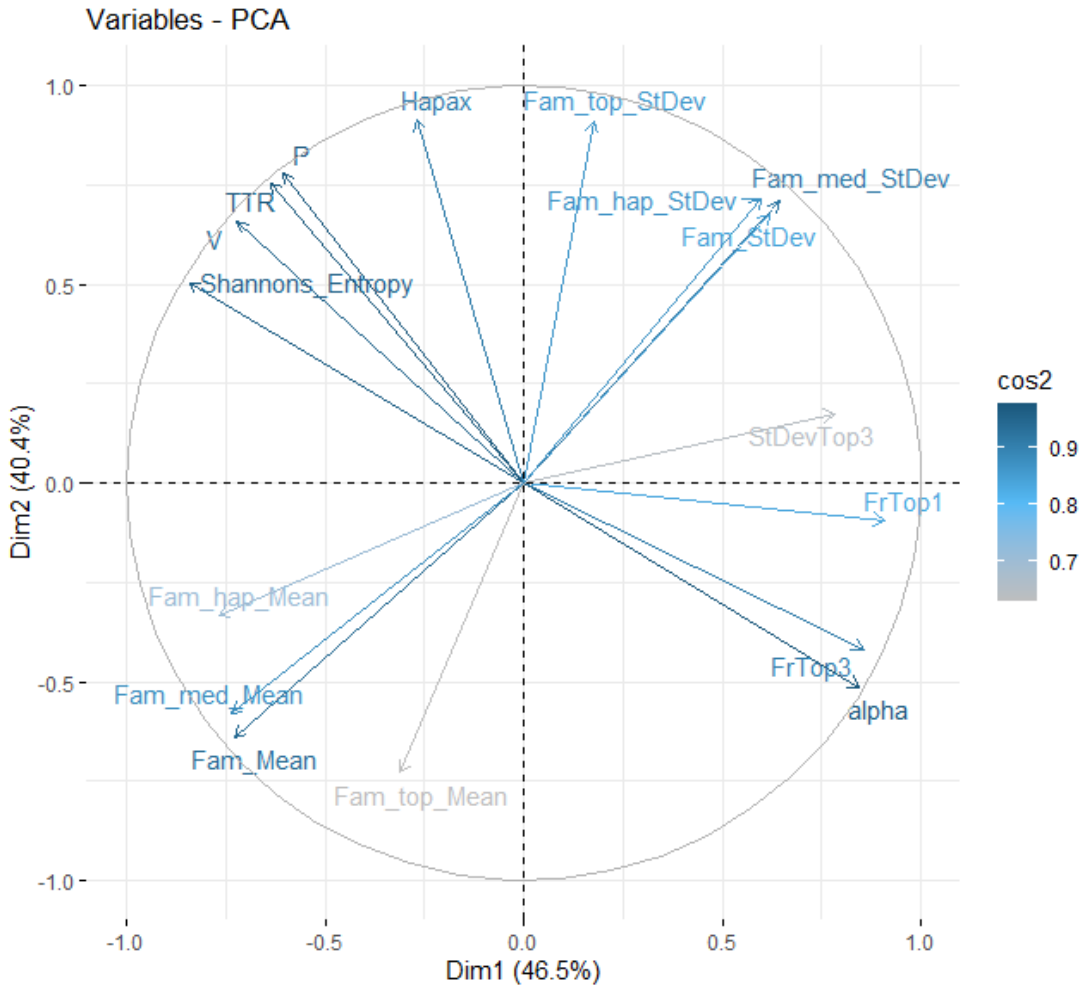
Morphological cxns are a third cluster

- ***mezzo*-X** and ***semi*-X** are very close, whereas ***non*-X** is a bit detached

***pseudo*-X** falls in between morphology and syntax (cfr. Van Goethem et al. 2024 on its debonding, adjective-like behavior)

RQ1) verify if evaluative morphological schemas actually pattern with syntactic schemas in terms of productivity

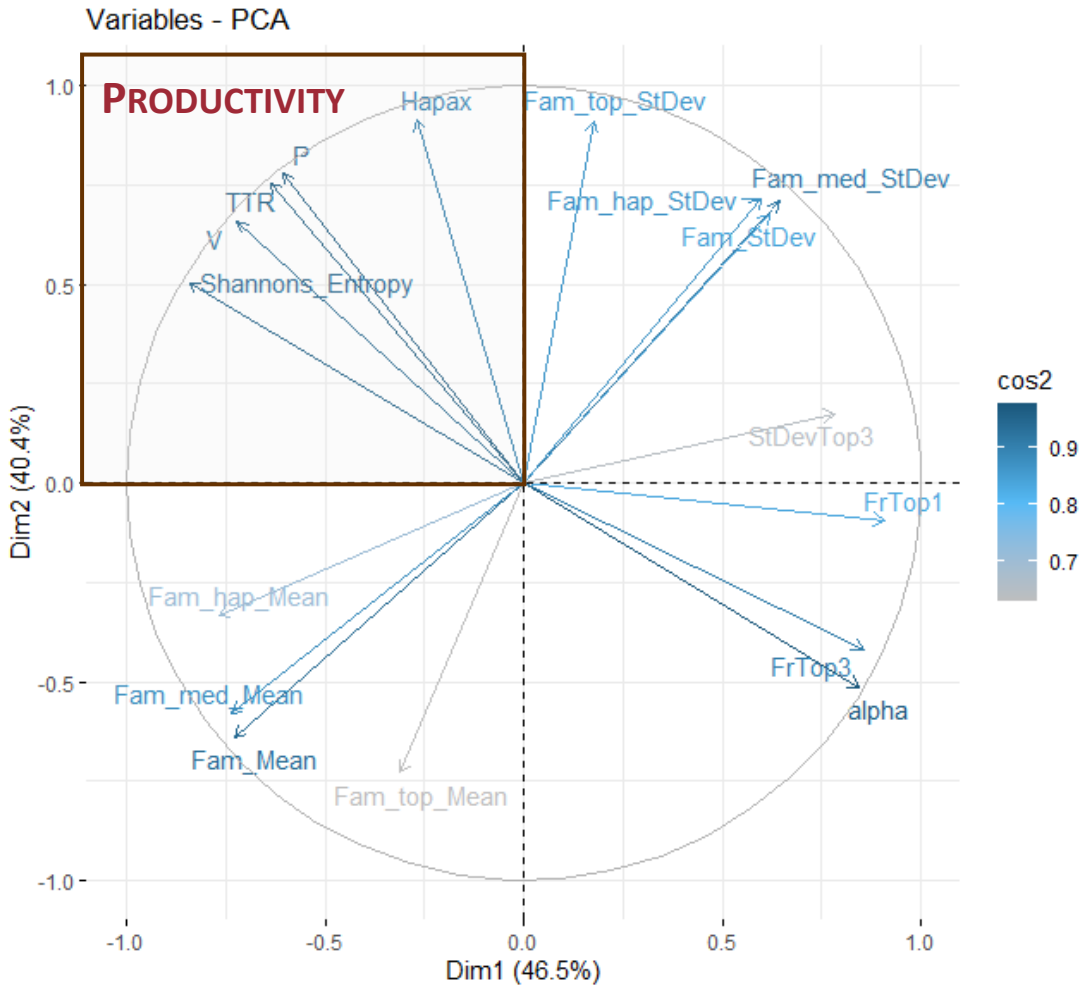
Putting things together: PCA in R



- We put together different types of information

RQ2) understand the relationship between productivity, entrenchment and familiarity in different types of cxns

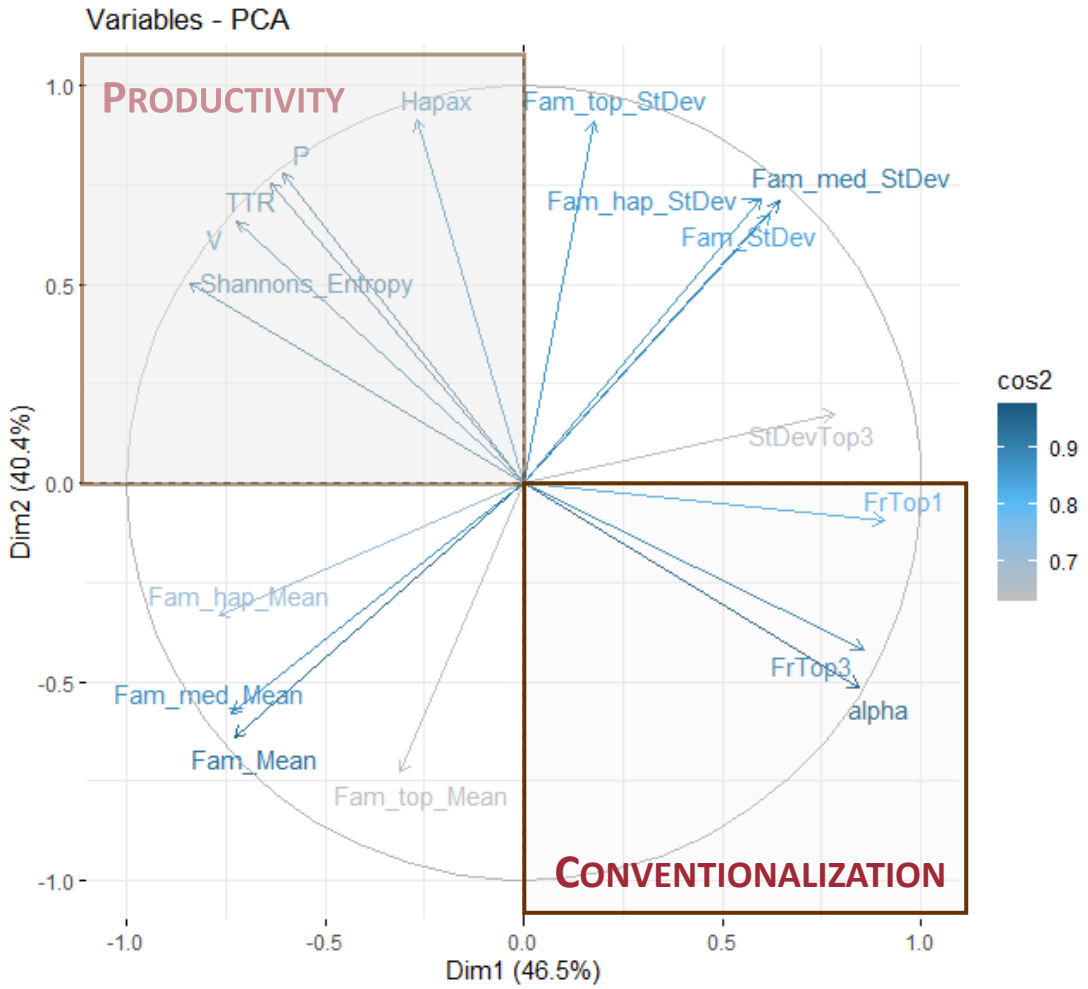
Putting things together: PCA in R



- We put together different types of information
  - Productivity measures (top left) correlate among each other

RQ2) understand the relationship between productivity, entrenchment and familiarity in different types of cxns

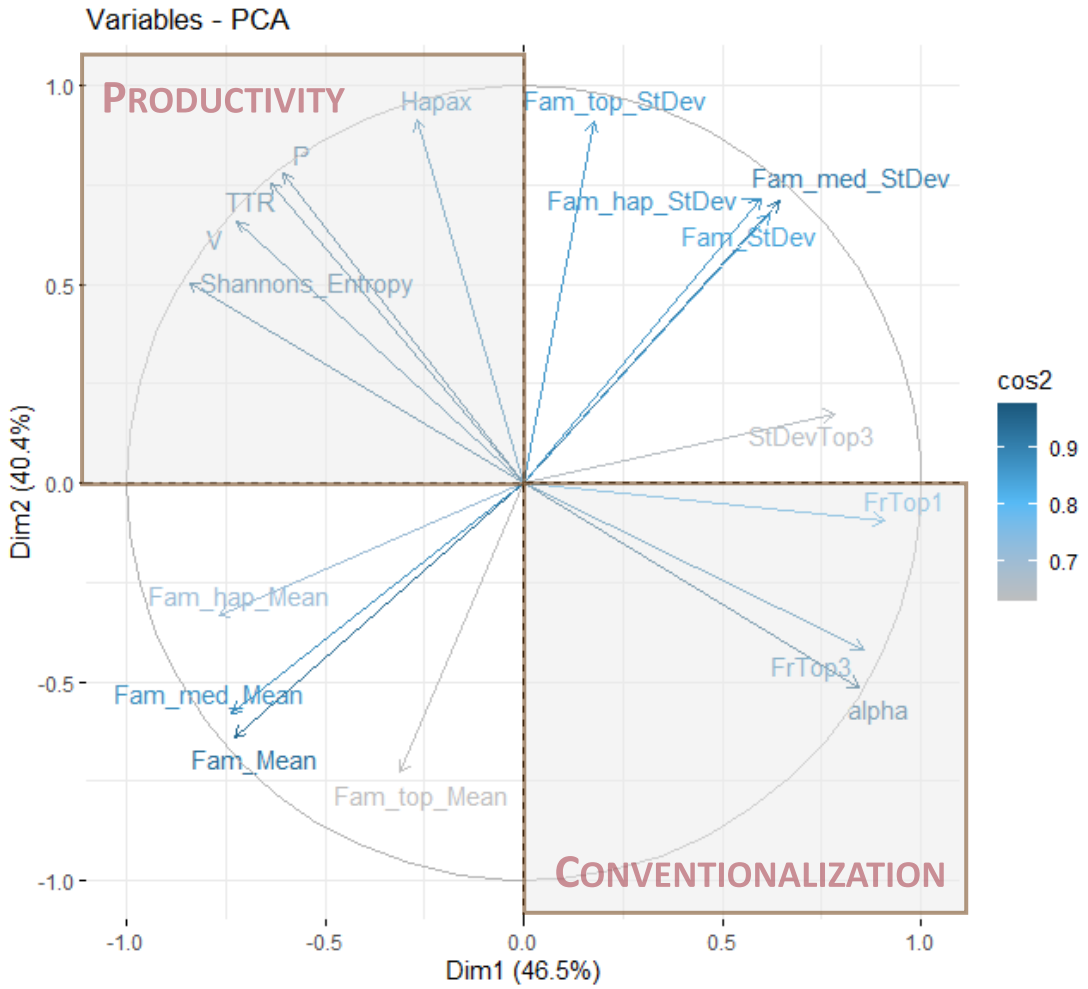
Putting things together: PCA in R



- We put together different types of information
  - Productivity measures (top left) correlate among each other
  - Conventionalization measures (bottom right) correlate among each other

RQ2) understand the relationship between productivity, entrenchment and familiarity in different types of cxns

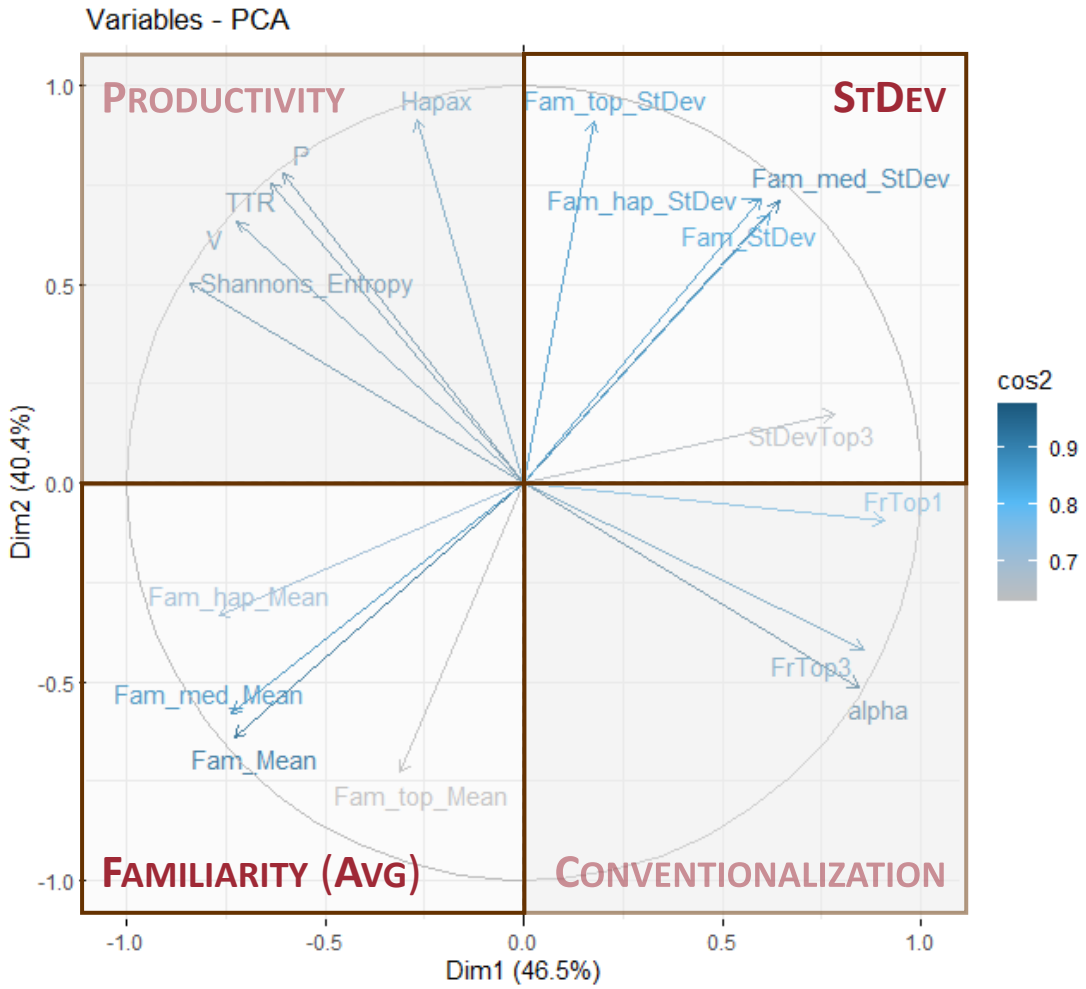
Putting things together: PCA in R



- We put together different types of information
  - Productivity measures (top left) correlate among each other
  - Conventionalization measures (bottom right) correlate among each other
- So, our study **confirms** the results by Van den Heede & Lauwers (2023)

RQ2) understand the relationship between productivity, entrenchment and familiarity in different types of cxns

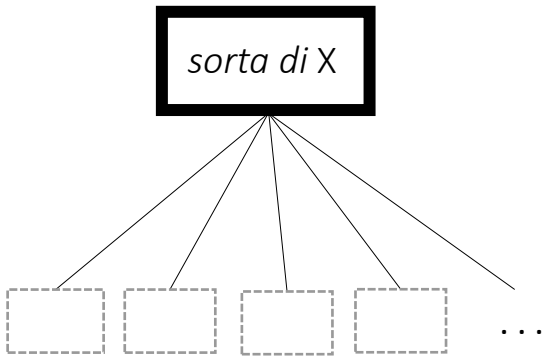
Putting things together: PCA in R



- We put together different types of information
  - Productivity measures (top left) correlate among each other
  - Conventionalization measures (bottom right) correlate among each other
- So, our study **confirms** the results by Van den Heede & Lauwers (2023)
- But we also add **familiarity**, which doesn't correlate with the other measures
  - ☞ Familiarity as another *dimension* of entrenchment

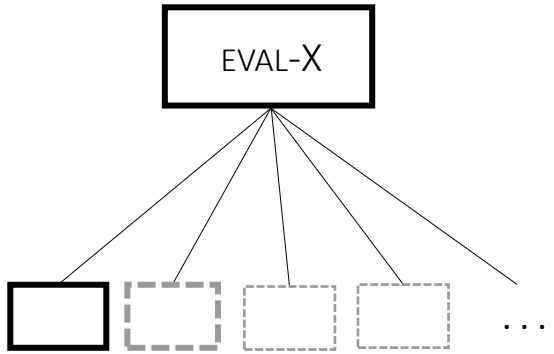
RQ2) understand the relationship between productivity, entrenchment and familiarity in different types of cxns

There is daughter and daughter…



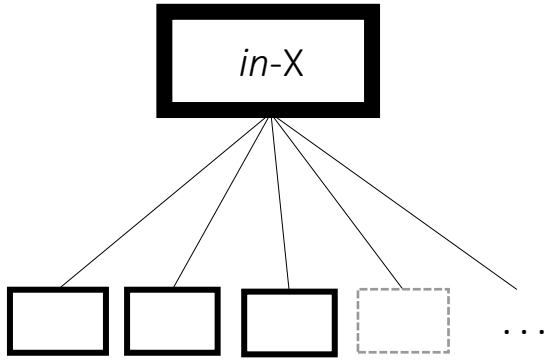
Syntactic cxns

- Very high number of daughters
- Daughters are constructs
- Daughters are homogeneous (low alpha)
- **Familiarity is medium-high**



Evaluative morphological cxns

- Medium number of daughters
- Daughters are mostly constructs
- Daughters are not homogeneous (high alpha)
- **Familiarity is medium-low**

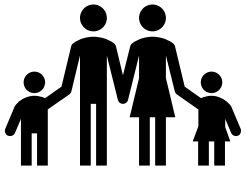


Derivational cxns

- Lower number of daughters
- Daughters are mostly stored
- Daughters are not very homogeneous
- **Familiarity is high**

RQ3) clarify how all this can be modelled in CxG, focusing on the role of networks and daughters

Family-arity matters



- Adding to the picture **familiarity** and '**family**' (namely, the type of **network** the cxns develop, the nature of the **daughters**, and the **relations** among them) allows to grasp evaluatives' behavior better
  - Familiarity can be high even if productivity scores are low (derivation), especially if the daughters are (stored) constructions that strengthen the entrenchment of the mother cxn
  - Familiarity can be rather high even if frequency is low (*sorta di* X expressions): in this case familiarity cannot be a correlate of the frequency/entrenchment of the single items but may hint at the entrenchment of the mother cxn that licences them (and perhaps of the acceptability of the filler)
  - The medium-low familiarity of evaluative morphological cxns may, instead, hint at a lower entrenchment of the mother cxn (see StDev), which yields a rather large number of types with unequal status (mostly constructs plus some conventionalized items)

	Productivity measures	**Conventionalization measures**	**Familiarity rating**	**Nature of daughters**
SYNTACTIC_EVALUATIVE	HIGH	LOW	MEDIUM-HIGH	CONSTRUCTS
MORPHOLOGICAL_EVALUATIVE	MEDIUM	HIGH	MEDIUM-LOW	MOSTLY CONSTRUCTS
DERIVATIONAL_MORPHOLOGY	LOW	MEDIUM	HIGH	CONSTRUCTIONS

Conclusions

- In order to unravel productivity (as a multi-dimensional concept) we need to take into account also **construction-external factors** (like type of network/family and daughters)
- Different methods and **different types of evidence** (corpus data, behavioral data, etc.) may be necessary

- **Limits and future steps**
  - Small sample size ☞ use larger sample?
    - Hapax-based measures less reliable (e.g., *in*- hapaxes employed in the familiarity test are not hapaxes in the whole corpus)?
  - Combine this picture with information about the bases (Xs): PoS, semantic sparsity, etc.
  - Take into account possible horizontal links between competing cxns (e.g., *mezzo*-X and *semi*-X)

Thank you for your attention!

Comments welcolme

francesca.masini@unibo.it

fpisciotta@unisa.it

References (selected)

Albair, Joshua. 2010. What is the state of evaluative affixes in contemporary English? *Début: the undergraduate journal of languages, linguistics and area studies 1*(1), 1-16.

Baayen, Harald R. 1991. Quantitative aspects of morphological productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1991*, 109-149. Dordrecht: Kluwer Academic Publishers.

Barðdal, Johanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.

Bauer, Laurie. 1997. Evaluative morphology: in search of universals. *Studies in Language 21*(3), 533-575.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Caldwell-Harris, Catherine, Jonathan Berant & Shimon Edelman. 2012. Measuring mental entrenchment of phrases with perceptual identification, familiarity ratings, and corpus frequency statistics. In Dagmar Divjak & Stefan Th. Gries (eds.), *Frequency effects in language representation*, 165-194. Berlin: De Gruyter.

Gaeta, Livio & Davide Ricca. 2003. Italian prefixes and productivity: a quantitative approach. *Acta Linguistica Hungarica 50*(1-2), 93-112.

Masini, Francesca & Silvia Micheli. 2023. The morphological expression of approximation: the emerging simil- construction in Italian. *Word Structure 13*(3). 371–402.

Masini, Francesca, Muriel Norde & Kristel Van Goethem (eds.) 2023. Approximation in Morphology. *Zeitschrift für Wortbildung / Journal of Word Formation*, Special Issue 7(1).

Pankratz, Elizabeth, Titus von der Malsburg, & Shravan Vasishth. 2022. Shannon entropy is a more comprehensive and principled morphological productivity measure than the standard alternatives. *PsyArXiv*. doi: 10.31234/osf.io/vkyug [last accessed: 01/12/2023]

Perek, Florent. 2020. Productivity and schematicity in constructional change. In Lotte Sommerer & Elena Smirnova (eds.), *Node and networks in Diachronic Construction Grammar*, 141-166. Amsterdam: John Benjamins.

Rossini Favretti, Rema, Fabio Tamburini & Cristiana De Santis. 2002. CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In Andrew Wilson, Paul Rayson & Tony McEnery (eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, 27-38. Munich: LINCOM-Europa.

Van den Heede, Margot & Lauwers, Peter. 2023. Syntactic productivity under the microscope: the lexical and semantic openness of Dutch minimizing constructions. *Folia Linguistica 57*(3). 723-761.

Van Goethem, Kristel, Muriel Norde & Francesca Masini (2023). The fate of ‘pseudo-’ words: a contrastive corpus-based analysis. *Languages in contrast*. Online First.

Zeldes, Amir. 2012. *Productivity in Argument Selection. From Morphology to Syntax*. Berlin: De Gruyter.

Ratings: variation by Age

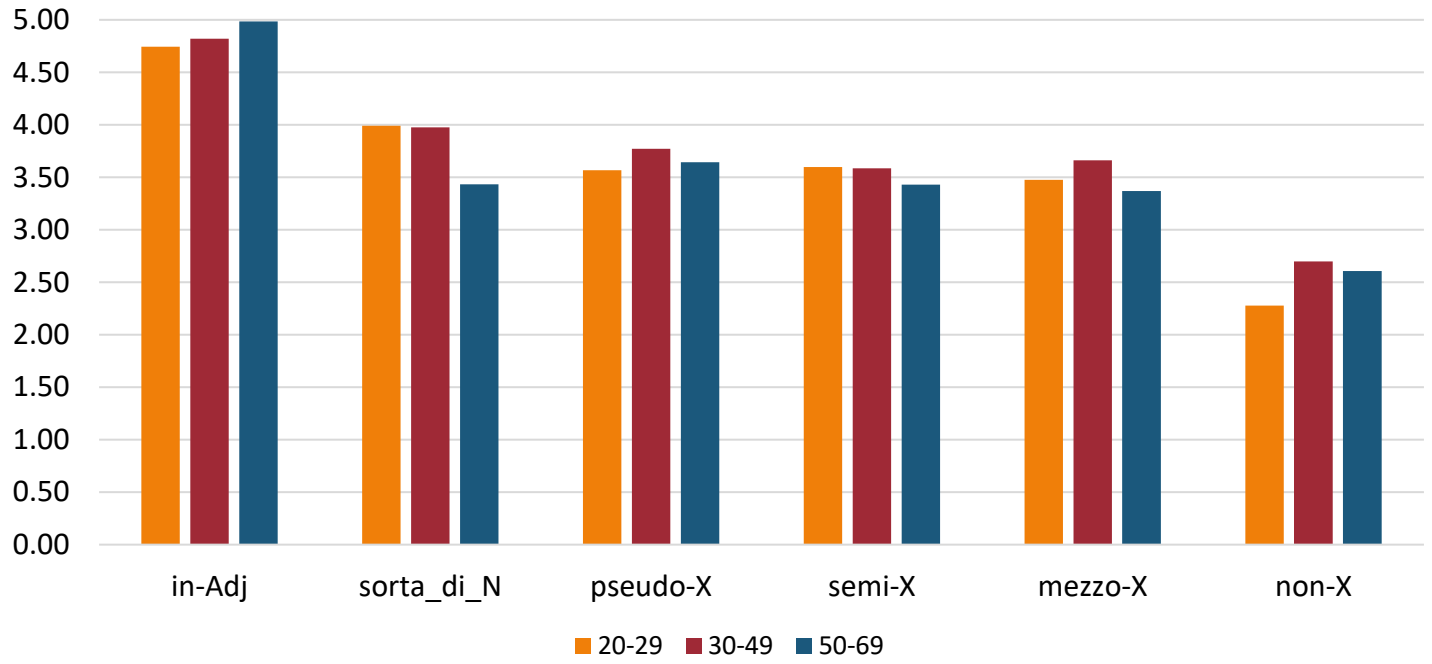
AgeGroup	n	AvgRating	StDev
20-29	88	3.38	1.83
30-49	87	3.49	1.78
50-69	25	3.38	1.93

Average Ratings of Cxns by Age

	20-29	30-49	50-69	Overall
in-Adj	4.74	4.82	4.98	4.81
sorta_di_N	3.99	3.98	3.43	3.91
pseudo-X	3.57	3.77	3.64	3.66
semi-X	3.60	3.59	3.43	3.57
mezzo-X	3.47	3.66	3.37	3.54
non-X	2.28	2.70	2.61	2.50

significantly higher

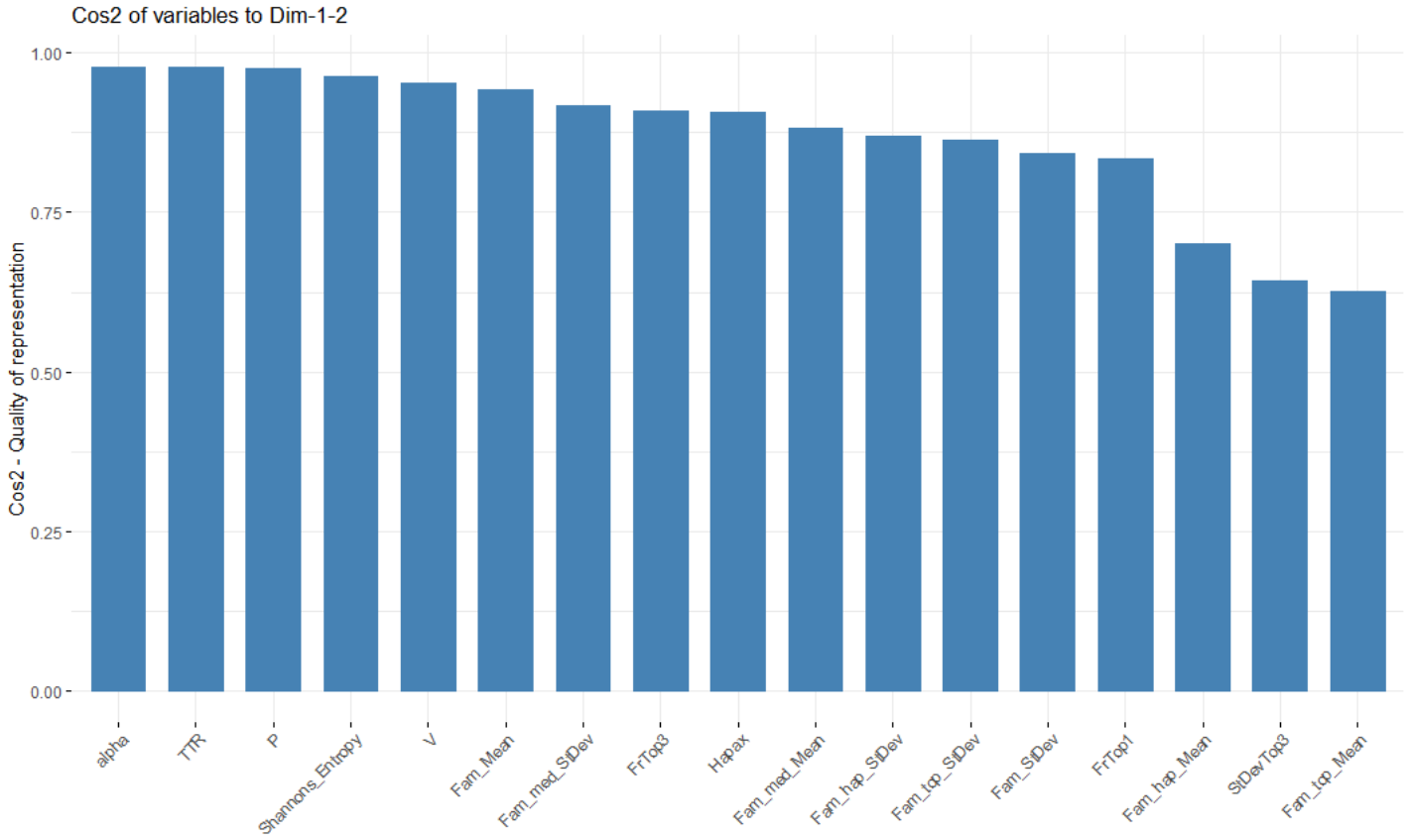
significantly lower



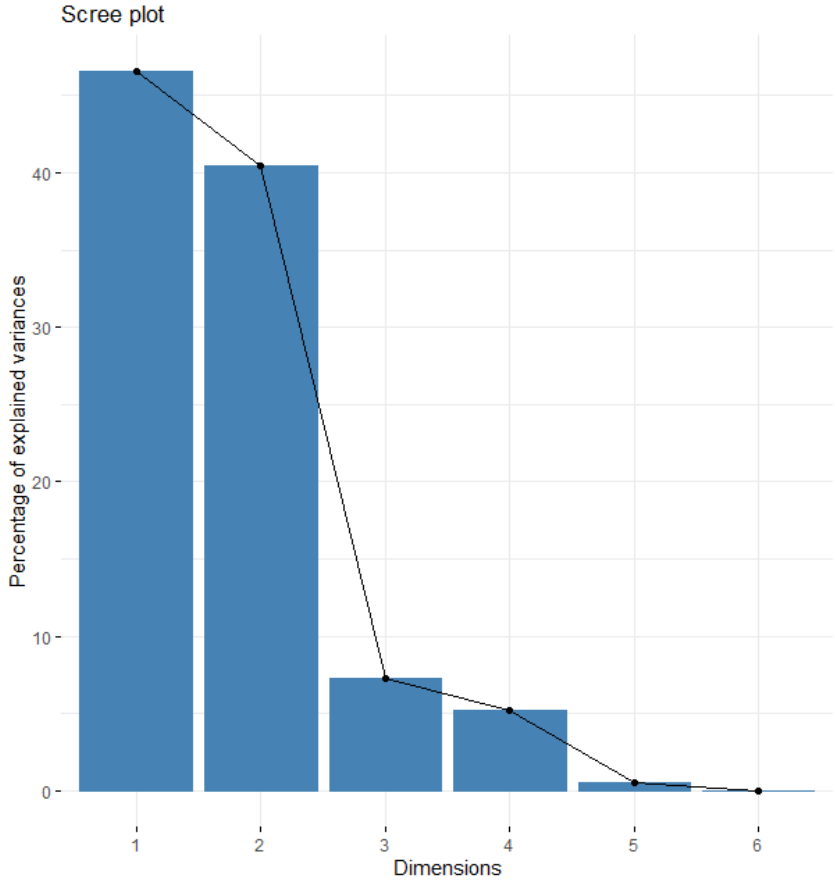
PCA loadings

	PC1 (46.54%)	PC2 (40.40%)	PC3 (7.31%)	PC4 (5.21%)	PC5 (0.53%)	PC6 (<0.0001)
V	-0.25666	0.250754	0.009522	0.221193	0.194566	0.210434
TTR	-0.22665	0.288138	0.100807	0.109712	-0.00376	-0.38413
P	-0.21555	0.297509	0.023314	0.154925	-0.16676	-0.10381
Hapax	-0.09622	0.348404	-0.16771	0.254764	0.079498	0.006849
Shannons_Entropy	-0.29926	0.192468	0.095629	0.128166	0.349214	0.311923
alpha	0.300338	-0.19615	-0.06922	-0.01987	0.416493	-0.25287
FrTop1	0.323072	-0.03602	0.265914	0.291507	0.150793	0.117125
FrTop3	0.304263	-0.15984	-0.04586	0.306177	0.26427	0.195159
StDevTop3	0.278146	0.066877	0.491574	0.252233	0.075379	-0.0275
Fam_Mean	-0.25944	-0.24433	-0.04436	0.243278	0.166468	0.034846
Fam_top_Mean	-0.11019	-0.27795	-0.32791	0.519582	-0.07187	-0.4804
Fam_med_Mean	-0.26224	-0.22139	-0.28371	-0.04151	0.44147	0.077216
Fam_hap_Mean	-0.27315	-0.12705	0.464451	0.185773	0.045497	0.050287
Fam_StDev	0.220598	0.257907	-0.34821	0.083581	0.118443	0.102755
Fam_top_StDev	0.062955	0.347764	0.086211	-0.34314	0.519073	-0.41203
Fam_med_StDev	0.229168	0.270361	-0.01059	0.302648	-0.1105	-0.22804
Fam_hap_StDev	0.212094	0.273314	-0.31336	0.096211	-0.09383	0.33274

Cos2 Variables



Explained Variance



Scores

	PC1 (46.54%)	PC2 (40.40%)	PC3 (7.31%)	PC4 (5.21%)	PC5 (0.53%)	PC6 (<0.0001)
sorta_di_N	-3.937882	2.78805332	-0.09522624	0.1508509	-0.30729150	-2.359224e-16
in-Adj	-1.891447	-4.69042198	0.71836436	-0.2095662	-0.06488900	1.443290e-15
semi-X	1.389472	-0.91880606	-2.13282474	-0.3311607	-0.03658137	-5.773160e-15
non-X	3.509635	1.37957546	1.00109968	-1.0210186	-0.17806036	-2.775558e-16
mezzo-X	2.167598	0.09573847	0.31221284	1.7583098	0.02028296	6.827872e-15
pseudo-X	-1.237375	1.34586079	0.19637411	-0.3474151	0.56653926	2.241263e-15

Familiarity Ratings ~ Corpus frequency (CORIS)

cxn	freqType	stimulus	n	pmw	LogFreq	AvgRating	StDev
sorta_di_N	top	sorta di museo	12	0.07	1.079181	3.87	1.33
sorta_di_N	median	sorta di villaggio	4	0.02	0.60206	3.99	1.22
sorta_di_N	hapax	sorta di scatola	13	0.08	1.113943	3.88	1.39
in-Adj	top	inutile	10363	62.96	4.015485	4.97	0.18
in-Adj	median	ingiusto	1885	11.45	3.275311	4.92	0.43
in-Adj	hapax	irreale	791	4.81	2.898176	4.53	0.94
semi-X	top	semi-nudo	274	1.66	2.437751	4.54	0.89
semi-X	median	semi-professionale	18	0.11	1.255273	4.12	1.25
semi-X	hapax	semi-magico	2	0.01	0.30103	2.05	1.55
mezzo-X	top	mezzo-sorriso	276	1.67	2.440909	4.47	0.85
mezzo-X	median	mezzo-spento	2	0.01	0.30103	3.03	1.74
mezzo-X	hapax	mezza-notizia	7	0.04	0.845098	3.13	1.52
pseudo-X	top	pseudo-religioso	20	0.12	1.30103	3.62	1.50
pseudo-X	median	pseudo-artista	2	0.01	0.30103	3.93	1.21
pseudo-X	hapax	pseudo-informazione	2	0.01	0.30103	3.45	1.38
non-X	top	non-vita	16	0.10	1.20412	2.75	1.62
non-X	median	non-lavoro	8	0.05	0.90309	2.29	1.66
non-X	hapax	non-regola	1	0.01	0	2.46	1.56