

A gem of a construction. Exploring cross-linguistic differences with corpus data

Isa Hendriks, Valentina Piunno, Kristel Van Goethem

8 October 2025

Master course “Digital linguistics”

University of Bergamo

Outline

1. The Evaluative Binominal Construction: State of the art
2. Research goal and questions
3. Data and method of analysis
4. Corpus results
5. Discussion and conclusions

The Evaluative Binominal Construction (EBC)

- **Examples:**

- DU *een schat van een kind* ‘a lovely child’ (lit. a treasure of a child)
- EN *a hell of a day*
- FR *cette saleté de médicament* ‘this dirty medication’ (lit. this dirt of medication)
- IT *quel genio di bambino* ‘that smart child’ (lit. that genius of child)

- **Structure:** [DET1 N1 PREP (DET2) N2]

- Semi-schematic construction
 - Only the preposition is fixed
 - All other slots are subject to variation

- **Non-compositional meaning:**

- “N2 showing to an extraordinary degree characteristic features of N1 (subjective-evaluative)” (translated from Verhagen 2005: 202)
 - N1 → extraordinary (physical or abstract) quality (often metaphorical value)
 - N2 → entity to which this quality is attributed (referential head)

State of the art

- **Studies on single languages:**

- Dutch (a.o. Paardekooper 1956, Verhagen 2005, **Van Goethem 2023**)
- English (a.o. Aarts 1994, 1998, **ten Wolde 2023**)
- French (a.o. Regula 1965, Larrivée 1994)
- Italian (Alinei 1971)
- Arabic (Vallaro 1993)

- **Cross-linguistic studies:**

- Germanic and Romance languages (a.o. Foolen 2004, Simone in press)
- Romance languages (a.o. Tutescu 1969, **Piunno 2025**)

→ red color = corpus-based studies

→ Corpus-based cross-linguistic studies are still scarce

Research goal & questions

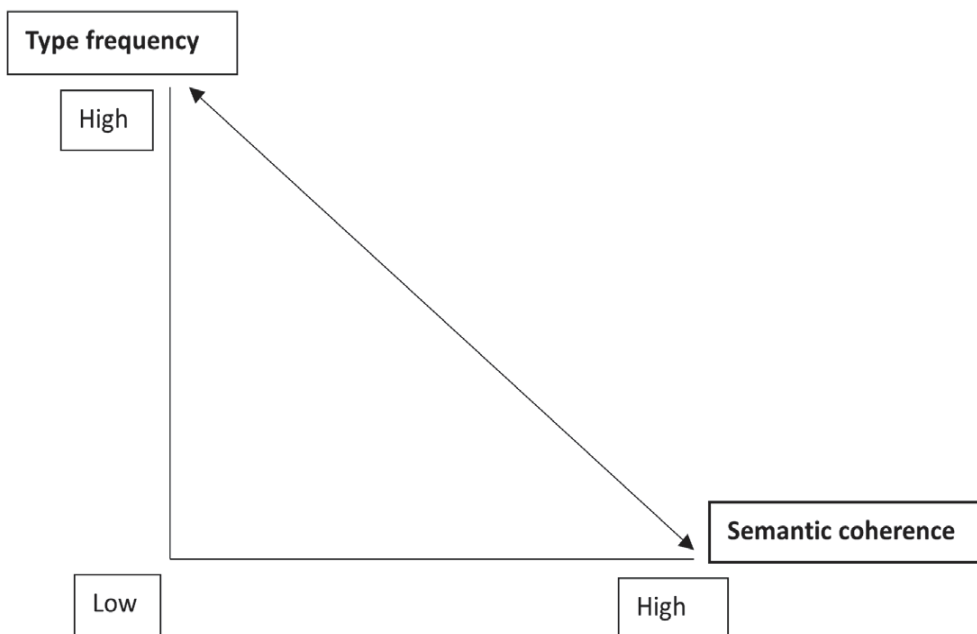
Research goal

- A corpus-based cross-linguistic constructionist analysis of the EBC (Dutch, English, French, Italian)

Research questions

- **RQ1:** What are the main **cross-linguistic similarities and differences** between the EBC in the four languages under study?
- **RQ2:** What is the relation between **productivity** and **semantic coherence** in the EBC in the four languages under study?
 - *semantic coherence*: internal consistency between the members of the schema (Barðdal 2008: 27)

Hypothesis: Inverse correlation between type frequency and semantic coherence



Barðdal (2008:35)

“[...] the higher the type frequency of a construction, the lower the degree of semantic coherence is needed for a construction to be productive. Conversely, the lower the type frequency of a construction, the higher degree of semantic coherence is needed for a construction to be extendable.” (Barðdal 2008:34)

“(...) the categories relatively close to the cline will show most signs of productivity, while categories further away from the cline will be either low/nonproductive (lower left-most sphere) or non-existing (upper right-most sphere).” (Barðdal 2008: 39)

Corpus data

Sketch Engine: <https://www.sketchengine.eu>

- TenTen corpus family
- Corpus data collected from the Internet

two Germanic languages (English and Dutch)

- nlTenTen20 (6,836,979,371 tokens) 2020
- enTenTen21 (52 billion words) 2021

two Romance languages (French and Italian)

- frTenTen23 (23.8 billion words) 2022-2023
- itTenTen20 (14,514,566,714 tokens) 2019-2020

Concordance function – CQL queries

Combination of quantitative and qualitative analysis

Data extraction

Extraction of the **top 15** EBC patterns per language

- General CQL Query:
 - e.g., English [word="a"] [tag = "noun.*"] [word="of"] [word="a"]
- Extract top 15 EBC patterns per language
- General analysis of frequency and semantic properties of top 15

Detailed analysis of the **top 15** EBC micro-constructions per language

- More specific CQL query including N1 lemma
 - e.g., English [tag="det.*"]? [lemma="hell"] [lemma="of"]
 - Italian: [DET 1="quello|quel|questo|quella|quelle|quegli|quei|questi|queste"]
- Random samples of 1000 occurrences per subpattern
- Manual check until 100 relevant occurrences per pattern
- Final datasets of 15x 100 relevant occurrences per language
= **6000 relevant occurrences in total**

Data annotation and analysis

- **Annotation in Excel in terms of:**
 - **Formal properties** (Det 1, Det 2, number, complexity N2)
 - **Evaluative semantics construction:** ameliorative, pejorative, augmentative, diminutive (or combination)
 - **Semantics N2:** person, animal, object, place, concept
 - **Lemma N2** ~ productivity scores: TTR, PP

RESULTS:

1. Cross-linguistic similarities and differences

Frequency: top 15

	Dutch	English	French	Italian
1	<i>een dijk van een</i>	<i>a hell of a</i>	<i>ce/cette putain de</i>	<i>questo fenomeno di</i>
2	<i>een pracht van een</i>	<i>a heck of a</i>	<i>ce/cette drôle de</i>	<i>quel cazzo di</i>
3	<i>een schat van een</i>	<i>a gem of a</i>	<i>cette saloperie de</i>	<i>questo capolavoro di</i>
4	<i>een joekel van een</i>	<i>a whale of a</i>	<i>ce con de</i>	<i>quel genio di</i>
5	<i>een draak van een</i>	<i>a beast of a</i>	<i>cette saleté de</i>	<i>quel cavolo di</i>
6	<i>een beest van een</i>	<i>a monster of a</i>	<i>ce monstre de</i>	<i>questo schifo di</i>
7	<i>een juweel van een</i>	<i>a devil of a</i>	<i>ce diable de</i>	<i>questo gioiello di</i>
8	<i>een kanjer van een</i>	<i>a mess of a</i>	<i>ce salaud de</i>	<i>questa meraviglia di</i>
9	<i>een droom van een</i>	<i>a cracker of a</i>	<i>ce connard de</i>	<i>quel maiale di</i>
10	<i>een parel van een</i>	<i>a steal of a</i>	<i>ce bijou de</i>	<i>quello porco di</i>
11	<i>een knaller van een</i>	<i>a rollercoaster of a</i>	<i>cet imbécile de</i>	<i>quel pazzo di</i>
12	<i>een dot van een</i>	<i>a bitch of a</i>	<i>cet idiot de</i>	<i>quell' idiota di</i>
13	<i>een boom van een</i>	<i>a rocket of a</i>	<i>cet abruti de</i>	<i>quella merda di</i>
14	<i>een beer van een</i>	<i>a brute of a</i>	<i>cette merde de</i>	<i>quello stronzo di</i>
15	<i>een kast van een</i>	<i>a fuck of a</i>	<i>cet enfoiré de</i>	<i>quel diavolo di</i>

Top 15: Semantics cxn

Dutch	English	French	Italian
een dijk van een	a hell of a	ce/cette putain de	quel fenomeno di
een pracht van een	a heck of a	ce/cette drôle de	quel cazzo di
een schat van een	a gem of a	cette saloperie de	quel capolavoro di
een joekel van een	a whale of a	ce con de	quel genio di
een draak van een	a beast of a	cette saleté de	quel cavolo di
een beest van een	a monster of a	ce monstre de	questo schifo di
een juweel van een	a devil of a	ce diable de	questo gioiello di
een kanjer van een	a mess of a	ce salaud de	questa meraviglia di
een droom van een	a cracker of a	ce connard de	quel maiale di
een parel van een	a steal of a	ce bijou de	quel porco di
een knaller van een	a rollercoaster of a	cet imbécile de	quel pazzo di
een dot van een	a bitch of a	cet idiot de	quell'idiota di
een boom van een	a rocket of a	cet abruti de	quella merda di
een beer van een	a brute of a	cette merde de	quello stronzo di
een kast van een	a fuck of a	cet enfoiré de	quel diavolo di

Ameliorative

Pejorative

Augmentative

Diminutive

Evaluative semantics of the construction as a whole

Dutch

- Mostly **ameliorative** and **augmentative** connotations
 - *een **pracht** van een kerstkaart* 'a beauty of a Christmas card'
 - *een **joekel** van een auto* 'a huge car' (lit. a whopper of a car)

English

- Mix of **pejorative** and **ameliorative** values:
 - *a **hell** of a job*
 - *a **gem** of a book*

French / Italian

- Clearly more **negative** connotations + **informal** language
 - FR *la **putain** de patrie* 'that fucking fatherland'
 - IT *quello **schifo** di città* 'that disgusting city'
- **Mirative** value of *drôle* (expressing surprise)
 - FR *Quelle **drôle** d'idée* 'what a crazy idea'

Metaphorical values of N1

Dutch / English

- Mostly (concrete/abstract) N1 with metaphorical value
 - DU een **dijk** van een stem (lit. a dike of a voice) 'a powerful voice'
 - ENG a **rollercoaster** of a journey

French / Italian

- Fewer cases of N1 with metaphorical value in the Romance languages
- Often nominalized adjective in N1-slot
 - FR quel(le) **imbécile** de N
 - IT quell'**idiota** di N

Formal properties: general tendencies

DET 1

- Wide variation in DET1 position in Germanic:
 - DU ***vier*** *schatten van kleinkinderen* ‘four darlings of grandchildren’
 - ENG ***my*** *hell of a house*
- Preference for a demonstrative/exclamative in Romance
 - FR ***ces*** *saletés de laveurs de vitres* ‘those filthy window cleaners’
 - IT ***quella*** *meraviglia di bambino* ‘that lovely child’

DET 2

- Indefinite article or absent in Germanic (e.g., *an absolute whale of time*)
- Less frequently present in Romance

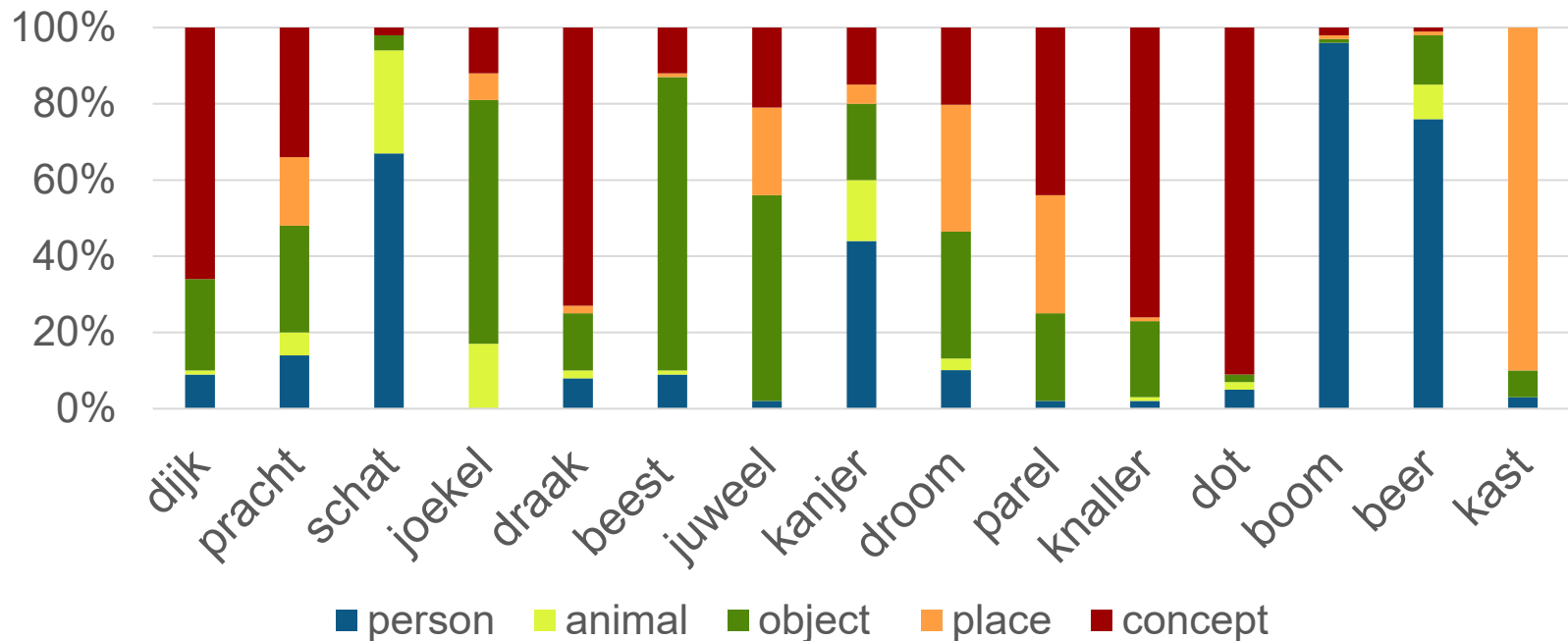
N2

- Not only simplex nouns, also compounds and NP
 - FR *ces saloperies de ***toilettes turques**** ‘these dirty Turkish toilets’
 - DU *een schat van een ***Engelse nanny**** ‘a sweetheart of an English nanny’

RESULTS:

2. Relation between productivity and semantic coherence

Dutch: Semantics N2



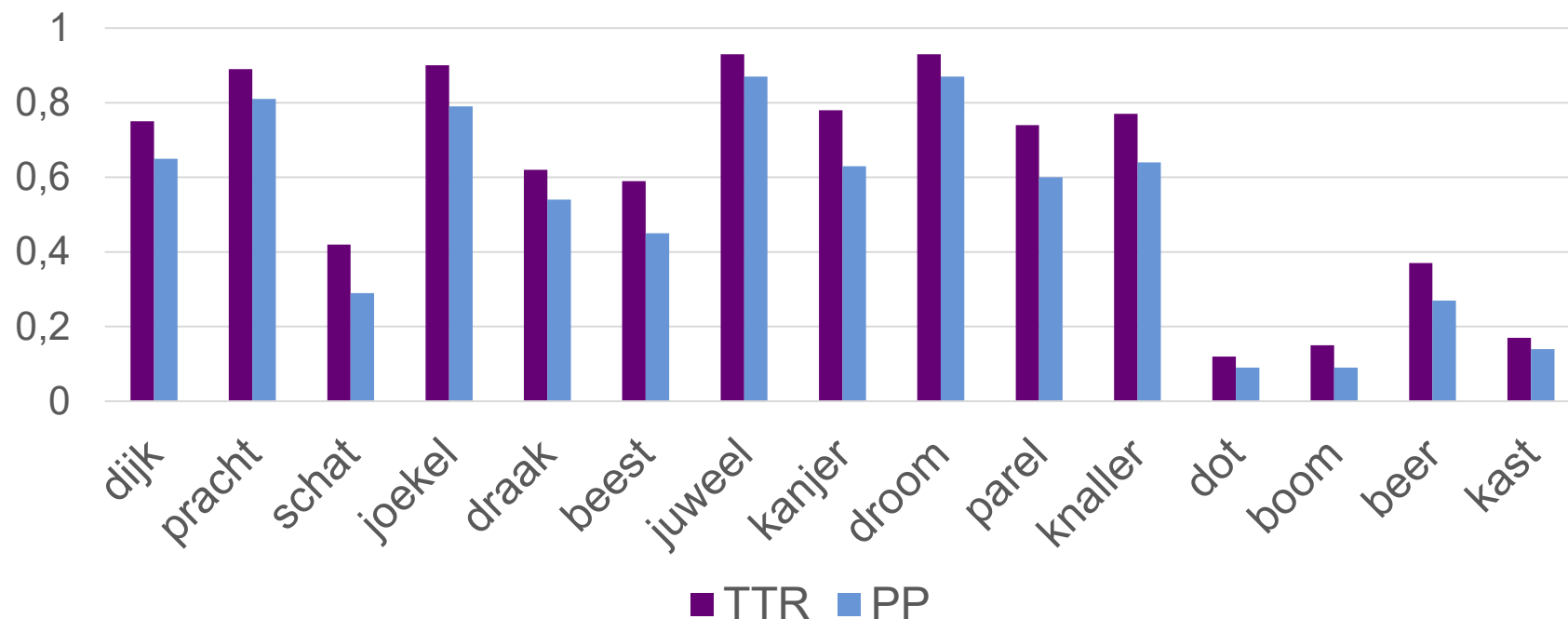
Examples with semantic variation

- *een **pracht** van een **verhaal*** (n=4) (lit. a beauty of a story) 'a beautiful story'
- *een **kanjer** van een **boek*** (n=5) (lit. a whopper of a book) 'a great book'
- *een **droom** van een **vakantie*** 'a dream of a holiday'

Examples with semantic specialization

- *een **dot** van een **kans*** (n=86) (lit. a blob of a chance) 'a golden opportunity'
- *een **boom** van een **kerel*** (n=41), ***vent*** (n=39) (lit. a tree of a man) 'a big guy'
- *een **kast** van een **huis*** (n=80) 'a mansion of a house, a huge house'

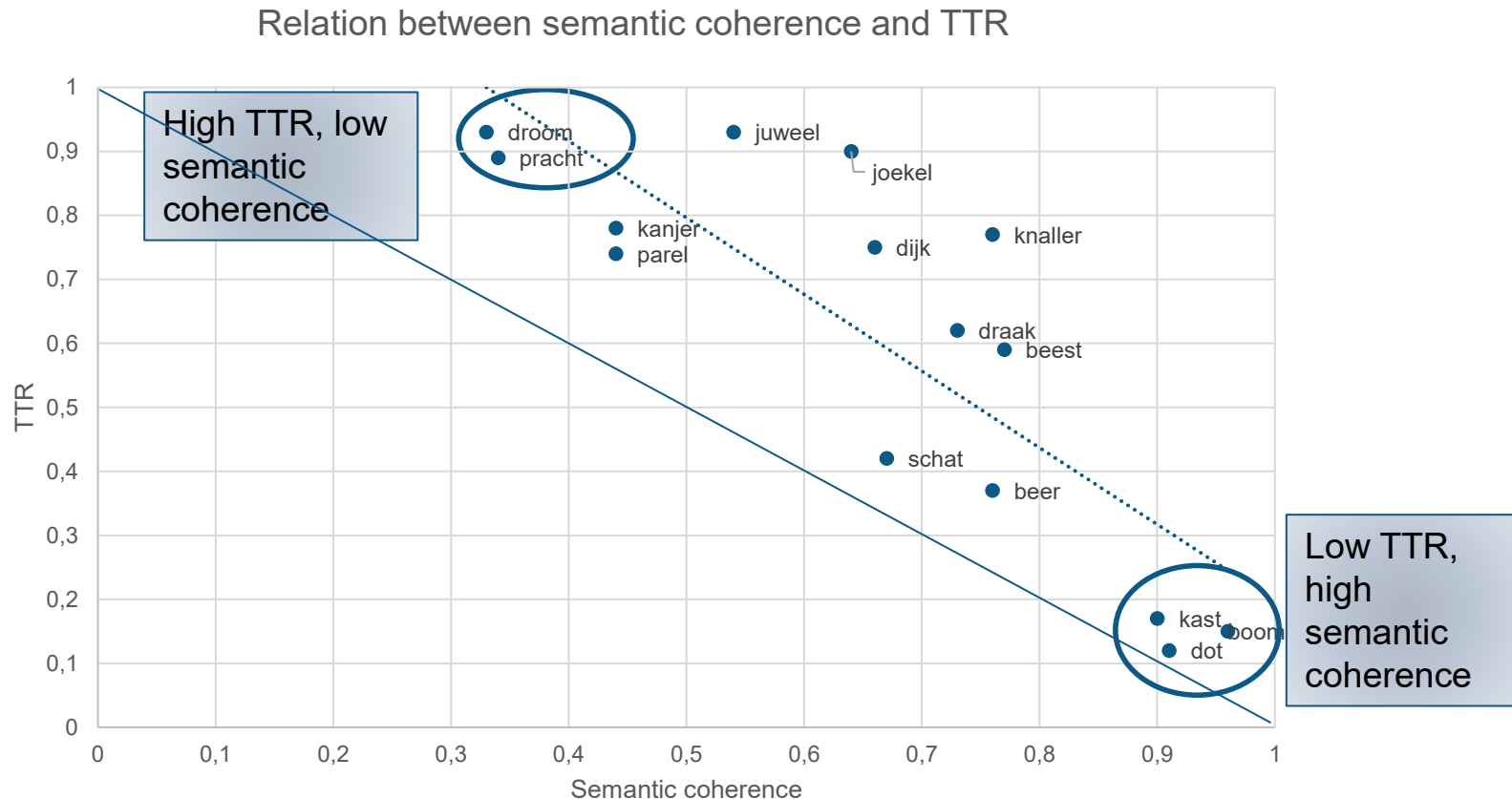
Dutch: Productivity scores



High productivity (hapaxes)

- *een joekel van een courgette* (lit. a whopper of a zucchini) ‘a giant zucchini’
- *een juweel van een doelpunt* ‘a gem of a goal’
- *een droom van een vrijgezellenfeest* ‘a dream of a bachelor party’

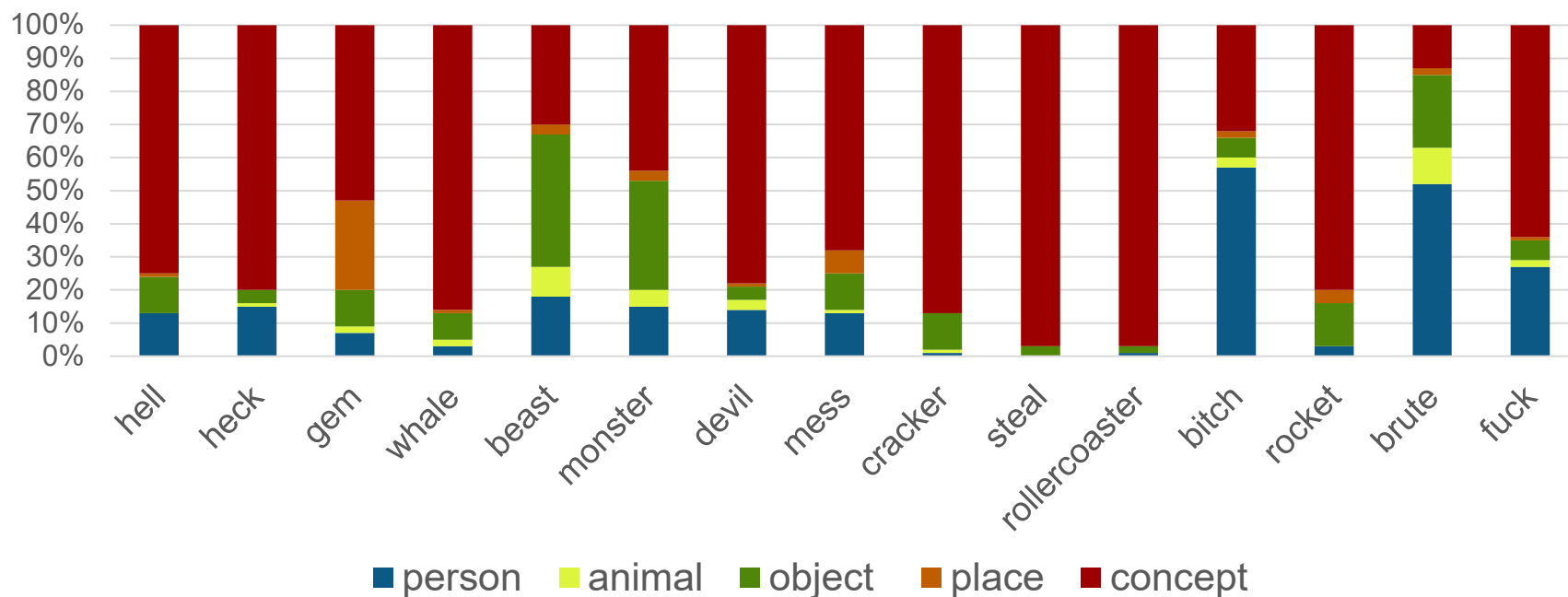
Dutch: Relation between productivity and semantic coherence



Spearman $r_s = -0.85$ ($p < .01$)

→ Strong negative relationship between TTR and semantic coherence

English: Semantics N2

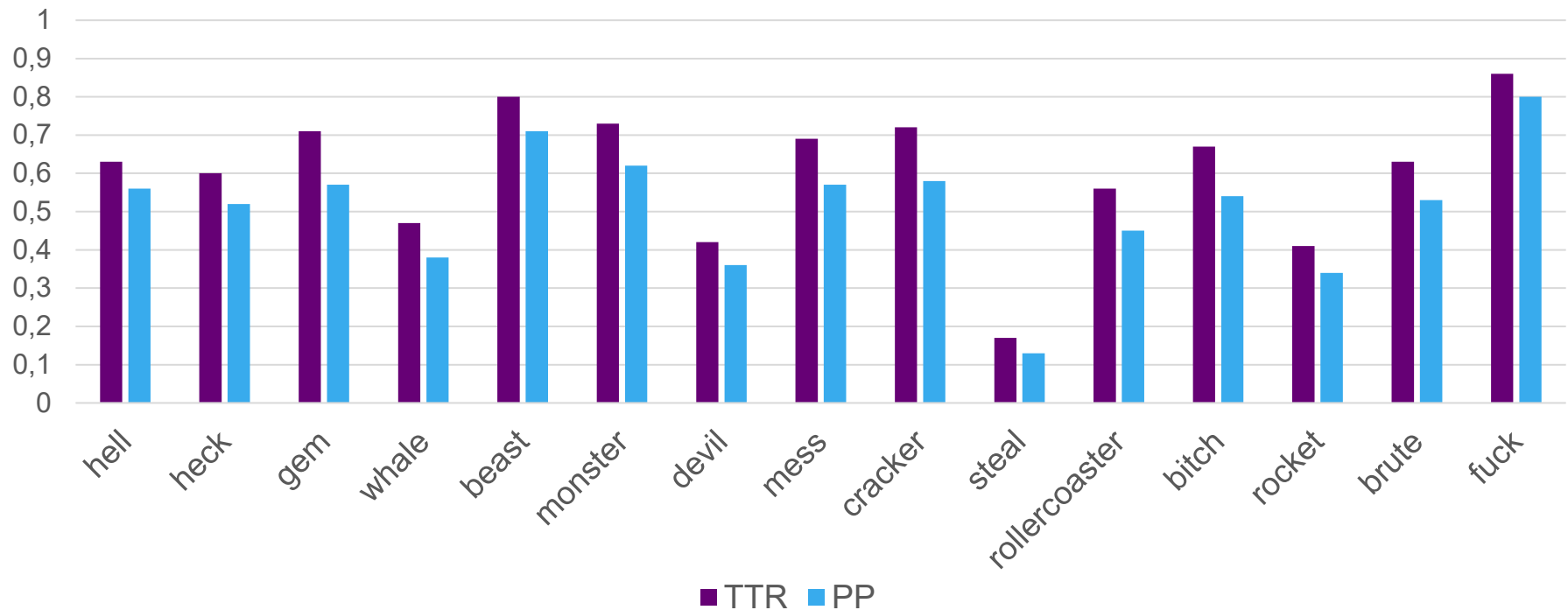


Examples with semantic variation: **Examples with semantic specialization:**

- a **beast** of a **machine** (n=8)
- a **monster** of a **man** (n=8)
- a **brute** of a **man** (n=23)

- a **cracker** of a **game** (n=10)
- a **steal** of a **deal** (n=58)
- a **rollercoaster** of a **year** (n=13)

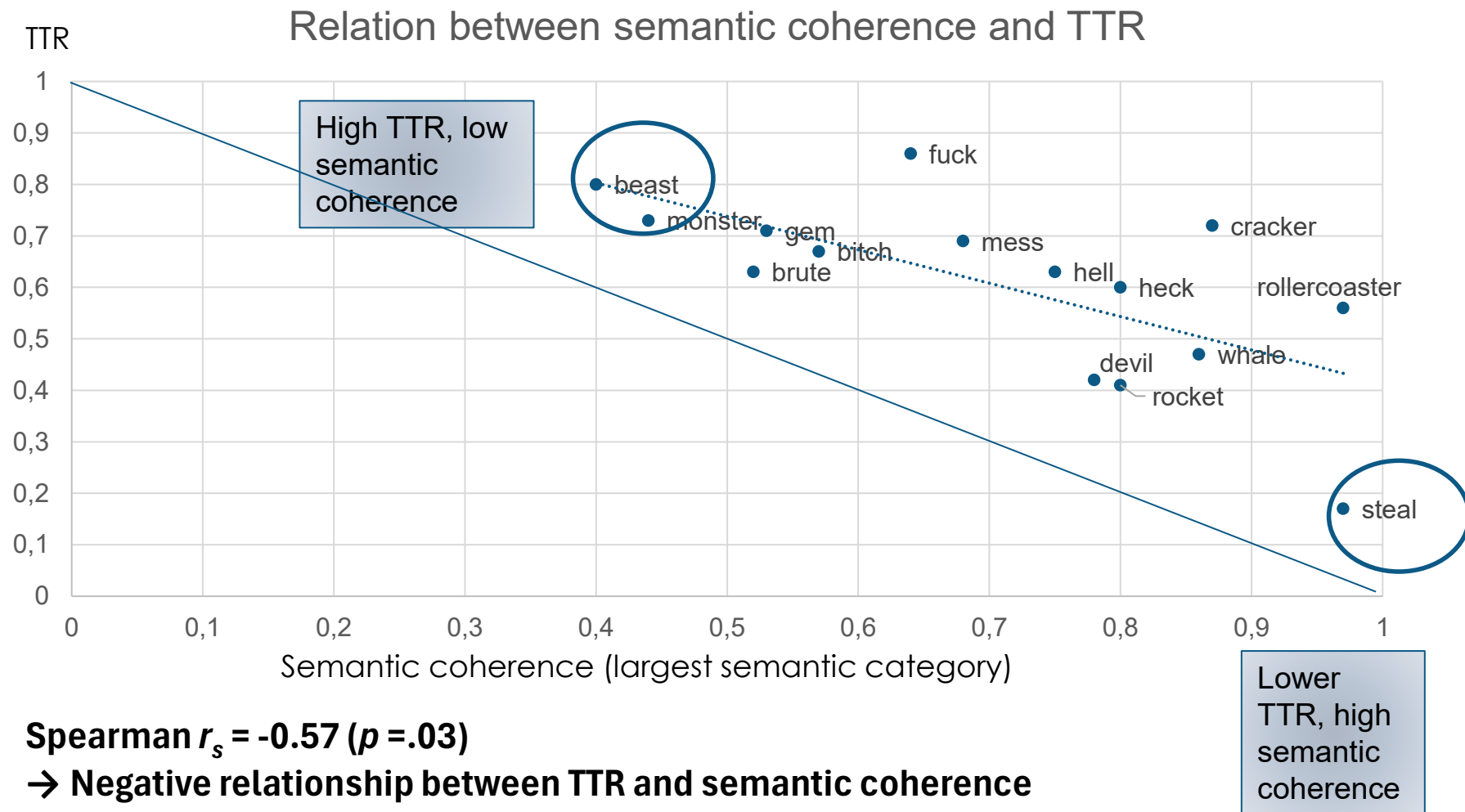
English: Productivity



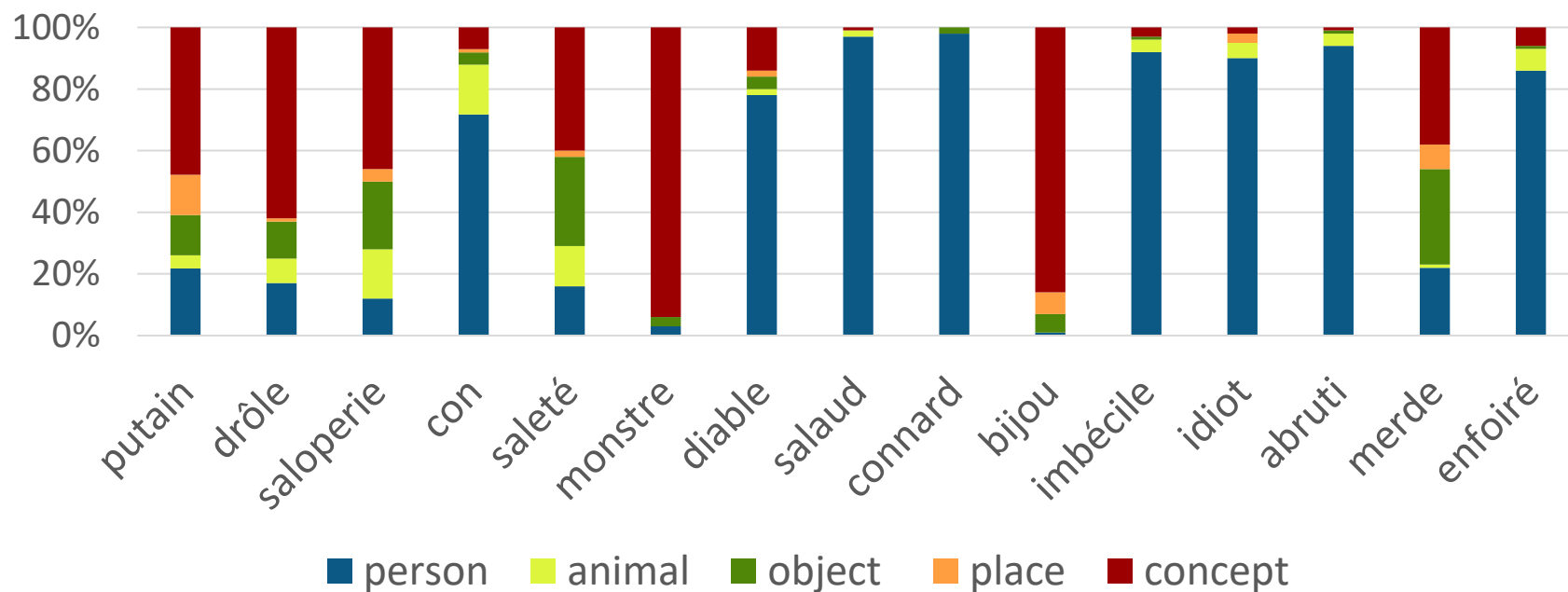
High productivity (hapaxes):

- a **beast** of a **kayak**
- one **monster** of a **migraine**
- some greedy **fuck** of a **chief**

English: Relation between productivity and semantic coherence?



French: Semantics N2



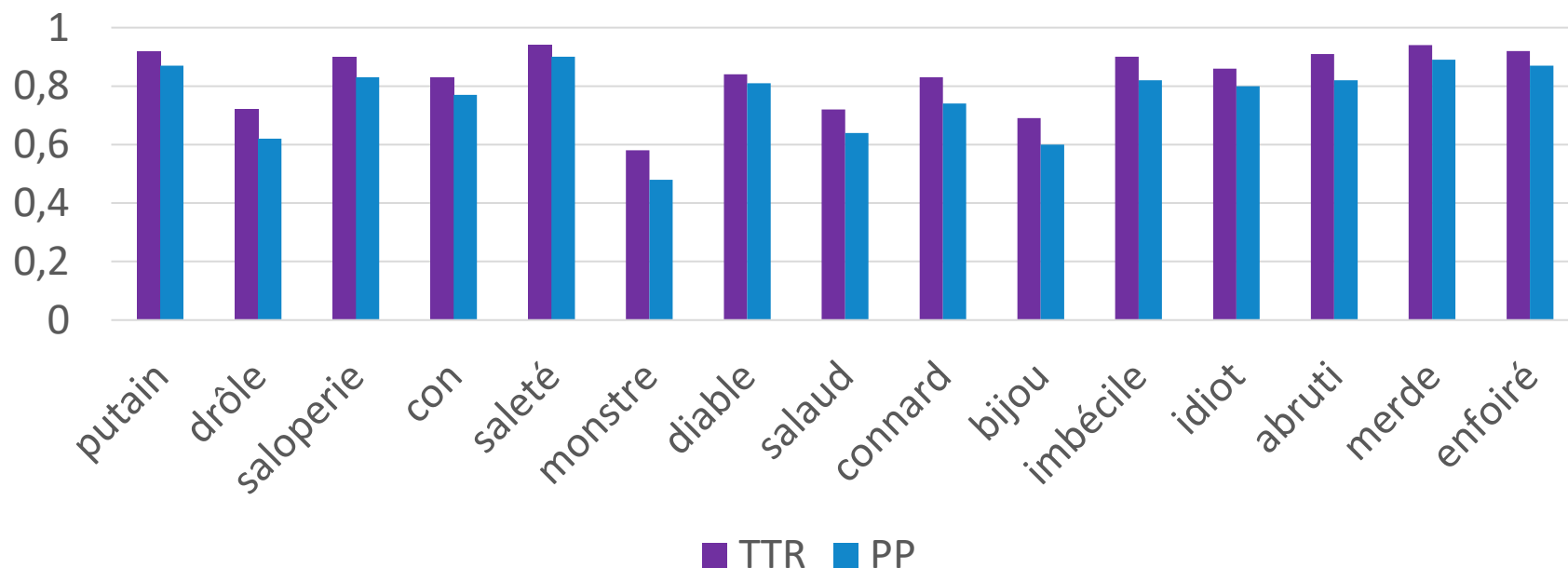
Examples of semantic variation

- *Putain de merde* ! (n=5) 'Fucking hell!'
- *Cette saleté de maladie* (n=4) 'this filthy disease'
- *Ces merdes de voisins* (n=3) 'those shitty neighbours'

Examples of semantic specialization

- *Mon connard de père* (n=6) 'my jerk of a father'
- *Salauds de pauvres* (n=13) 'bastards of poor people'
- *Ce monstre de puissance* (n=19) 'this monster of power'

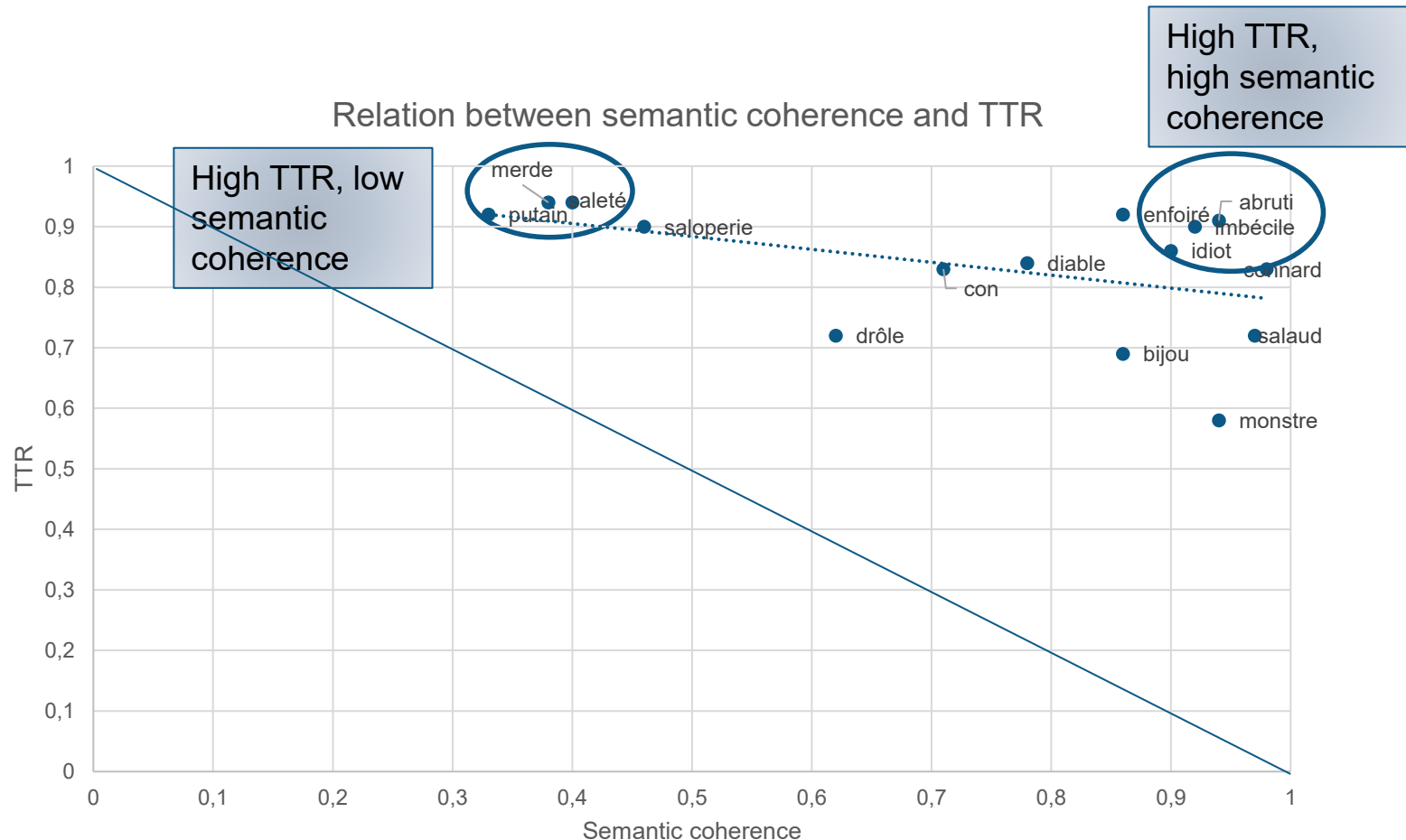
French: Productivity



High productivity (hapaxes):

- *Ce **putain** de **romantisme*** 'this fucking romanticism'
- *Une **saleté** de **bactérie*** 'a nasty bacterium'
- *Cette **merde** de **Redbull*** 'this crap of Redbull'

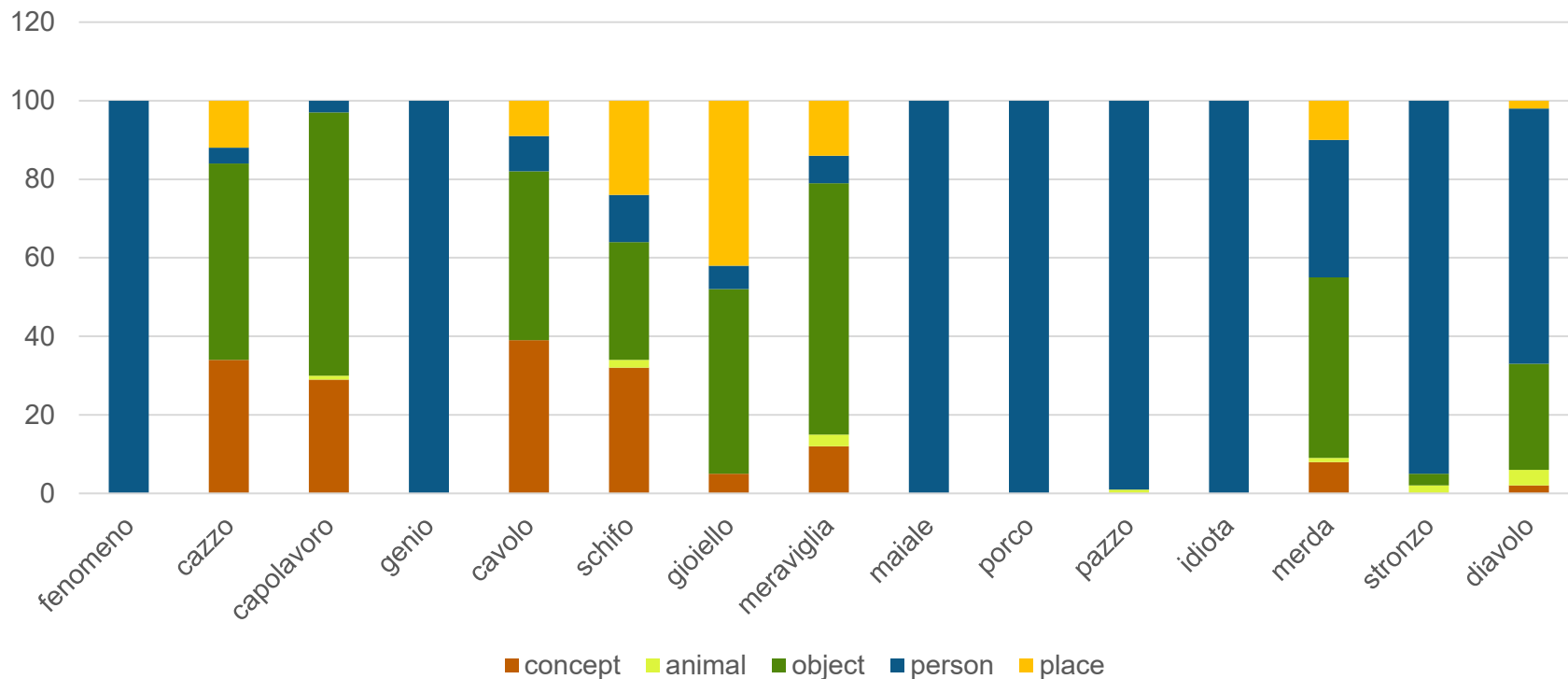
French: Relation between productivity and semantic coherence



Spearman $r_s = -0.53$ ($p = .04$)

→ Negative relationship between TTR and semantic coherence (but lower significance score than in Dutch/English)

Italian: Semantics N2



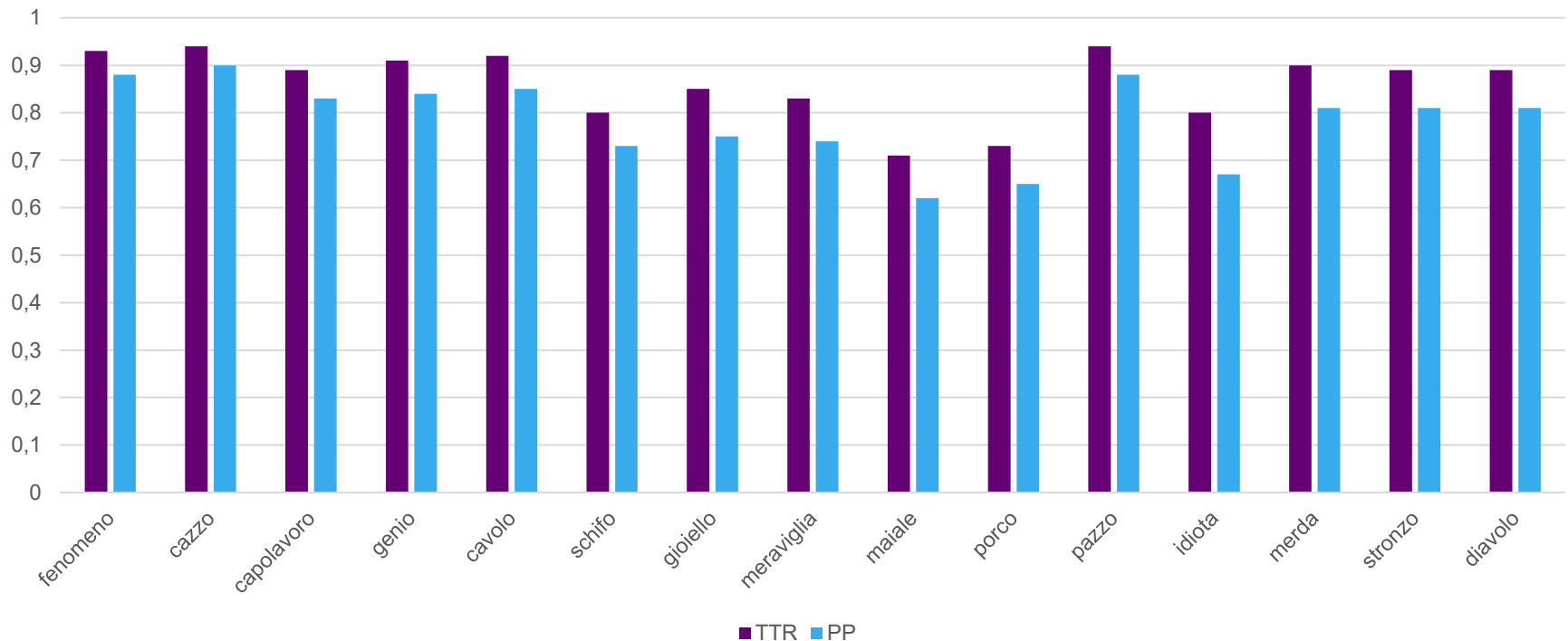
Examples with semantic variation

- *questo **schifo** di **mondo*** (n=9)
'this disgust of world'
- *quel **gioiello** di **città*** (n=6)
'that jewel of a city'
- *questa **meraviglia** di **torta*** (n=5)
'this wonder of a cake'

Examples with semantic specialization

- *quel **fenomeno** di **Cancello*** (n=4)
'that phenomenon of my Cencello'
- *quel **genio** di **mia madre*** (n=3)
'that genius of my mother'
- *quell'**idiota** del **mio ex*** (n=5)
'that idiot of my ex husband'

Italian: Productivity



- High productivity (hapaxes):

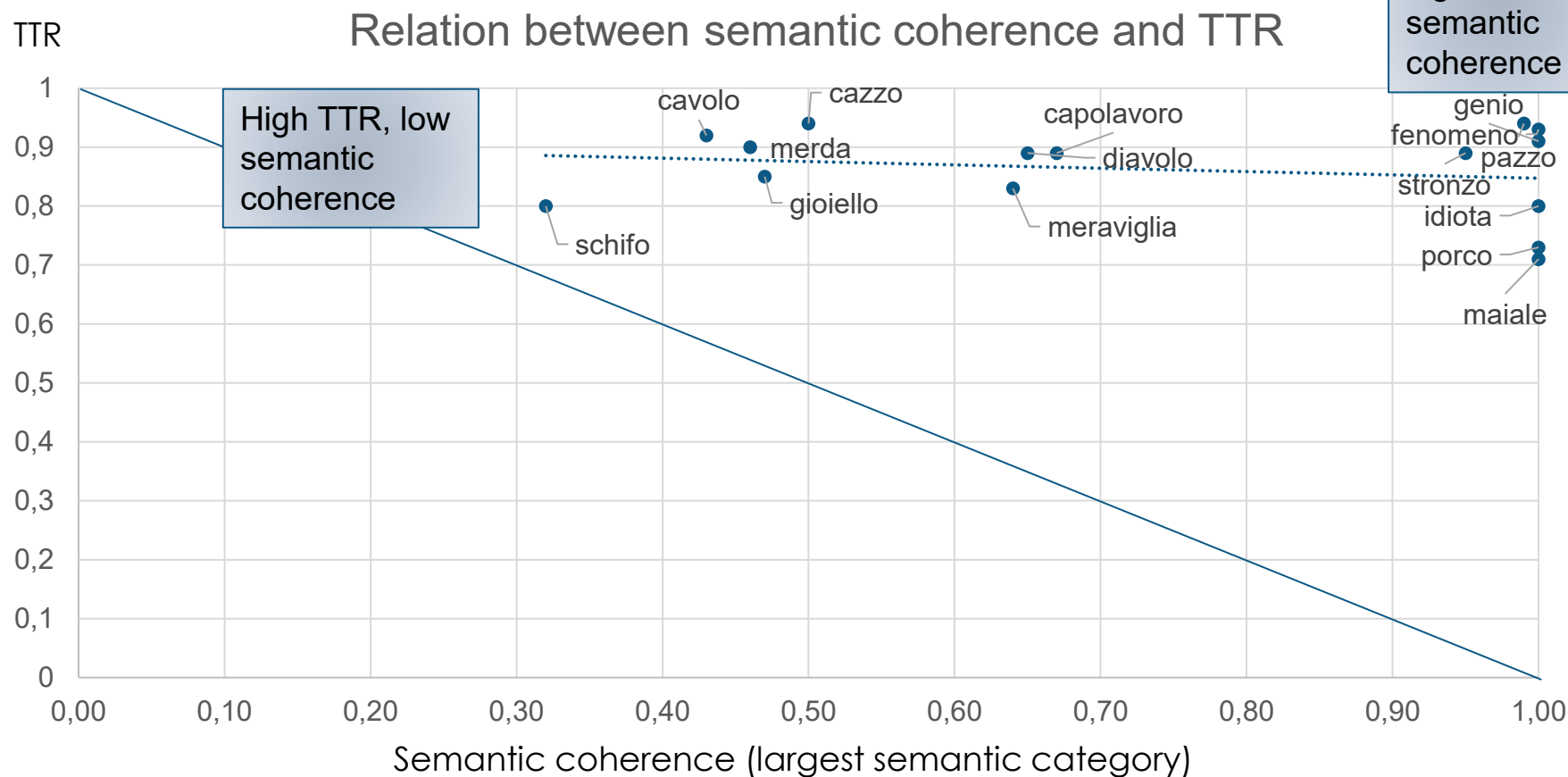
- *quel **genio** di ministro*
- *questo **schifo** di strada*
- *quella **meraviglia** di auto*

‘that genius of ministry’

‘this disgust of a road’

‘that wonder of a car’

Italian: Relation between productivity and semantic coherence?



Spearman $r_s = -0.16$ ($p = .580$)

→ No relationship between TTR and semantic coherence

Discussion and conclusions

- **RQ1: Cross-linguistic similarities and differences?**

- Comparable schematic pattern [DET1 N1 PREP (DET2) N2]
- Comparable non-compositional evaluative semantics
- But also: subtle cross-linguistic differences:
 - **Formal differences:** more variation in DET1 position in Germanic than in Romance; DET2 mostly absent in Romance
 - **Semantic specialization:** ameliorative/augmentative connotation in Dutch, mixed profile in English, pejorative in French and Italian

!!! In addition, **intra-linguistic variation** of the semi-filled micro-constructions with respect to semantic specialization and productivity

Discussion and conclusions

- **RQ2: Relation between productivity and semantic coherence?**
 - **Inverse correlation between productivity and semantic coherence confirmed for Germanic languages**
 - High productivity and low semantic coherence (= flexible schemas)
 - DU [*een pracht/droom van een N*] (ameliorative evaluation)
 - ENG [*a beast of a N*] (mostly ameliorative evaluation)
 - Low productivity and high semantic coherence (= fixed schemas)
 - DU [*een boom van een PERSON*]
 - ENG [*a whale of a CONCEPT*]

Discussion and conclusions

- **RQ2: Relation between productivity and semantic coherence?**
 - **But also, exceptions to this tendency:**
 - In **Italian** there is no inverse correlation between productivity and semantic coherence
 - All analyzed micro-constructions are productive, regardless of the degree of semantic coherence
 - IT [*quel genio di* PERSON]: high productivity + high semantic coherence
 - **French** has an in-between status:
 - Inverse correlation confirmed, but large group of micro-constructions that combine high TTR and high degree of semantic coherence
 - FR [*ce(t) idiot/imbécile/enfoiré/abruti/connard de* PERSON]
- **A semantically constrained construction can still be productive!**
(↔ Barðdal 2008: 39)

Future areas of investigation

Competing constructions

- For instance: competing evaluative compounds in Dutch (cf. Van Goethem 2023)
 - *De kanosport is een **prachtsport** (...) 'beautiful sport'*
 - *Geen **joekel-tattoo** maar gewoon een kleintje (...) 'giant tattoo'*

Expressive pleonastic constructions in French

- ***Putain de saloperie de tout !** 'Fucking fucking everything'*
- ***Saloperie de saloperie de café de merde !** 'Fucking awful coffee!'*
- ***Où elle est cette bordel de merde de putain de ceinture orange à la con?** 'Where the fuck is this fucking orange belt?'*

Grammaticalization of taboo words into intensifiers

- ***Puoi chiudere quella cazzo di bocca** 'you can close that fucking mouth'*
- ***A hell/heck of a lot of money***
- ***Le prochain chapitre est putain de long** 'the next chapter is fucking long'*

References

- Aarts, B. (1994). The syntax of binominal noun phrases in English. *Dutch Working Papers in English Language and Linguistics* 30, 1-28.
- Aarts, B. (1998). Binominal noun phrases in English. *Transactions of the Philological Society* 9(1), 117-158.
- Alinei, M. (1971), Il tipo sintagmatico *quel matto di Giorgio*. In M. Medici & R. Simone (eds.), *La grammatica trasformazionale italiana*, Roma, Bulzoni: 1-12.
- Barðdal, J. (2008). *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam: John Benjamins.
- Foolen, A. (2004). Expressive binominal NPs in Germanic and Romance languages. In: G. Radden & K.-U. Panther (red.), *Studies in linguistic motivation*. Berlin: De Gruyter, 75-100.
- Larrivée, P. (1994). Quelques hypothèses sur les structures syntaxique et sémantique de *Ce fripon de valet*. *Revue québécoise de linguistique* 23(2), 101-113.
- Paardekooper, P.C. 1956. Een schat van een kind. *De nieuwe taalgids* 49. 93-99.
- Piunno (in press) *Quel genio di lingua*. Analisi quantitativa e qualitativa delle costruzioni fraseologiche valutative dell'italiano e confronto con lo spagnolo e il francese. In E. Lavric, C. Konecny, C. Konzett-Firth & M. Messner (eds.). *Actes du IXe Colloque International 'Linguistique contrastive germano-romane et intraromane'*, Berlin: Frank & Timme.
- Regula, M. (1965). Zur Analyse der Fügung *ce fripon de valet*. *Linguistica* 7/1, 57-59
- ten Wolde, E. (2023). *The English Binominal Noun Phrase: A Cognitive-Functional Approach*. Cambridge: Cambridge University Press.
- Simone, R. (in press), On Light Words. In A. Pompei, E. Lombardi Vallauri & V. Piunno (eds.) *Temi e argomenti. Omaggio a Lunella Mereu*, Rome: Roma Tre Press.
- Tutescu M. (1969). Le type nominal *ce fripon de valet*. *Revue de linguistique romane* 33, 299-316.
- Van Goethem, K. (2023). *Een pracht van een constructie vs. een prachtconstructie*. Een vergelijkende corpusstudie van de syntactische en morfologische Expressieve Binominale Constructie. *Nederlandse Taalkunde* 28(2), 325-348.
- Vallaro, M. (1992), A proposito della presenza in arabo del tipo sintagmatico «quello stupidodel professore» , *Quaderni di Studi Arabi*, 11 , 117-138.
- Verhagen, A. (2005). Constructiegrammatica en 'usage-based' taalkunde. *Nederlandse Taalkunde* 10, 197-222.

