Linking across the syntax-lexicon continuum

horizontal links between denominal psych-verbs and light verbs constructions in Italian

Flavio Pisciotta
University of Salerno
fpisciotta@unisa.it

Challenging Construction Grammar: New insights from morphology – Monte Verità, April 9, 2024

DILLS DOTTORATO IN STUDI LETTERARI, LINGUISTICI E STORICI

Overview

- Introduction: LVCs and denominal verbs
- The dataset of psych-predicates
- A paradigm approach to psych-predicates
- Inside the causative cell: competing patterns
- Discussion
- Final remarks

Introduction: LVCs and denominal verbs

- Light verbs constructions (LVCs) are **multiword expressions** involving a verb and a **noun** (or an adjective) → the non-verbal element is the predicative one, the verb is said to be semantically empty

  dare un bacio 'give a kiss' ~ *baciare* 'kiss'

  → Functional overlap between LVCs and morphologically related denominal synthetic verbs (SVs)

- Different degrees of overlap between morphologically related LVCs and SVs:

  - **Event-denoting nouns:**

    LVCs and SVs are not totally synonyms → LVCs denote a single, bounded instance of SV

fare	*una*	*passeggiata*		*passeggiare*
do.INF	ART.IND.SG	walk.SG	⟺	kiss.INF
'to take a walk'				'to walk'

Introduction: LVCs and denominal verbs

- Light verbs constructions (LVCs) are **multiword expressions** involving a verb and a **noun** (or an adjective) → the non-verbal element is the predicative one, the verb is said to be semantically empty

  dare un bacio 'give a kiss' ~ *baciare* 'kiss'

  → Functional overlap between LVCs and morphologically related denominal synthetic verbs (SVs)

- Different degrees of overlap between morphologically related LVCs and SVs:

  - **State-denoting nouns**:

    LVCs are instead more semantically similar to the corresponding SVs

fare	*paura*		*impaurire*
do.INF	fear.SG	⟺	scare.INF
'to scare'			'to scare'

Introduction: competition in the psych-domain

- A particular case: predicates formation from nouns expressing psychological states (e.g., *fear*, *joy*, *anxiety*)
- Different types of psych-verbs: causative vs. non causative; stative vs. eventive (Jackendoff 1990)

  → Three main event types: **stative**, **inchoative** and **causative** verbs (Croft 1991; Talmy 2000)

	SV		LVC	
a.	*simpati**zzare*** 'sympathize'	~	***avere/provare*** *simpatia* (lit. 'have/feel sympathy')	[stative → feel N]
b.	***im****paurir**si*** 'get frightened'	~	***prendere*** *paura* (lit. 'take fear')	[inchoative → begin to feel N]
c.	*angosci**are*** 'distress'	~	***mettere/dare*** *angoscia* (lit. 'put/give anguish')	[causative → cause X to feel N]

- All of them can be expressed by means of many different morphological and LVC patterns

  → this leads to the creation of (sometimes numerous) synonyms

Introduction: competition in the psych-domain

- **A complex situation of competition across the syntax-lexicon "borders"…**
  - Many different morphological and syntagmatic patterns potentially expressing the same semantics
  - Sometimes blocking prevents predicate formation, but seemingly in an unpredictable way (Masini 2019)

    avere *paura* 'have fear' ~ °*paur**are*** but *provare gioia* 'feel joy' ~ *gio**ire*** 'rejoyce'

- **…and across different event types**

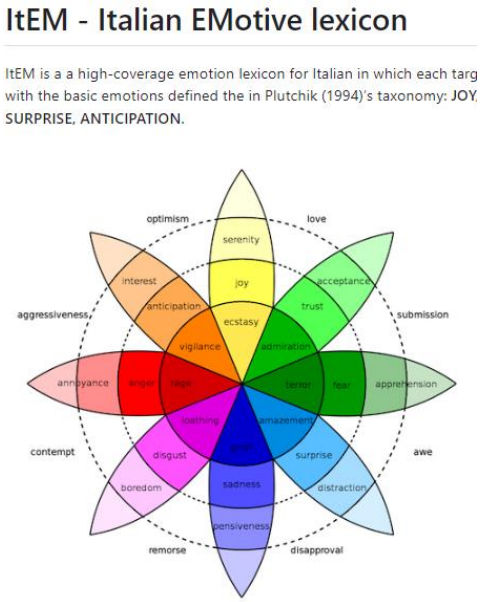
Question: How is it possible to formalize (and address) such a situation in CxG/CxM terms?

→ we will present data and results from ongoing research to theoretically address this challenge

A dataset of psych-predicates

- Pisciotta & Masini (submitted) → Collection of a (as complete as possible) list of psych-predicates
- Lexicographic and corpus data

1. **86 underived psych-nouns** from ItEm (Italian Emotional lexicon, Passaro et al. 2015), in order to:



A dataset of psych-predicates

- Pisciotta & Masini (submitted) → Collection of a (as complete as possible) list of psych-predicates
- Lexicographic and corpus data

1. **86 underived psych-nouns** from ItEm (Italian Emotional lexicon, Passaro et al. 2015), in order to:
2. Extract denominal SVs from GRADIT dictionary (De Mauro 2007)

A dataset of psych-predicates

- Pisciotta & Masini (submitted) → Collection of a (as complete as possible) list of psych-predicates
- Lexicographic and corpus data

1. **86 underived psych-nouns** from ItEm (Italian Emotional lexicon, Passaro et al. 2015), in order to:
2. Extract denominal SVs from GRADIT dictionary (De Mauro 2007)
3. Search the following V N patterns on itWaC Italian Web Corpus (Baroni et al. 2009):

STATIVES	INCHOATIVES	CAUSATIVES
avere N 'have N'	*prendere* N 'take N'	*dare* N 'give N'
essere in N 'be in N'	*farsi* N 'do oneself N'	*mettere* N 'put N'
provare N 'feel N'		*fare* N 'do N'
sentire N 'feel N'		

A dataset of psych-predicates

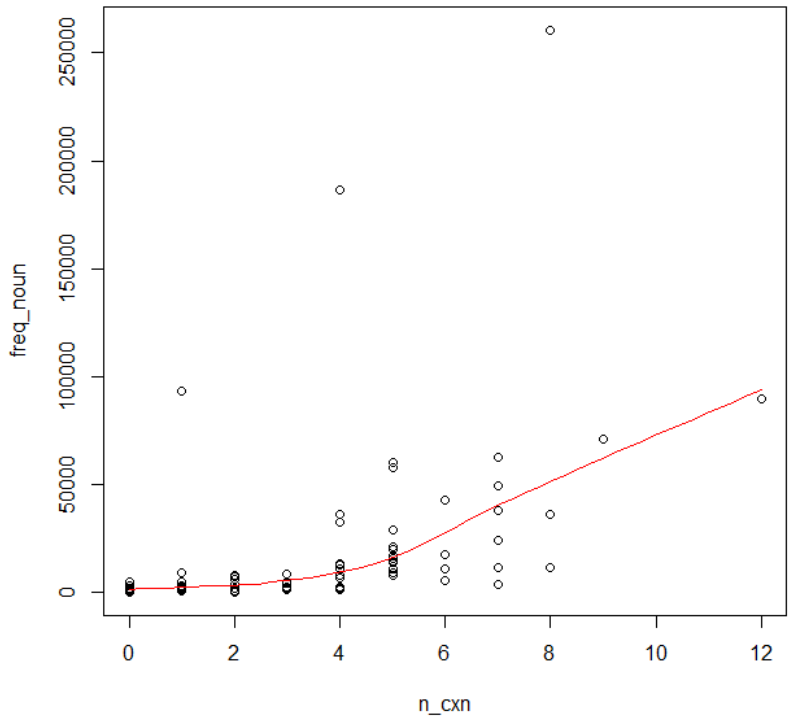
- The final dataset, after cleaning and checking the constructions (false positives, semantics):

206 LVC-types - 75 SV-types

noun (freq.)	**stative predicate** (= feel N)		**inchoative predicate** (= begin to feel N)		**causative predicate** (= cause X to feel N)	
	SV	LVC	SV	LVC	SV	LVC
interesse 'interest' (260748)	*interess**arsi***	***avere** interesse*; ***provare** interesse*	*interess**arsi***	***prendere** interesse*	*interess**are***	***dare** interesse*
amore 'love' (186567)		***avere** amore*; ***provare** amore*; ***sentire** amore*	***in**namor**arsi***			
dubbio 'doubt' (93128)		***essere in** dubbio*				
paura 'fear' (89449)		***avere** paura*; ***provare** paura*; ***sentire** paura*	***im**paur**irsi***	***prendere** paura*	***im**paur**ire***; *spaur**ire***	***fare** paura*; ***mettere** paura;* ***dare** paura*
pena 'pain' (71264)	*pen**are***	***essere in** pena*; ***avere** pena*; ***provare** pena*; ***sentire** pena*		***prendersi** pena*		***fare** pena*; ***dare** pena*

A dataset of psych-predicates

- What's the distribution of the types (wrt nominal bases and event types)?
  - Number of predicates created, by noun frequency (extracted from itWaC) → tau = 0.59, p <0.001



  - LVCs and SVs by event type → p < 0.001

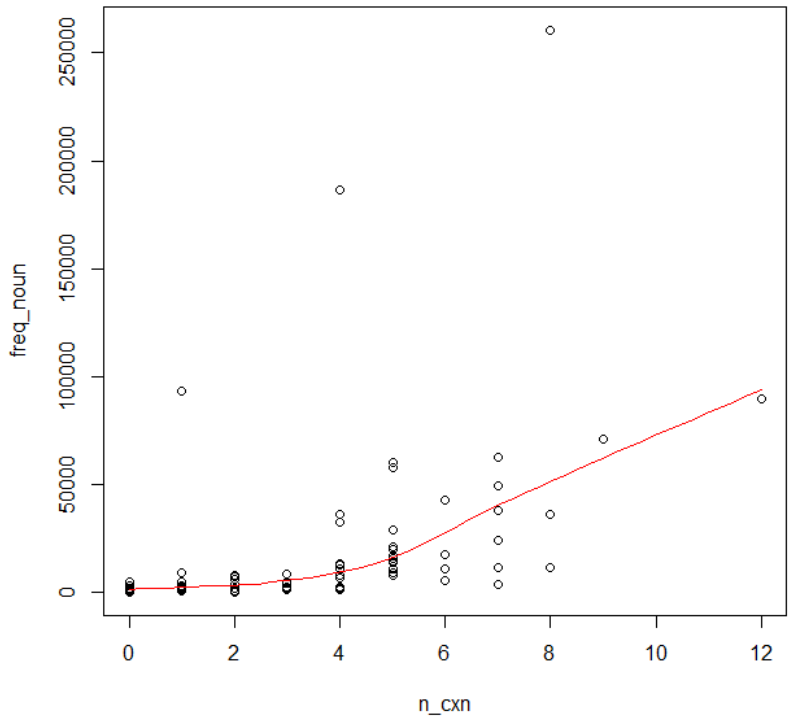
	LVCs	**SVs**	*total*
statives	123	17	*140*
inchoatives	6	26	*33*
causatives	76	32	*108*
total	*206*	*75*	*281*

Std. residuals
Pos. association
Neg. association
No association

Mainly driven by LVCs (tau = 0.61)
The correlation is low with SVs (tau = 0.29)

A dataset of psych-predicates

- What's the distribution of the types (wrt nominal bases and event types)?
  - Number of predicates created, by noun frequency (extracted from itWaC) → tau = 0.59, p <0.001



  - LVCs and SVs by event type → p < 0.001

	LVCs	**SVs**	*total*
statives	123	17	*140*
inchoatives	6	26	*33*
causatives	76	32	*108*
total	*206*	*75*	*281*

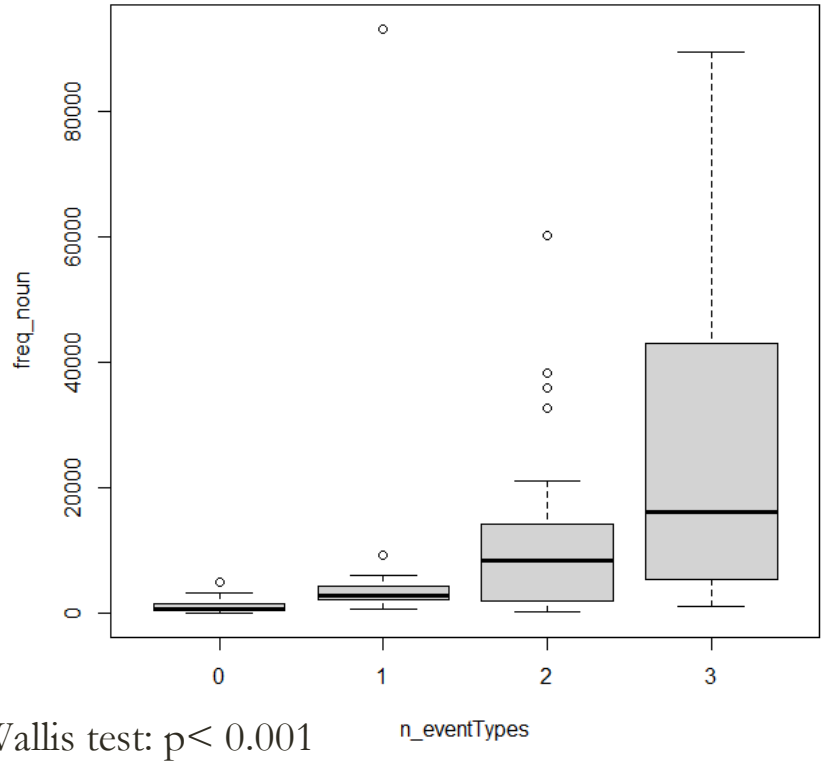
Std. residuals
Pos. association
Neg. association
No association

Multiple synonymic LVCs per noun
Basically no synonymic coradical SVs (blocking?)

A dataset of psych-predicates

- The predicates do not distribute randomly: they tend to express all of the three possible meanings

The more a noun is frequent, the more is likely that it will be used to express all of the three event types



Kruskal-Wallis test: $p < 0.001$

- There seems to be a tendency to fill all the "meaning cells"

→ The full range of psych-events emerges when we look at more frequent psych-nouns

coraggio 'courage'

feel courage	**begin to feel courage**	**cause X to feel courage**
avere *coraggio*	***prendere*** *coraggio*	***fare*** *coraggio*
		in*coraggi****are***

- Can this situation be fruitfully described as a paradigm?

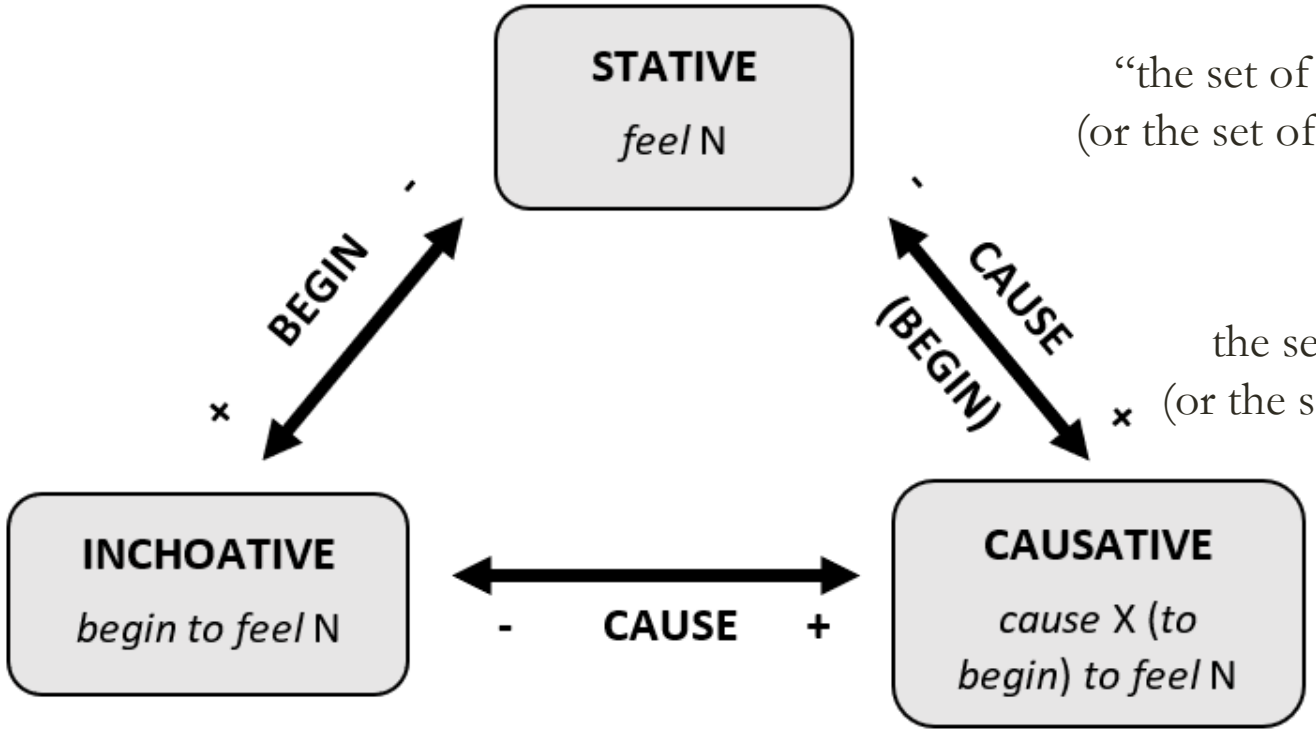
A paradigm approach to psych-predicates

- Paradigms in CxG are highly abstract hyperconstructions that emerge as generalizations of grammatical structures (Diewald & Politt 2022; Leino 2022 ; Diessel 2023). They have been mainly used to formalize inflectional morphology.
- However, our approach is inspired by the concept of derivational paradigms (Štekauer 2014; Hatout & Namer 2019; Melloni & Dal Maso 2022)
- Derivational paradigms can be intended in multiple ways:
  - Derivational "series of morphologically related forms which share a base or base-type" (Bauer 1997)
    e.g., *employ* → *employ**er***, *employ**ment***, *employ**ee***, *employ**able*** …
  - Content-based "paradigms are determined by the semantic content of the lexemes they contain" (Hatout & Namer 2022)
    e.g., Verb → Action_N, Agent_N, etc…

We will start by assuming the second perspective…

A paradigm approach to psych-predicates

- Set of cells representing the three event types (linked by **paradigmatic links**, van de Velde 2014)



Paradigm in an abstract sense

"the set of combinations of morphosyntatic properties or features (or the set of 'cells') realized by inflected forms of words (or lexemes) in a given word-class […] in a language"
(Carstairs-McCarthy 1994: 739)

the set of combinations of semantic properties or features (or the set of 'cells') realized by deriving words from words of a given word-class in a language

← by deriving verbs from psych-nouns

A paradigm approach to psych-predicates

- A paradigm approach allows us to unveil the constructional cooperation between LVCs and SVs:
  - Stative predicates → mainly formed via **LVCs**
  - Inchoatives → mainly formed via **SVs** (anticausative morphology)
  - Causatives → mixed situation

	LVCs	**SVs**	*total*
statives	123	17	*140*
inchoatives	6	26	*33*
causatives	76	32	*108*
total	*206*	*75*	*281*

- More than just a derivational paradigm

  → both morphological schemas and MWEs equally cooperate to express the range of meanings

- Treating uniformly morphological and MWE schemas can be proficuous:
  - Critical points of derivational paradigms (differential exponence, overabundance) (e.g., Melloni & Dal Maso 2022) actually become a way to frame competition at different levels of abstraction!

Inside the causative cell: competing patterns and forms

- Competition among psych-predicates can be found at different (though related) levels of abstraction

  - Competition between patterns → multiple schemas express the same semantics

"Differential exponence [is] a formal inconsistency in the realization forms of paradigmatic slots, where the same morphosyntactic and semantic categories are instantiated by different exponents" (Melloni & Dal Maso 2022) → resulting in affix rivalry in derivational paradigms

e.g., *fare* N, N-*are*, N-*izzare*, *dare* N, *mettere* N, *in/ad-* N *–are/ire* all creating causative predicates

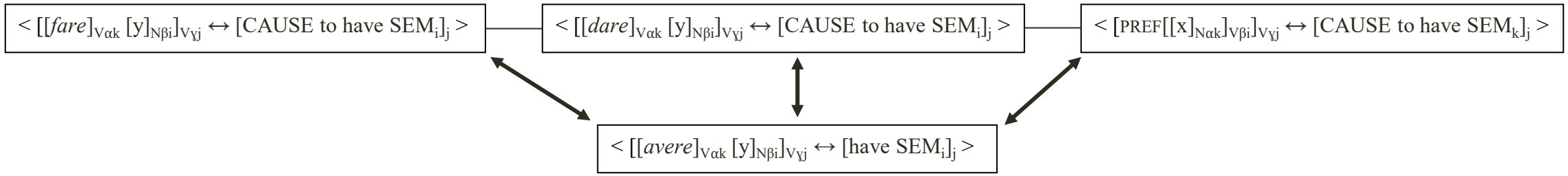
  - Competition between forms → co-radical predicates express the same semantics

"Overabundance is the situation in which two (or more) inflectional forms are available to realize the same cell in the inflectional paradigm of a lexeme" (Thornton 2019)

e.g., *fare paura* (lit. do fear'), *mettere paura* (lit. 'put fear'), *impaurire* all meaning 'to scare'

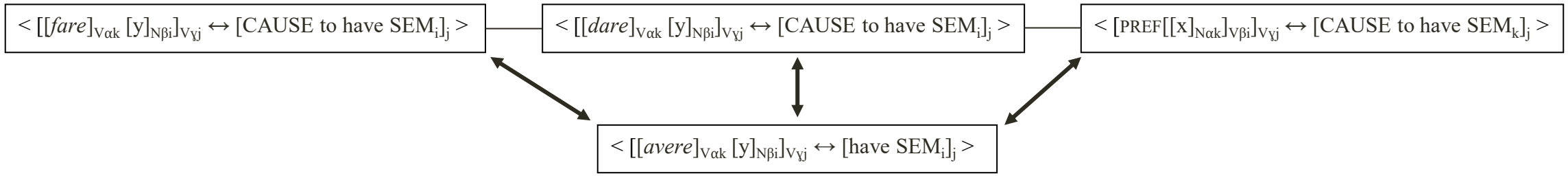
Competition between multiple exponents

- Each of the patterns stands in paradigmatic opposition with patterns filling other cells



Competition between multiple exponents

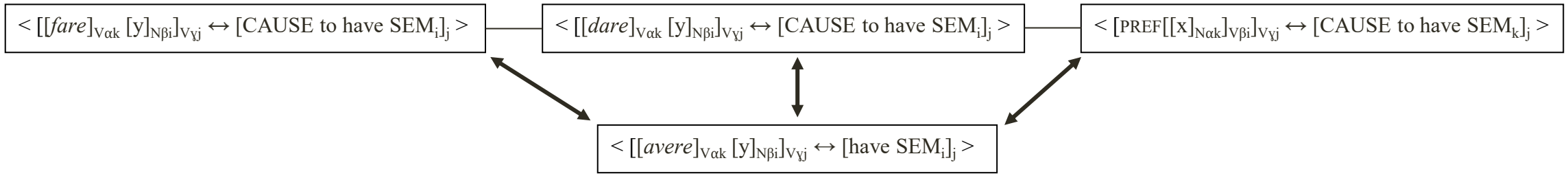
- Each of the patterns stands in paradigmatic opposition with patterns filling other cells



- Attempts to find regularities in the LVC choice → semantics of nouns (e.g., Sanromàn Vilas 2003)

Competition between multiple exponents

- Each of the patterns stands in paradigmatic opposition with patterns filling other cells



- Attempts to find regularities in the LVC choice → semantics of nouns (e.g., Sanromàn Vilas 2003)
- Research in progress on this matter: we present some preliminary data on causatives
- First step: Multiple distinctive collexeme analysis (Gries and Stefanowitsch 2004) based on the frequencies of predicates in itWaC, to obtain a list of the nouns associated with the causative schemas

Competition between multiple exponents

- Semantic classification of the nouns associated to the different patterns (mixed results):
  - *fare* N → associated with DISGUST nouns

    fare schifo 'do disgust'
    fare orrore 'do horror/disgust'
    fare ribrezzo 'do repugnance'
    fare impressione 'do shock'
    fare antipatia 'do antipathy'
  - *mettere* N → associated with FEAR nouns

    mettere paura 'put fear'
    mettere angoscia 'put anguish'
    mettere ansia 'put anxiety'
    mettere inquietudine 'put concern'
    mettere soggezione 'put awe'
- SVs → **valence of the emotion** (conversion more towards positive side, parasynthesis negative one)

  ↓ ↓

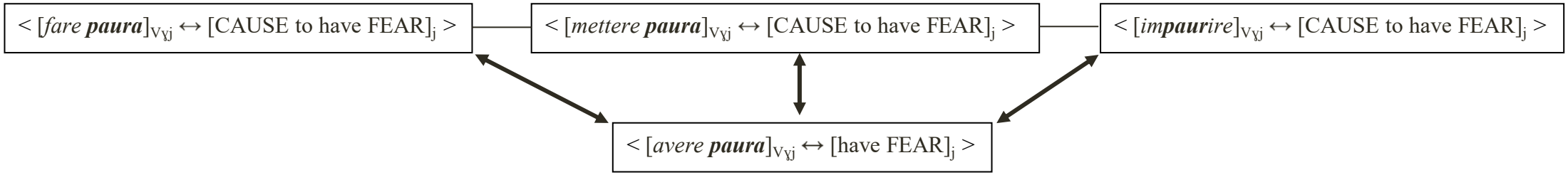
  entusiasmare 'excite', *interessare* 'interest' *infastidire* 'annoy', *intimorire* 'intimidate'

…but there might be more relevant factors, e.g., formal ones (number of syllables, initial phoneme, etc.) for the choice between SV patterns. **Are they relevant on the LVC (and LVC vs SV) side?**

- As for patterns, at this stage we found different niches for LVCs (filler noun semantics), but maybe we need some other criteria to evaluate pattern competition in and between syntax and morphology.

Competition between "co-radical" forms

- Competition between "co-radical" forms arises when semantically similar schemas' slots get filled with the same noun.
- As for patterns, competing co-radical forms are linked by allostructional links, while each of them stands in opposition with forms realizing other cells:



- Competition between forms has been more often addressed, particularly between synthetic and analytic predicates (Sanromán Vilas 2009; Bonial & Pollard 2020) → contextual and cotextual distribution
- Pisciotta & Masini (submitted) performed a mixed-effects analysis of the factors regulating the choice between 10 competing co-radical causative psych-SVs and LVCs (n = 419, written and spoken data)

Competition between "co-radical" forms

- Apart from the effect of register, significant factors are related in a way or another to the non-prototypicality of LVCs as morphological verbs:
  - LVCs are chosen in order to modify the filler noun (morphological vs MWE nature of the cxns)
    e.g., *mettere* ***molta*** *ansia*, lit. 'put **a lot of** anxiety'
  - LVCs are rarely used in non-finite forms, and cannot be used as past participles (LVCs are defective?)
    e.g., *un uomo infastidito*/****dato fastidio*** 'an annoyed man'
  - LVCs allow the speaker to omit the indirect object experiencer argument more often than SVs
    → LVCs already display a "direct object" (the predicative noun) in their argument structure
    e.g., *I lupi fanno paura Ø* lit. 'Wolves do fear'/??*I lupi impauriscono Ø* '??Wolves frighten'
- Thus, also at lower levels of abstraction, we find a division of labour between morphological and syntactic cxns, motivated by structural differences between SVs and LVCs

Summing up: levels of abstraction in our paradigm

set of semantic features realized by forming predicates from psych-nouns

<stative event>, <inchoative event>, <causative event>

content paradigm (Stump 2016)

(Semi-)schematic cxns associated to the semantic content of paradigms cells

<[*fare* N], [*mettere* N], [*in*-N-*ire*] …>

Fully specified cxns realizing the semantic content of paradigms cells

<[(*fare paura*), (*fare schifo*)], …>

realized paradigm (Stump 2016)

Discussion: paradigms across lexicon and grammar

- We found some advantages in accomodating both morphological and multiword cxns in the same paradigm:
  - In our case, it helped to formalize the complex heterogeneity of psych-predicates and their relationship
  - More generally, it allows us to keep together strategies that **cooperate at all the levels of abstraction** in the expression of the same semantic domain
    - Interestingly, the division of labour is found not only between competing forms, but also in the expression of non-overlapping meanings (LVCs → statives; SVs → inchoatives, etc.)
  - Phenomena (e.g., overabundance) seen as non-canonical mappings in inflectional paradigms are actually the norm between derivational and syntactic schemas (and framed as competition)

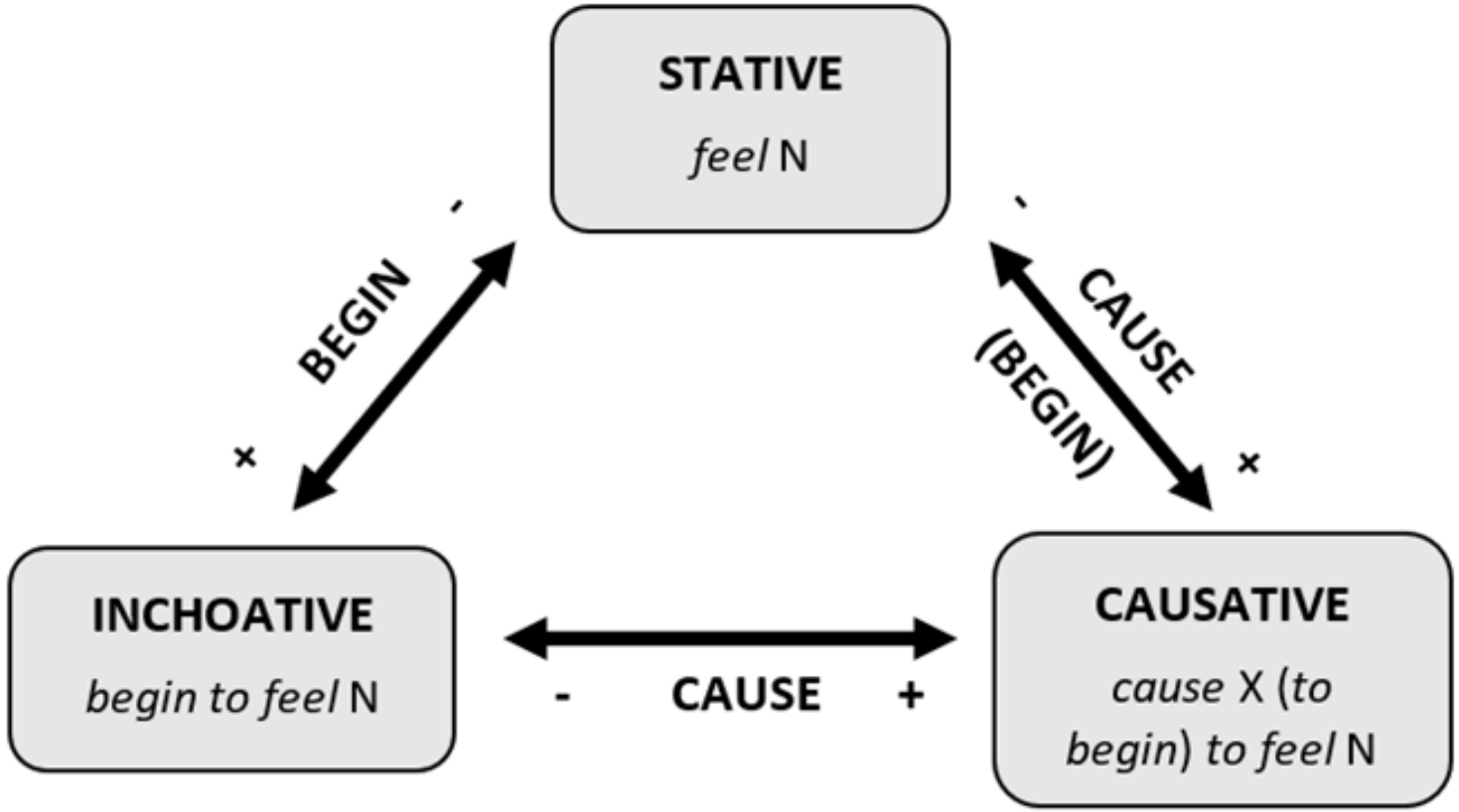
  → **Supporting the idea of continuum**, which justifies a unitary treatment

Discussion: paradigms across lexicon and grammar

- A practical advantage is that CxG already has the machinery to formalize such a paradigmatic approach:
  - **Paradigms** → generally applied separately to morphological (e.g., inflectional) and syntactic phenomena, they can represent higher-level generalizations, even very abstract ones (such as content-based paradigms' cells)
- Paradigms can be internally structured via vertical and, crucially, **horizontal relations**:
  - **Paradigmatic links** can be used between exponents and realizations of different cells
  - **Allostructional links** can be used between exponents and realizations of the same cell

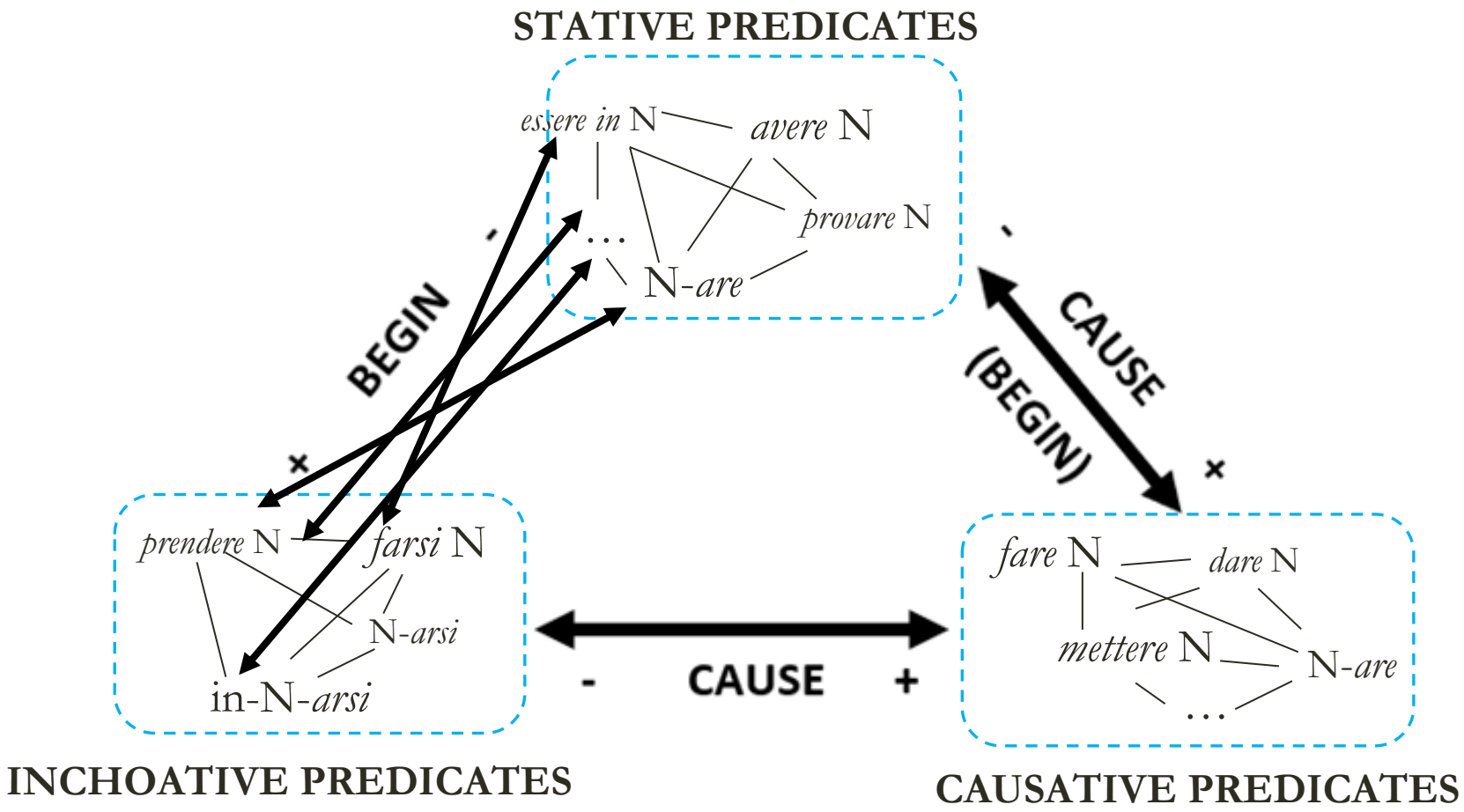
Discussion: paradigms across lexicon and grammar

- By assuming this perspective, our "abstract paradigm" can be thought as a simplified view of a complex network, in which paradigms cells are clusters of constructions connected by allostructional links.



Discussion: paradigms across lexicon and grammar

- By assuming this perspective, our "abstract paradigm" can be thought as a simplified view of a complex network, in which paradigms cells are clusters of constructions connected by allostructional links.



Final remarks: some cautious words

- Note that our unified approach to does not deny the **different behaviour and phenomena pertaining to structurally different constructions** (i.e., that lie at different points of the continuum)

- In particular, we find:
  - Differences in the entrenchment → SVs are overall more frequent and stable, while we find less frequent LVCs that are probably constructs and not constructions:

    e.g., *dare timore* 'to intimidate' (lit. 'give awe')

  - Presence of blocking among SVs → we generally find only one SV per meaning, while the creation of synonyms is not blocked among LVCs

  - Different degrees of transparency and separability → LVCs are more separable and structurally transparent than SVs

- While this motivates constructional cooperation, it poses some methodological challenges, especially when some factors are relevant only on one of the two sides.

Thank you for your attention!

Questions and (critical) comments are welcome

Flavio Pisciotta
e-mail:
fpisciotta@unisa.it

Mixed-effects model: data

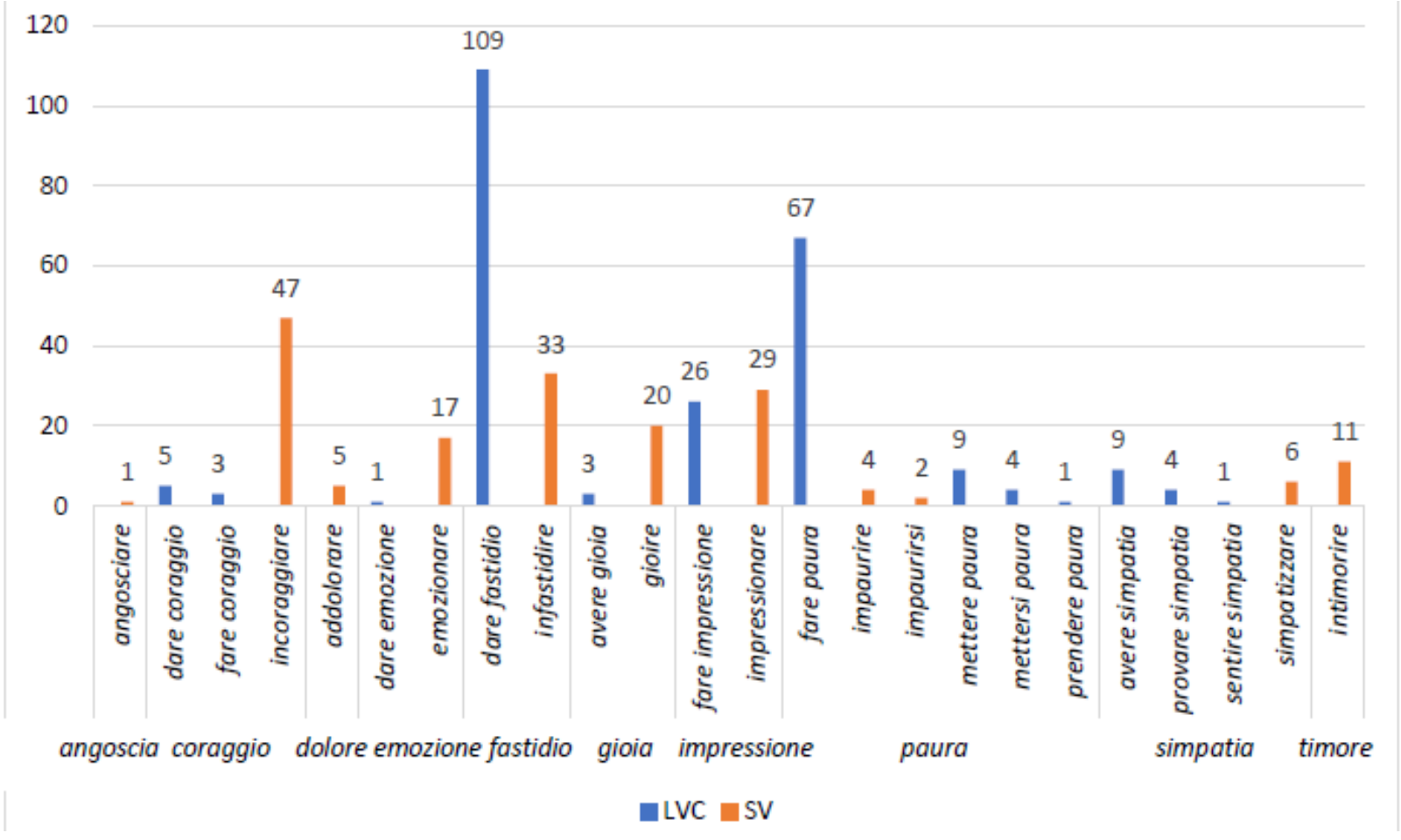
noun	SV	LVC	event type
paura 'fear'	*impaurire, spaurire*	*fare/mettere/dare paura*	causative
	impaurirsi	*prendere paura*	inchoative
dolore 'pain'	*addolorare*	*dare dolore*	causative
emozione 'excitement'	*emozionare*	*dare emozione*	causative
gioia 'joy'	*gioire*	*avere/provare/sentire gioia*	stative
coraggio 'courage'	*incoraggiare*	*fare/dare coraggio*	causative
impressione 'impression'	*impressionare*	*fare/dare impressione*	causative
timore 'fear'	*intimorire*	*dare/fare timore*	causative
fastidio 'bother'	*infastidire*	*dare fastidio*	causative
simpatia 'sympathy'	*simpatizzare*	*avere/provare simpatia*	stative
angoscia 'anguish'	*angosciarsi*	*essere in/provare angoscia*	stative

500 total occurrences extracted from:
Written Italian → CORIS (309)

Spoken Italian → KIParla (117)
LIP (30)
Radiocast (44)

Mixed-effects model: data

After cleaning the dataset → 419 occurrences



Mixed-effects model: factors

Variable	Levels
Corpus	• Written • Spoken
Modification (adjectives or adverbs of quantity or quality)	• Modified • Non_Modified
Verb_Form	• Finite • Non Finite_Infinitive • Non Finite_Other (participles, gerunds)
Text_genre	• Dialogic_speech (face-to-face and telephone conversations) • Monologic_speech (university lessons, public speeches) • Broadcast (radio and TV speech) • Fiction_prose • Press • NonFiction_prose (academic and legal prose) • Websites (blog posts)
[causatives only] *Sem_SubjStimulus* *(Subject-Stimulus semantics)*	• Egophoric (i.e., non-3rd person stimulus, as in (*Io*) *vi do coraggio* '**I** give you courage') • Animate • Inanimate • Clause&Clause_referred (i.e., ***Camminare al buio** fa paura* '**Walking in the dark** is scary') • Unknown/Unspecified (i.e., the stimulus is not specified, as in *Ho visto un bambino impaurito* 'I saw a frightened child',)
[causatives only] *Sem_NonSubjExperiencer* *((In)direct object-Experiencer semantics)*	• Egophoric (i.e., non-3rd person experiencer *I rumori **mi** infastidiscono* 'Noises bother **me**') • Animate • Inanimate • Zero (i.e., the experiencer is not specified, as in *L'ignoto fa paura* 'The unknown is frightening (lit. frightens)')

Mixed-effects model: results (all predicates)

Construction ~ Text_genre + Verb_Form + Modification + (1|Noun) + (1| Event_type)

Predictor	Levels	Estimate	Std.Error	z_ratio	p_value	
Verb_Form *** (p <.001)	Finite	0.1529	0.4962	-3.6977	0.0002	***
	NonFinite_Infinitive	1.0610	0.5227	-1.7730	0.0762	.
	NonFinite_Other	4.7493	0.9544	2.8935	0.0038	**
Text_genre *** (p <.001)	Broadcast	2.9456	0.4481	2.1377	0.0325	*
	Dialogic_speech	0.6463	0.3782	-3.5470	0.0004	***
	Fiction_prose	2.3026	0.3208	0.9814	0.3264	
	Monologic_speech	-0.3611	1.0304	-2.2796	0.0226	*
	NonFiction_prose	3.1191	0.6516	1.7363	0.0825	.
	Press	2.1809	0.3556	0.5432	0.5870	
	Websites	3.0809	0.7030	1.5548	0.1200	
Modification ** (p = .009)	Modified	1.4869	0.1975	-2.5356	0.0112	*
	Non_Modified	2.4886	0.1975	2.5356	0.0112	*

Model Performance	Classification accuracy	AIC	BIC	R^2 (cond.)	R^2 (marg.)
	0.84	351.971	401.368	0.654	0.289

Mixed-effects model: results (causatives)

Construction ~ Text_genre + Verb_Form + Modification + Sem_NonSubjExperiecer + (1|Noun)

Predictor	Levels	Estimate	Std.Error	z_ratio	p_value	
Sem_NonSubjExp ***	Animate	3.2237	0.3942	2.9795	0.0029	**
(p <.001)	Egophoric	1.6111	0.4210	-1.0405	0.2981	
	Inanimate	2.6693	0.9414	0.6589	0.5100	
	Zero	0.6922	0.4870	-2.7861	0.0053	**
Verb_Form *** (p <.001)	Finite	0.5185	0.4868	-3.1444	0.0017	**
	NonFinite_Infinitive	1.0936	0.5392	-1.7722	0.0764	.
	NonFinite_Other	4.5351	0.9225	2.6949	0.0070	**
Text_genre *** (p <.001)	Broadcast	2.8649	0.4916	1.6594	0.0970	.
	Dialogic_speech	0.8075	0.4472	-2.7760	0.0055	**
	Fiction_prose	2.0228	0.3914	-0.0671	0.9465	
	Monologic_speech	-0.2406	1.1385	-2.0111	0.0443	*
	NonFiction_prose	2.6205	0.8601	0.6643	0.5065	
	Press	2.2360	0.4320	0.4328	0.6651	
	Websites	4.0323	0.9371	2.1163	0.0343	*
Modification * (p = .035)	Modified	1.5717	0.2324	-2.0537	0.0400	*
	Non_Modified	2.5264	0.2324	2.0537	0.0400	*

Model Performance	Classification accuracy	AIC	BIC	R^2 (cond.)	R^2 (marg.)
	0.88	270.394	325.031	0.760	0.320