e-ISSN 2785-3187

Balcania et Slavia
Vol. 5 – Num. 2 – December 2025

Binominal Constructions and Categorization: An Experiment on Categorial Membership in Italian and Russian

Alessia Lacroce
Università degli Studi di Torino, Italia

Beatrice Bernasconi
Università di Torino, Italia

Abstract Recent research has shown that binominal constructions such as [N1 *of* N2] and [N1 N2GEN] can function as classifiers in languages without dedicated classifier systems. In these constructions, N1 plays a classifier-like role, while N2 is the classified noun. Although much work has focused on the types of N2s allowed, semantic groupings are rarely defined, and when they are, they often reflect the researcher's interpretation rather than shared cognitive patterns. To investigate this, we conducted a categorization experiment with L1 speakers of Italian and Russian to test the validity of a 21-category classification identified based on corpus data. Participants were asked to classify lexical items into proposed categories or suggest new ones, with each category including both prototypical and peripheral members. Results were analyzed through quantitative methods and qualitative observations.

Keywords Binominal constructions. Categorization. Cross-Linguistic experiment. Quantitative analysis

Index 1 Introduction. – 2 Methodology. – 3 Results and Discussion. – 4 Conclusion.

Edizioni Ca'Foscari

Peer review

Submitted 2025-10-09
Accepted 2025-12-07
Published 2026-03-31

Citation Lacroce, A.; Bernasconi, B. (2025). "Binominal Constructions and Categorization: An Experiment on Categorial Membership in Italian and Russian". *Balcania et Slavia*, 5(2), 29-46.

DOI 10.30687/BES/2785-3187/2025/01/002

1 Introduction

When dealing with nominal classification, languages can be divided into two main types. Classifier languages (e.g., Chinese, Thai) obligatorily mark nominal classification through a closed set of grammaticalized forms called 'classifiers' (Tai, Wang 1990; Aikhenvald 2003). Non-classifier languages (e.g., English, Italian, Russian) lack such devices but may resort to alternative strategies for cases where classification is felt necessary (e.g., Russian numerals, Sussex 1976).[1]

Since the 1970s, linguists have identified binominal constructions of the type 'N1 *(of)* N2' as a possible systematic strategy for nominal categorization in many languages (Allan 1977; Cinque 2006; Cotta Ramusino 2016; Xu 2017; Benigni 2022; Benigni, Latos 2023, 2024). In these patterns, N1 plays a role comparable to that of a classifier, while N2 is the categorized noun. Two structural types are recurrent: (i) constructions where N2 is introduced by a preposition (e.g., N1 *of* N2) as in *three bales of hay*, and (ii) constructions where N2 appears in the genitive case, as in Ru. *tjuk-Ø sen-a*$_{\text{GEN}}$ 'a bale of hay'.

The co-occurrence of specific types of nouns in the two slots of the construction depends on their semantic relation. For instance, collective N1s preferably co-occur with groups of animate entities; arrangement N1s with inanimate sets are organized into a specific shape; measure words with measurable entities, such as liquids, mass-like items, or surfaces. Most studies within this line of research examine the lexical types of N2 admitted by a given classifier-like noun to identify regularities in their selection (Tai, Wang 1990; McEnery, Xiao 2007; Verveckken 2015). Less frequently, however, researchers have worked on the identification of a cross-linguistic list of semantic categories to group the various N2s (Benigni 2022; Benigni, Latos 2023, 2024; Lacroce 2023). Since such categories are researcher-defined, they risk reflecting interpretive bias rather than actual cognitive processes of categorization. Moreover, as stated by Clark:

> semantic categories do not map to each other across languages in any direct fashion. […] The boundaries for lexico-semantic categories are likewise seldom clear-cut as one moves from one language to another. (Clark 2017, 407)

1 The two authors equally contributed to the conceptualization, development, and overall realization of this article. For Italian academic purposes only, Alessia Lacroce was responsible for Sections 1, 3.2, and 3.3, while Beatrice Bernasconi was responsible for Sections 2.2, 3.1, and 4. Both authors share equal responsibility for Section 2.1.

For some lexical items, categorization membership is relatively straightforward and intuitively shared among speakers. For others, however, categorization proves more complex, since their semantic features may overlap with two or more possible classes. This phenomenon reflects the inherently fuzzy nature of linguistic categories (Mervis, Rosch 1981; Taylor 2003), which lack clear-cut boundaries and include less prototypical entities at their edges. Such fuzziness is particularly relevant in the study of binominal constructions, where peripheral members challenge the stability of semantic groupings and reveal points of divergence in speaker judgments.

In Lacroce 2023, a corpus-based analysis of binominal constructions in Italian was carried out to demonstrate the classifier-like function of such patterns. The analysis yielded a data-driven list of operational semantic categories to describe the items that can fill the N2 slot of the Italian construction N1 *di* N2 'N1 of N2'. However, this classification may still be biased by the type of data considered, the researcher's perspective, and their personal mental representation of items. Therefore, this paper aims to verify the validity of such a categorization by presenting the results of an experiment conducted with L1 speakers of Italian and Russian. The study will address the following research questions:

- RQ1: Do L1 speakers of Italian and Russian consistently agree in assigning items to a finite set of categories?
- RQ2: What is the informants' agreement rate in categorizing a central vs. a peripheral member of a class?
- RQ3: Which principles underlie the process of class inclusion of lexical items in the two languages?

More broadly, our findings contribute to understanding how non-classifier languages encode categorization. Italian and Russian represent two interesting testing grounds: both are non-classifier languages, yet they differ significantly in morphological type and in the morphosyntactic realization of N1-N2 relations.

The remainder of this article is structured as follows. Section 2 provides details on the methodology and the experimental design. Section 3 summarizes and discusses the results of the experiment. Finally, Section 4 draws some conclusions.

2 Methodology

2.1 The Experiment

In order to design the categorization experiment, we first analyzed the collocational profile of three Italian and Russian classifier-like nouns to identify the stimuli among a list of authentically occurring N2s. For Italian, we focused on *mucchio* 'heap', *mazzo* 'bunch', *pila* 'pile'. The three nouns are among the most productive items and are very common in the Italian language. Moreover, none of them has fully grammaticalized into a quantifier (as is the case of *un sacco di* 'a lot of'). For Russian, we selected their semantically comparable counterparts (namely, *kuča* 'heap', *buket* 'bunch', *kopna* 'pile').

Relying on data from the *COrpus di Riferimento dell'Italiano Scritto* (CORIS) for Italian, and the main corpus (*osnovnoj*) of the *Nacional'nyj Korpus Russkogo Jazyka* (NKRJa) for Russian, we extracted the hundred most frequent types for each construction (i.e., *mucchio di* N, *mazzo di* N, pila di N, *buket* N_{GEN}, *kuča* N_{GEN}, *kopna* N_{GEN}), obtaining a dataset of 300 collocates per language.

Then, we grouped the attested N2 collocates into higher-level categories based on their semantic features. At this stage, we relied on the adjacent linguistic co-text to collocate ambiguous items. Drawing on the classification proposed by Lacroce (2023) within the corpus-based analysis of binominal constructions in Italian, a list of 21 semantic categories was employed to classify each term. The choice of semantic categories was data-driven and aimed at ensuring comprehensive coverage of the attested items, while avoiding an overly fine-grained classification that would have resulted in a longer and more fragmented list. For this reason, the categories cover a wide range of domains, from animate to inanimate entities, from more concrete to more abstract ones: 1. Animals, 2. Body parts, 3. Buildings and places, 4. Bundles and waves, 5. Celestial bodies, 6. Clothing and linen, 7. Containers, 8. Flowers, herbs, and plants, 9. Food, 10. Fruits, vegetables, and legumes, 11. Generic, 12. Human activities, 13. Human beings, 14. Liquids, 15. Minerals and rocks, 16. Money, 17. Paper and related materials, 18. Powders, 19. Tools, 20. Waste, 21. Wood.

The list of annotated N2s from the corpora was used as a source for the experiment's stimuli. For each of the 21 categories, two lexical items were selected for a total of 42 stimuli per language. To capture the typical properties of internal category structure, we

chose[2] one central member, i.e., a prototypical exemplar that clearly instantiates the category (e.g., It. *ragazzi* 'boys' for Human beings) and one peripheral member, namely, a less prototypical item that possibly shares features with more than one category (e.g., It. *cuscino* 'pillow' may be classified both within Tools and Clothing and linen). This design allows us to investigate both the stability of prototypical membership and the fuzziness of category boundaries. In selecting the stimuli, we attempted to maintain comparability between the Italian and Russian versions of the test; hence, when possible, we selected semantically similar or equivalent items (for literature in support of pairing conceptually similar items across languages, cf. Haspelmath 2010; Baker 2011). Table 1 shows the full list of items provided for both languages. Twenty items across ten categories are cross-linguistically equivalent and are highlighted in grey.

Table 1 List of stimuli included in the two versions (Italian and Russian) of the experiment

Categories	Italian		Russian	
	Central members	**Peripheral members**	**Central members**	**Peripheral members**
Animals	*vacche* 'cows'	*pulci* 'fleas'	*voly* 'oxes'	*bakterii* 'bacteria'
Body parts	*capelli* 'hair'	*pelli* 'skins'	*volosy* 'hair'	*kosti* 'bones'
Buildings and places	*case* 'houses'	*stazioni* 'stations'	*doma* 'houses'	*ruiny* 'ruins'
Beams and waves	*raggi* 'rays'	*scosse* 'shakes'	*svet* 'light'	*ul'trazvuki* 'ultrasounds'
Celestial bodies	*stelle* 'stars'	*satelliti* 'satellites'	*solnce* 'sun'	*asteroidy* 'asteroids'
Clothing and linen	*abiti* 'clothes'	*cuscini* 'pillows'	*bel'ë* 'linen'	*poduški* 'pillows'
Containers	*scatole* 'boxes'	*pneumatici* 'tires'[1]	*meški* 'bags'	*komnaty* 'rooms'
Flowers, herbs, and plants	*violette* 'violets'	*paglia* 'hay'	*gladiolusy* 'gladioli'	*pšenica* 'wheat'
Food	*lasagna* 'lasagna'	*crema di sesamo* 'sesame cream'	*xleb* 'bread'	*kukuruza* 'corn'
Fruits, vegetables, and legumes	*meloni* 'melons'	*rucola* 'arugula'	*jabloki* 'apples'	*gorox* 'peas'
Generic	*roba* 'stuff'	*doni* 'gifts'	*veščі* 'stuff'	*kačestva* 'qualities'
Human activities	*viaggi* 'travels'	*domande* 'questions'	*rabota* 'work'	*voprosy* 'questions'

2 As one anonymous reviewer pointed out, our a priori selection of certain items may represent a prototypical instance of the researcher's own interpretive bias(es) possibly influencing the results and outcomes of the research. While we acknowledge this, we aimed to select items that could be considered central or peripheral members as objectively as possible, to provide a clear starting point. In any case, the experiment results offered a valuable testing ground for potential biases.

Human beings	*ragazzi* 'boys'	*clientele* 'clienteles'	*rebjata* 'boys'	*angely* 'angels'
Liquids	*acqua* 'water'	*olio* 'oil'	*vino* 'wine'	*kraski* 'paint'
Minerals and rocks	*sassi* 'rocks'	*ghiaia* 'gravel'	*kamni* 'rocks'	*glina* 'clay'
Money	*banconote* 'banknotes'	*debiti* 'debts'	*den'gi* 'money'	*dolgi* 'debts'
Paper and related materials	*giornali* 'newspapers'	*poesie* 'poems'	*gazety* 'newspapers'	*dokumenty* 'documents'
Powders	*farina* 'flour'	*sale* 'salt'	*pepel* 'ash'	*pesok* 'sand'
Tools	*chiavi* 'keys'	*biciclette* 'bicycles'	*zontiki* 'umbrellas'	*mašiny* 'cars'
Waste	*spazzatura* 'garbage'	*cadaveri* 'corpses'	*navoz* 'manure'	*trupy* 'corpses'
Wood	*legna* 'wood'	*tavole* 'boards'	*brëvna* 'logs'	*ščepki* 'splinters'

1 We recognize that seeing tires as 'containers for air' may be very counterintuitive. However, this choice was driven by the question whether shape similarity to standard containers played a role when deciding category membership for this term. If we look only at the objective, and binary properties of containers (for instance, an object with an empty space inside), then a tire can indeed be included within the category. However, we included it precisely to test whether an interpretation that is logically acceptable can nevertheless be functionally and cognitively unmotivated, and how such an item is identified.

The experiment consisted of two questionnaires, one per language, administered entirely in the respective target language (category labels are reported here in English for clarity and readability). Each questionnaire began with instructions and a short section to collect sociodemographic information (age, gender, nationality, education, experience in linguistics, and first language). Since participation was restricted to individuals over 18 years of age who were L1 speakers of Italian or Russian, this information was essential for verifying eligibility.

Stimuli were presented individually in random order, each accompanied by a list of categories from which participants were asked to select one. If none of the options were deemed appropriate, participants could choose an 'Other' option and propose a new label. To minimize ordering effects and ensure independent judgments, participants were not allowed to return to previous questions.

The Italian questionnaire[3] was created and distributed via LimeSurvey[4] in April 2022. Participants were recruited during university activities (lectures, student meetings, etc.), through personal contacts and social networks. The Russian version was designed using Google Forms,[5] with participants recruited via

3 The Italian questionnaire is available at this link: `http://www.servizididatticiscuolalfl.it/Ling_UniRoma3/index.php/151745?lang=it`.

4 `www.limesurvey.org`.

5 The Russian questionnaire is available at this link: `https://forms.gle/WDxGWzG47ENaX8yw5`.

Prolific;[6] responses were collected in August 2024. In total, we obtained 70 valid responses for Italian and 67 for Russian. For analysis, we retained 67 responses per language to ensure balanced datasets.[7]

2.2 Statistical Methods

We addressed our research questions with three complementary statistical analyses. First, we examined the overall agreement among Italian and Russian participants in assigning items to categories (RQ1) via an inter-rater reliability test, which assesses the extent to which participants consistently assigned a given lexical item to the same category, and whether this agreement exceeded chance level. We used Fleiss' Kappa (Fleiss 1971), computed separately for Italian and Russian, with categories treated as nominal, via the R package *irr* (*kappam.fleiss*). The κ value ranges from 0 (chance agreement) to 1 (perfect agreement), with negative values indicating less-than-chance agreement. We interpret κ following Landis and Koch (1977): 0.00-0.20 slight agreement; 0.21-0.40 fair agreement; 0.41-0.60 moderate agreement; 0.61-0.80 substantial agreement; 0.81-1.00 almost perfect agreement.

To test whether central vs. peripheral category membership predicts expected assignment (RQ2), we fitted a mixed-effects binominal logistic regression using the *glmer* function of the R package *lme4*. The dependent variable was 'match' (1 = expected category; 0 = otherwise; 'Other' is considered as 0) while the predictors included 'centrality' (central vs. peripheral) as a fixed effect and random intercepts for 'participant ID' and 'item'. Separate models were fitted for each language and compared. This test evaluates the probability that a given noun was assigned to the expected category above chance, while accounting for individual variability among participants and items and contrasting central and peripheral members.

To have a more detailed overview of which single items were assigned to their expected category above chance and which, on the contrary, were assigned to different ones, we ran exact binominal tests per item (separately by language), applying a False Discovery Rate (FDR) correction.

Lastly, for non-matching responses, we identified the most frequent choices to determine how informants tended to categorize

6 www.prolific.com.

7 The questionnaire results are available at the following link: https://doi.org/10.17605/OSF.IO/3AQ9W.

these items instead. Through a qualitative description of these cases, we address the issue of which principles guide speakers in assigning category membership to less prototypical items (RQ3).

3 Results and Discussion

3.1 Testing the Role of Item Centrality: Fleiss' Kappa and Mixed-Effects Logistic Regression

The first research question we wanted to address was to what extent Italian and Russian speakers share the same categories both within and between the two groups, and whether they agreed in assigning the items to the intended category. The inter-rater reliability test was run for the purpose and shows that, for both languages, there is a moderate overall agreement among the informants ($0.41 < \kappa < 0.60$). However, Italian participants performed a bit better than Russian ones: $\kappa_{IT} = 0.609$ (z-score = 806; p-value < 0.001); $\kappa_{RU} = 0.572$ (z-score = 761; p-value < 0.001).

Considering the κ value for each category, in the Italian experiment, thirteen categories showed substantial agreement ($\kappa > 0.60$), four of which reached almost perfect agreement ($\kappa > 0.80$): Animals, Buildings and places, Minerals and rocks, and Celestial bodies. As for Russian, ten categories displayed substantial agreement ($\kappa > 0.60$), three of which exceeded $\kappa > 0.80$, namely, Wood, Buildings and places, and Minerals and rocks. For the latter two, high agreement is consistent with Italian participants' rates.

Within the Italian data, only two categories fall into the range of slight agreement ($\kappa < 0.20$), namely, Generic and Powders. The next lowest agreement occurred for Tools ($\kappa = 0.415$) and Human activities ($\kappa = 0.442$). In the Russian data, the same four categories yielded the lowest agreement, with Generic showing the weakest result ($\kappa = 0.181$), followed by Tools ($\kappa = 0.295$), Human activities ($\kappa = 0.306$), and Powders ($\kappa = 0.357$). Thus, speakers of both languages exhibit strong consensus for some categories while showing limited consensus for others. Interestingly, the patterns of low agreement are consistent between the two languages. Overall, a cognitively plausible pattern emerges; categories with the highest agreement tend to be specific, narrow, and concrete, whereas those with lower agreement are generally broader, less specific, and more abstract.

We then complemented Fleiss' Kappa with a mixed-effects logistic regression for each language to examine how agreement varies based on item centrality.

The mixed-effects logistic regression on the Italian data revealed a robust effect of 'centrality' ($\beta = 2.58$, SE = 0.61, $z = 4.20$, $p < 0.001$).

The odds of assigning an item to its expected category were thirteen times higher for central items than for peripheral ones. The intercept was not significant (β = –0.10, SE = 0.43, p = 0.82), indicating that peripheral items did not differ from chance level once variability was taken into account. Random effects indicated that variance was substantially larger across items (σ^2 = 3.66) than across participants (σ^2 = 0.34), suggesting that item difficulty accounted for most of the heterogeneity in classification accuracy.

The same model run on Russian data showed likewise a significant effect of centrality (β = 1.68, SE = 0.48, z = 3.49, p < 0.001) with central items over five times more likely to be correctly assigned than peripheral items (odds ratio = 5.4). As in Italian data, the intercept was not significant (β = 0.01, SE = 0.34, p = 0.97), implying that peripheral items did not differ from chance level once random variation was accounted for. Random effects showed considerably greater variance across items (σ^2 = 2.29) than across participants (σ^2 = 0.14). This pattern indicates that variability was driven primarily by item differences in both languages.

So, to answer RQ2, the centrality of an item with respect to a category is a significant predictor for categorization in both languages. Central items were significantly more likely to be assigned to the expected categories than peripheral ones (thirteen times more likely in the Italian data and five times in the Russian data), which provides experimental evidence of the fuzziness of categories. Such a tendency is visible in Figure 1, which plots the probability of an expected answer (y-axis), as predicted by the logistic regression model, based on the centrality of the item (x-axis, peripheral vs. central).

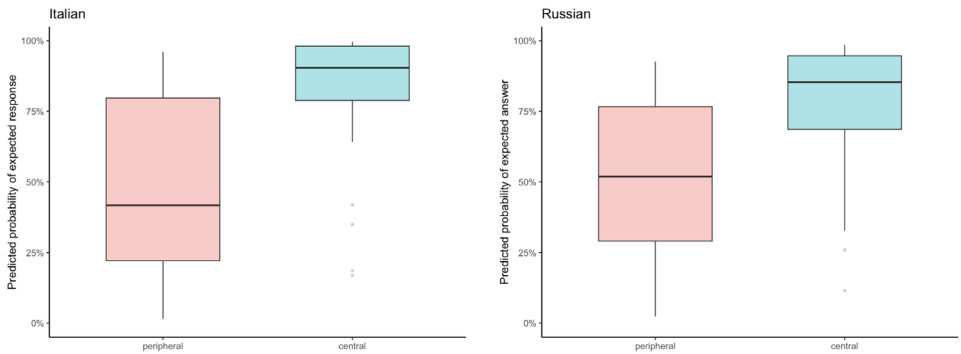

Figure 1 Plots of model-based item probabilities by centrality in Italian (left) and Russian (right) data

In the plots, each box represents the distribution of item-level predicted probabilities for peripheral (red) and central (green) nouns. The clear separation between the two boxes in both languages visually conveys the strong effect of centrality: central items have uniformly high predicted accuracy, clustering at the upper end of the scale, while peripheral items show wide variability with medians close to chance level and several items falling below it.

3.2 A More Detailed View of Participants' Responses: Accuracy Measures and Binominal Test

The results of the logistic regression model were further explored by verifying, through a binominal test, whether the response accuracy for each item is above chance. Figures 2 and 3 display the distribution of response accuracy for central and peripheral items in the two languages (Italian on the left, Russian on the right). The colored bars in the plots represent 95% confidence intervals: narrower bars mean more consistent responses, while wider bars relate to more uncertainty in the estimates. These figures show in more detail what is 'inside' the boxes in Figure 1.

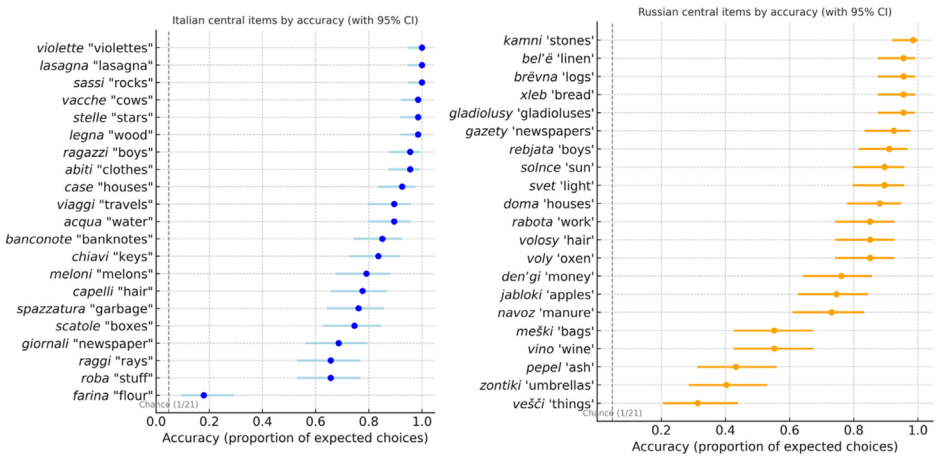

Figure 2 Response accuracy of Italian and Russian central items

In Figure 2, most Italian central items (20/21) show an accuracy above 0.6. Only *farina* 'flour' was assigned to the expected category (Powders) with difficulty. As for Russian, fifteen items were successfully assigned to the expected categories (accuracy > 0.6), whereas six – namely, *meški* 'bags' for Containers, *vino* 'wine' for

Liquids, *pepel* 'ash' for Powders, *zontiki* 'umbrellas' for Tools, and *veščі* 'things, stuff' for Generic – showed more inconsistencies. In both languages, most of the central items can be considered as robust category prototypes, with the Italian dataset performing slightly better than the Russian one. However, some items (one in Italian and six in Russian) that were originally identified as central members were not perceived as such. For instance, Ru. *pepel* 'ash' was mostly categorized (31 responses) as Waste; similarly, It. *farina* 'flour' was very easily categorized (50 responses) as Food.[8] This suggests that these items may be more peripheral within their expected categories and, instead, behave as central items of others.

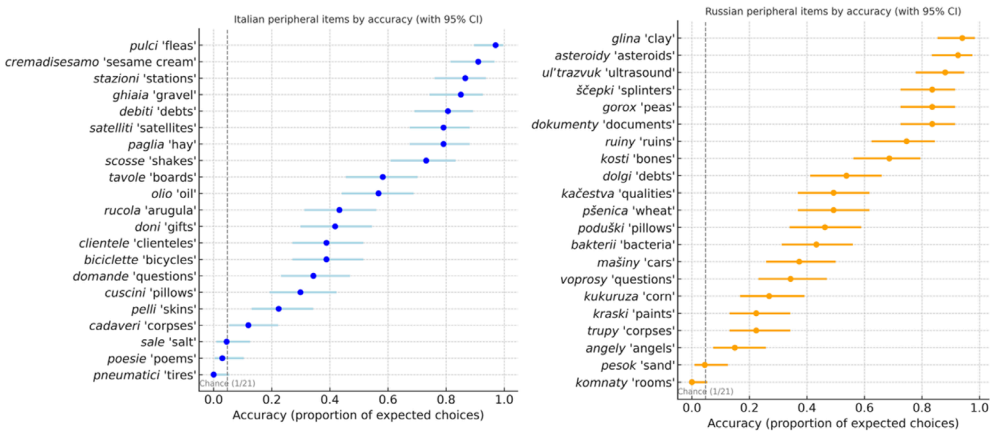

Figure 3 Response accuracy of Italian (on the left) and Russian (on the right) peripheral items

In Figure 3, the scattered distribution of accuracy values attests to the greater variability of peripheral items. In both Italian and Russian, 13 out of 21 items register a rather low accuracy (< 0.6), meaning that participants assigned them to a wider range of categories, thus confirming their peripheral status. However, 8 out of 21 items per language were consistently classified in the expected category (items with accuracy > 0.6). For example, It. *pulci* 'fleas' – although referring to insects – was almost unanimously categorized as Animals, and only 2 participants assigned it to a new category. There could be a double reason for this answer: a) a lack of knowledge that fleas

8 As an anonymous reviewer suggested, this finding supports a dynamic theory of lexical meaning (Pustejovsky 1995). The reclassification of the Italian and Russian items may reflect the greater cognitive salience of certain *qualia* roles in specific contexts. For example, the fact that flour was readily classified as Food shows that its telic role (namely, its purpose) is especially salient.

are insects, which induced speakers to classify them as Animals on the grounds that fleas typically infest mammals and animals, and are living beings; b) the respondent was too 'lazy' to propose a new category and opted for the closest possible one. In the same way, Ru. *glina* 'clay' was successfully assigned to Minerals and rocks by the majority of participants, though it could have also activated other categorizations and been assigned to Liquid or Containers (1 response for the latter), which would probably be the case if the noun were framed under a specific context. These results again question the *a priori*-defined status of these items as peripheral within their expected categories.

3.3 What Principles Drive Categorization?

Let us now look at all original peripheral items with accuracy lower than 0.6 to see how they were categorized. These items can be divided into three groups, both in Italian and Russian, based on the statistical significance of their accuracy as calculated by a binominal. In the first group, reported in Table 2 (Italian items in green and Russian ones in orange), accuracy is near zero and not statistically significant ('Above chance' column = FALSE), indicating that no one or almost no one chose the expected category, while most selected a new dominant one.

Table 2 List of Italian (green) and Russian (orange) near-zero accuracy items

	Item	Expected category	Accuracy	Dominant unexpected category	Dominant unexpected category count	Above chance (FDR)
Italian	*pneumatici* 'tires'	Containers	0	Tools	38	FALSE
	poesie 'poems'	Paper and related materials	0.029	Human activities	40	FALSE
	sale 'salt'	Powders	0.044	Minerals and rocks	33	FALSE
Russian	*komnaty* 'rooms'	Containers	0	Buildings and places	65	FALSE
	pesok 'sand'	Powders	0.044	Minerals and rocks	58	FALSE

The classification of these items allows us to reflect on the principles that drove their inclusion into the new categories. In principle, It. pneumatici 'tires' and Ru. *komnaty* 'rooms' may share some physical features with Containers, i.e., having a space enclosed by rigid walls, with an opening and closing point; and functional similarity (wheels as containers of air, rooms as containers of people). Nevertheless, the

experimental data show that these similarities to Containers were overridden by more salient, functional, or experiential attributes. The categorization of *pneumatici* as Tools (56.7% of answers) is probably driven by their relationship with cars: 23,8% of informants suggested their own category for *pneumatici*, defining them as "part of a vehicle".[9] Similarly, the inclusion of Ru. *komnaty* in Buildings and places reflects the identity of rooms as spatial parts of a larger structure.

This demonstrates that, when multiple attributes are available, the basic-level categorization is determined by the most salient and frequent feature in real-world human interaction, in line with the cognitive economy principle of basic categorization (Rosch 1978). The relationship between a member and its category is usually a taxonomical one ("X is a Y", e.g, *salt is a mineral*), but such a relation is established following some specific principles. For *pneumatici* and *komnaty*, two might apply, namely, 'purpose or function' and 'meronymy' (i.e., part-whole relationship) principles. In addition, the items It. *sale* 'salt' and Ru. *pesok* 'sand' were classified as Minerals and rocks by virtue of the people's knowledge that both belong to that domain of substances, overriding the physical resemblance to entities with atomic particles that could justify their membership in Powders. In such cases, we could say that people rely on an 'encyclopedic' principle according to which items are categorized based on intrinsic properties, such as constitution, material, origin, or source. Finally, It. *poesie* 'poems' is classified not as an object (like paper) but as a more abstract Human activity (59.7%). This example shows that for some items, categorial membership is based on an abstract feature more than a concrete one, hence on what we could define as a 'concrete vs. abstract' principle.

The second group, reported in Table 3, includes items with accuracy above zero that is statistically significant but still very low.

9 Hence, referring back to the question raised in footnote 3, although a tire could in principle be construed as a 'container for air' or a pillow as a 'container for feathers', which indeed reflects their essential nature, these potential properties are simply ignored and do not contribute to determining their proper category membership. The classification, therefore, takes place according to both cognitive and logical parameters, not solely on the basis of fixed physical characteristics.

Table 3 List of Italian and Russian items with statistically significant accuracy above zero

	Item	Expected category	Accuracy	Dominant unexpected category	Dominant unexpected category count	Above chance (FDR)
Italian	cadaveri 'corpses'	Waste	0.119	Human beings	42	TRUE
	pelli 'skins'	Body parts	0.223	Animals	29	TRUE
	cuscini 'pillows'	Clothing and linen	0.298	Tools	20	TRUE
	domande 'questions'	Human activities	0.343	Generic	19	TRUE
Russian	*angely* 'angels'	Human beings	0.149	OTHER	17	TRUE
	kraski 'paints'	Liquids	0.223	Tools	23	TRUE
	trupy 'corpses'	Waste	0.223	Human beings	33	TRUE
	kukuruza 'corn'	Food	0.268	Fruits, vegetables, and legumes	39	TRUE
	voprosy 'questions'	Human activities	0.343	Generic	29	TRUE

Here, the expected category is recognized as a good fit but competes with one or more preferred ones, ranking second or lower. The categorization of Italian and Russian items in this group reveals speakers' preference for functional or encyclopedic source-based principles over meronymy. As an example, here are the frequency tables for It. *pelli* 'skins' and Ru. *trupy* 'corpses'.

Table 4 Frequency table of answers for It. pelli and Ru. trupy

Item	Category	Raw fq.	Item	Category	Raw fq.
It. *pelli* 'skins'	Animals	29	Ru. *trupy* 'corpses'	Human beings	33
	Body parts	15		Waste	15
	Clothing and linen	10		Animals	5
	Generic	7		OTHER	5
	OTHER	3		Human activities	3
	Tools	1		Generic	3
	Human beings	1		Body parts	3
	Waste	1			

For both items, the expected category (Body parts and Waste) is the second-highest rated. For It. *pelli* 'skins', the dominant choice is Animals (29 responses), along with the 12 responses for Clothing and linen, suggesting that the plural form of *pelli* activates a commercial context, prioritizing the item's origin and use over its anatomical

status. This preference for a functional principle is echoed in the reclassification of It. *cuscini* 'pillows' and Ru. *kraski* 'paints' as Tools.

As for 'corpses' (It. *cadaveri*, Ru. *trupy*), the speakers' dominant choice is Human beings (It. 42 vs Ru. 33 rates) rather than Waste. In this case, too, a functional principle is at stake. Corpses are less likely to be viewed as waste because the link to human identity maintains its importance even after death. A similar cultural value emerges with Russian *angely* 'angels', where the relevance of the link to the religious sphere forces the speaker's choice toward the 'Other' option.

Lastly, in the case of It. *domande* and Ru. *voprosy* 'questions', speakers of both languages chose Generic over Human activities. This gives a hint of how ambiguity is handled in both languages: speakers appear to use Generic as an 'escape' category to resort to when unsure about their answer. In Italian, this is evident in the categorization of *cadaveri* where Generic ranked second (9 responses), followed by the expected Waste category (8 responses).

Finally, the third group includes items with higher accuracy but fewer dominant unexpected categories, as reported in Table 5.

Table 5 List of Italian and Russian peripheral items with higher accuracy

	Item	Expected category	Accuracy	Dominant unexpected category	Dominant unexpected category count	Above chance (FDR)
Italian	*biciclette* 'bicycles'	Tools	0.388	OTHER	19	TRUE
	clientele 'clienteles'	Human beings	0.388	Human activities	28	TRUE
	doni 'gifts'	Generic	0.417	Human activities	35	TRUE
	rucola 'arugula'	Fruit, vegetables, legumes	0.432	Food	24	TRUE
	olio 'oil'	Liquids	0.567	Food	28	TRUE
	tavole 'boards'	Wood	0.582	Generic	13	TRUE
Russian	*mašiny* 'cars'	Tools	0.373	Human activites	16	TRUE
	bakterii 'bacteria'	Animals	0.432	OTHER	14	TRUE
	poduški 'pillows'	Clothing and linen	0.462	Generic	12	TRUE
	kačestva 'qualities'	Generic	0.492	Human beings	12	TRUE
	pšenica 'wheat'	Flowers, herbs, and plants	0.492	Food	16	TRUE
	dolgi 'debts'	Money	0.537	Human activites	20	TRUE

For *tavole*, the expected Wood category (41.1%) follows the encyclopedic principle (source/material). However, this choice competes with Generic (20.8%) – the option we identified as an index of uncertainty – and Tools (19.4%), which reflects a choice guided by the functional principle. A similar pattern appears with food items such as It. *olio* 'oil' and *rucola* 'arugula', where the expected source-based category (Liquids and Fruit, vegetables, and legumes, respectively) competes with the functional property of edibility (Food). Finally, for vehicles (It. *biciclette* 'bicycles' and Ru. *mašiny* 'cars'), the expected Tools category competes with participants' suggested options, related to function or motion: Russian choices were split between Human activities (23.8%), justified by the action of driving, and a newly proposed category Means of transportation (28.3%). Italian speakers proposed the Means of transportation option, too, confirming a cross-linguistic preference for a functional domain over a more generic classification.

Finally, it is worth noticing that this group includes more complex abstract items, such as It. *clientele* 'clienteles' and *doni* 'gifts', or Ru. *dolgi* 'debts' and *kačestva* 'qualities'. Here, a double competition arises: first, between the concrete, expected category and an abstract, function-based one (e.g., *clientele* 'clienteles' and *dolgi* 'debts', where the concrete source-related category of Human beings and Money, respectively, is paralleled by the abstract, function-based class Human activities); second, between the generic, expected class and a specific one, as for *doni* 'gifts' and *kačestva* 'qualities' classified as both Generic (expected category) and Human beings or Human activities.

To answer RQ3, our experiment allowed for the identification of a set of principles that drive speakers' categorization: (i) a functional principle grouping items by purpose; (ii) a meronomy principle identifying the item as a part of a given whole; (iii) an encyclopedic principle that relies on shared world knowledge about the intrinsic properties of the item to define it; and (iv) an abstraction principle favoring abstract and general over concrete categorization. However, the qualitative nature of this interpretation implies the need for additional experimental evidence to support and validate these proposed principles and check for a possible hierarchy among them. Although no clear-cut tendencies emerged from the present results, future research could explicitly test this hypothesis by examining whether certain properties (e.g., *qualia* or similar ontological properties) emerge as more salient than others under controlled conditions.

4 Conclusion

In this paper, we aimed to verify the validity of a semantic classification conceived for the analysis of binominal constructions as classifier-like strategies in non-classifier languages. To do so, we ran a categorization experiment with 67 speakers of Italian and Russian, providing them with a list of items to categorize within 21 pre-defined categories. By examining the experiment results, we have demonstrated that the proposed categories were successfully recognized by speakers, as they generally found them apt to classify the given items. Through quantitative methods, we have demonstrated that: (i) participants are sufficiently consistent in classifying items in both languages, with a slightly better performance by L1 Italian speakers (RQ1); (ii) there is a statistically significant difference in the categorization of central vs. peripheral members of the proposed categories: despite some exceptions, central items were overall more easily categorized within the expected category than peripheral items in both languages (RQ2). Lastly, by looking at participants' responses in a more qualitative way, we were able to identify a set of principles that seem to guide categorial membership, namely, the principle of *function*, the principle of *meronymy* (or part-whole relationship), the *encyclopedic knowledge* principle, and the *concrete vs. abstract* principle (RQ3). Future research should include the collection of further evidence for such principles and the submission of a wider set of items on a larger informant base.

Bibliography

Aikhenvald, A.Y. (2003). *Classifiers: A Typology of Noun Categorization Devices*. 3rd ed. Oxford: Oxford University Press. http://doi.org/10.1093/obo/9780199772810-0007.

Allan, K. (1977). "Classifiers". *Language*, 53(2), 285-311. https://doi.org/10.2307/413103.

Baker, M. (2011). *In Other Words: A Coursebook on Translation*. 2nd ed. London: Routledge. https://doi.org/10.4324/9780203832929.

Benigni, V. (2022). "Binominal Constructions with Metaphorical Quantifiers in Russian: *Vsplesk Èmocij i Volna Pozitiva*". *Studi Slavistici*, 19(2), 169-92. https://doi.org/10.36253/studi_slavis-12323.

Benigni, V.; Latos, A. (2023). "Una Montagna di Errori: Costruzioni Binominali con Classificatori Metaforici". *Studia Universitatis Hereditati Znanstvena Revija Za Raziskave in Teorijo Kulturne Dediščine*, 11(1), 11-31. https://doi.org/10.26493/2350-5443.

Benigni, V.; Latos, A. (2024). "Metaphorical Binominal Constructions in the Domain of Water: A River of Words. Evidence from Italian, Polish and Russian". Baicchi, A.; Broccias, C. (eds), *Constructional and Cognitive Explorations of Contrastive Linguistics*. Cham: Springer International Publishing, 129-51. https://doi.org/10.1007/978-3-031-46602-1_8.

Cinque, G. (2006). "Are All Languages Numeral Classifier Languages?". *Rivista di Grammatica Generativa*, 31, 119-22.

Clark, E.V. (2017). "Semantic Categories in Acquisition". Cohen, H.; Lefebvre, C. (eds), *Handbook of Categorization in Cognitive Science*, 397-421. https://doi.org/10.1016/b978-0-08-101107-2.00017-8.

Cotta Ramusino, P. (2016). "A Proposito di Quantificatori Indefiniti in Polacco". Benigni, V.; Gebert, L.; Nikolaeva, J. (a cura di), *Le Lingue Slave tra Struttura e Uso*. Firenze: Firenze University Press, 79-92. https://doi.org/10.36253/978-88-6453-328-5.05.

Haspelmath, M. (2010). "Comparative Concepts and Descriptive Categories in Crosslinguistic Studies". *Language*, 86(3), 663-87. https://doi.org/10.1515/9783110607963-004.

Lacroce. (2023). "Linguistic Classification and its Mechanisms: The Case of Binominal Constructions in Non-Classifier Languages" [Unpublished D. Phil. Dissertation]. Rome: University of Roma Tre.

Landis, J.R; Koch, G.G. (1977). "The Measurement of Observer Agreement for Categorical Data". *Biometrics*, 33(1), 159-74. https://doi.org/10.2307/2529310.

McEnery, T.; Xiao, R. (2007). "Quantifying Constructions in English and Chinese: A Corpus-Based Contrastive Study". *Proceedings of the Corpus Linguistics Conference CL2007*. UK: University of Birmingham.

Mervis, C.B.; Rosch, E. (1981). "Categorization of Natural Objects". *Annual Review of Psychology*, 32, 89-115. https://doi.org/10.1146/annurev.ps.32.020181.000513.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge: MIT Press. https://doi.org/10.7551/mitpress/3225.001.0001.

Rosch, E. (1978). "Principles of Categorisation". Rosch, E.; Lloyd, B.B. (eds), *Cognition and Categorisation*. Hillsdale: Lawrence Erlbaum, 27-48. https://doi.org/10.4324/9781032633275-4.

Sussex, R. (1976). "The Numerical Classifiers of Russian". *Russian Linguistics*, 3(2), 145-55. https://doi.org/10.1007/bf03545839.

Tai, J.; Wang, L. (1990). "A Semantic Study of the Classifier *tiao*". *Journal of the Chinese Language Teachers Association*, 25(1), 35-56.

Taylor, J.R. (2003). *Linguistic Categorization*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199266647.001.0001.

Verveckken, K.D. (2015). *Binominal Quantifiers in Spanish: Conceptually-Driven Analogy in Diachrony and Synchrony*. Berlin; Boston: Mouton de Gruyter. https://doi.org/10.1515/9783110406733.

Xu, Z. (2017). *English Quasi-Numeral Classifiers: A Corpus-Based Cognitive-Typological Study*. Bern: Peter Lang. https://doi.org/10.3726/b10934.