Many hands on this study: Development of a metonymy comprehension task

Maity Siqueira, Tamara Melo, Sergio Duarte Júnior, Laura Baiocco, Caroline Girardi Ferrari, Nichele Lopes

PUC-SP

DELTA

Many hands on this study: Development of a metonymy comprehension task

Muitas mãos neste trabalho: desenvolvimento de uma tarefa de compreensão de metonímia

Maity Siqueira[1]
Tamara Melo[2]
Sergio Duarte Jr[3]
Laura Baiocco[4]
Caroline Girardi Ferrari[5]
Nichele Lopes[6]

Abstract

Metonymy is a pervasive linguistic/cognitive mechanism in everyday language. Still, studies assessing its comprehension are rare in the literature. Our goal is to present the elaboration of a verbal metonymy comprehension task in Brazilian Portuguese as well as test it using pilot studies. The task was based on the cognitive linguistics framework and psychometric guidelines were considered. The construction of the task considered psycholinguistic variables and included a conventionality task and an expert analysis. Three studies ($N = 290$, ages ranging 2 to 54 years old) were conducted until we reached a final version. The study resulted in a task composed of one practice item and six test items. Each item includes a sentence-stimulus and two questions (one open-ended and one closed-ended). Together, the items make up a raw score that represents metonymy comprehension. Data obtained from the pilot studies show that metonymy comprehension improves gradually with age and that there is a significant difference between items in both types of questions. This is in consonance with the literature, and thus points towards the instrument's construct validity. The study results in a task that presents enough sensitivity to gauge comprehension of this phenomenon at different ages.

Keywords: metonymy, metonymy comprehension, figurative language, psycholinguistic task elaboration, cognitive linguistics

¹ Federal University of Rio Grande do Sul. Porto Alegre – Brazil.

https://orcid.org/0000-0002-8775-4563. E-mail: maity.siqueira@ufrgs.br

² Federal Institute of Santa Catarina, Itajaí – Brazil.

https://orcid.org/0000-0001-9545-5355. E-mail: tameoli@gmail.com

³ Federal University of Rio Grande do Sul. Porto Alegre – Brazil

https://orcid.org/0000-0002-8641-7338. E-mail: sergioduartejunior@yahoo.com.br

⁴ The University of Arizona. Tucson – USA

https://orcid.org/0000-0002-5221-257X. E-mail: laurabaiocco@arizona.edu

⁵ Federal University of Rio Grande do Sul. Porto Alegre – Brazil.

https://orcid.org/0000-0003-3853-3673. E-mail: carolinegirardiferrari@yahoo.com.br

⁶ Federal University of Rio Grande do Sul. Porto Alegre – Brazil.

https://orcid.org/0000-0002-3424-9387. E-mail: lopesnichele@gmail.com

Resumo

Metonímia é um mecanismo linguístico/cognitivo pervasivo na linguagem cotidiana. Entretanto, estudos de avaliação desse fenômeno são escassos. Nosso objetivo é apresentar a elaboração de uma tarefa de compreensão de metonímia em português brasileiro e seus resultados piloto. Referenciais teóricos da linguística cognitiva e da psicometria foram considerados. A construção da tarefa considerou variáveis psicolinguísticas e incluiu uma tarefa de convencionalidade e análise de juízes. Três estudos ($N = 290$, entre 2 e 54 anos) foram conduzidos até a versão final da tarefa. O estudo resultou em uma tarefa com um item de treino e seis de teste. Os itens incluem uma frase de estímulo e duas perguntas (aberta e fechada). Consideramos que o escore dos itens representa a compreensão de metonímia. Os resultados dos estudos piloto indicam uma melhora gradual na compreensão de metonímia com a idade e uma diferença significativa entre os itens nas questões abertas e fechadas. Isso está em consonância com a literatura, sugerindo evidências de validade de construto ao instrumento. O estudo resulta em um instrumento que apresenta sensibilidade suficiente para aferir a compreensão do fenômeno em diferentes idades.

Palavras-chave: metonímia, compreensão de metonímia, linguagem figurada, elaboração de tarefa psicolinguística, linguística cognitiva

1. Introduction

As well established by cognitive linguistics (CL), figurativeness is a pervasive phenomenon not only in language, but also in thought and action. In daily speech, we often use metonymies, metaphors, irony, and so on, mostly without even noticing. When someone says *'We need more hands for this job'*, for instance, one is not literally asking for hands, but for help to do something. When we say on an occasion that the atmosphere was tense, we are not actually referring to tension or the atmosphere, but to an uncomfortable situation. We also repeatedly use other tropes on a daily basis. Yet, the prevailing notion, not only for the lay public, but also for many psychologists and linguists, is that language is still predominantly literal. Figurative phenomena are, therefore, fundamental pieces for language comprehension and use. The goal of this article is to describe the elaboration of a metonymy comprehension task.

The instrument we describe here was developed according to the perspective of CL and following psychometric guidelines. This is to say that our understanding of metonymy stems from CL and that procedures carried out in task elaboration were designed by the area in psychology that addresses test elaboration. Under the CL approach, figurative language is understood as a core mechanism for language and thought. Therefore, metonymy is a cognitive process in which there is a unidirectional mapping from one conceptual domain (the source) to another (the target) within the same cognitive model. First, let us try to delimit these concepts. "A conceptual domain is any coherent organization of experience" (Kövecses, 2010, p. 4). Cognitive models, on the other hand, are somewhat stable mental representations that mirror how we understand reality. Such cognitive models are considered idealized — they do not instantiate specific occurrences of a given experience but are created by means of generalizations of sets of experiences (Lakoff, 1987).

Traditionally, when metonymy is brought up, one thinks of figures of speech. From Greek rhetoric to more recent accounts, metonymy has been basically "regarded as a *stand for* relation in which the name of one thing is used to refer to another thing" (Panther & Thornburg, 2007, p. 237). In CL, however, metonymies are primarily conceived of as conceptual mappings (Barcelona, 2000), which can be expressed through verbal language, among other forms of communication (such as gestures or images). Note that what CL adds to the definition of metonymy is the idea of a cognitive domain (a conceptual mapping), beyond

a purely linguistic phenomenon (the substitution of a name). Despite the fact that CL considers two dimensions (a conceptual and a linguistic one), in the instrument we describe here, the linguistic facet is the perceptible one[1]. A sentence such as *'I've never even seen his face'*, for instance, has a linguistic expression derived from the conceptual metonymy PART FOR WHOLE. This conceptual mapping can also be identified in non-linguistic instantiations, as is the case when someone asks their friend to show them someone's photo and is satisfied with a picture of the person's face. PART FOR WHOLE (in this case, FACE FOR THE PERSON) is, in fact, a typical example of metonymy (Figure 1).

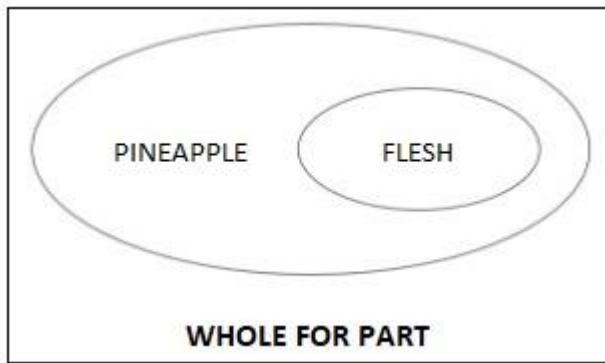**Figure 1** – Conceptual metonymy PART FOR WHOLE
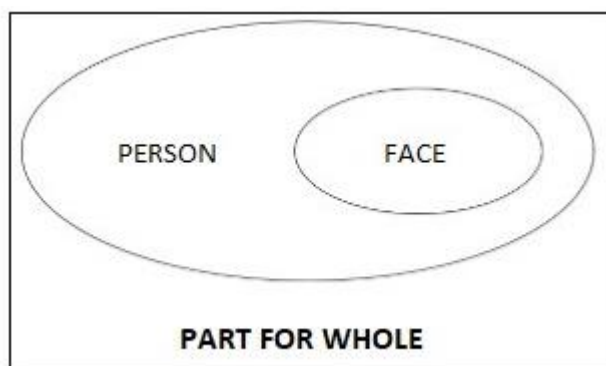
Still, there are other kinds of metonymy that express different relations between the domains: such as WHOLE FOR PART and PART FOR PART. Few people would easily recognize that the sentence *'My son ate a whole pineapple today'* contains a figurative expression. Yet it is understood that the child has eaten all edible parts of the pineapple, not its peel nor its crown. In this case, we are facing a WHOLE FOR PART metonymy (Figure 2).

**Figure 2** – Conceptual metonymy WHOLE FOR PART

---

[1] This approach, unlike traditional ones, allows us to have access to underlying conceptual mappings, which allowed us to control for this potential confounding variable during item elaboration. For example, we have avoided including items with the same conceptual mappings. Otherwise, we would be testing only this specific mapping. Nevertheless, what is tested is the linguistic facet, not the mappings, and therefore the task need not be carried out by someone who is aligned with CL.

An example of the PART FOR PART metonymy (more specifically, CONTAINER FOR CONTENT) would be '*My son never finishes his plate*', which refers to the food that is on the plate and not the plate itself (Figure 3).

Figure 3 – Conceptual metonymy PART FOR PART
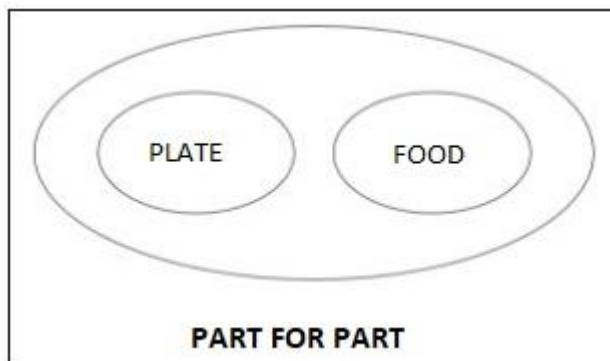
Although it is such a frequent phenomenon in language, metonymy is not as broadly investigated as metaphor and other figurative phenomena. We have been able to find two articles that assess its processing, using the eye-tracking technique. Frisson and Pickering (1999) have found that familiar metonymies are comprehended at a very similar rate to literal expressions by young adults. Even though their investigation regards adult speakers and a divergent empirical method in relation to ours, this could be an indication of an effortless ability for metonymy that launches early in the life cycle. More recently, Köder and Falkum (2020) have compared an online eye-tracking task and a picture selection task and found an early sensitivity to metonymy from the age of three and a gradual improving capacity with age, in a U-shaped curve. In another eye-tracking study, Xie (2022) corroborated their

findings and also found evidence for a U-shape tendency, suggesting a literal preference by 4-and 5-year-olds.

Regarding children's comprehension, some psycholinguistic studies have been developed to assess metonymy comprehension, mostly in the English language. Among these works, Nerlich, Clark and Todd (1999), Rundblad and Annaz (2010b), Falkum, Recasens, and Clark (2017), and Zhu (2021) investigate metonymy comprehension by typically developing individuals. These studies are the ones that most resemble the format of the test presented here, in that they address metonymy comprehension in the first language and by typically developing subjects by means of a similar method as ours. These studies are unanimous in showing that metonymy is comprehended at a fairly early stage of figurative language acquisition and its comprehension improves with age. Falkum et al. (2017) report an established competence to not only comprehend, but also produce and explain metonymy by age three. As for context, Nerlich, Clark and Todd (1999) state that when clues are present, young children tend to comprehend metonymies better. In the same vein, Frisson and Pickering (2007) reinforce the importance of context for the comprehension of unfamiliar metonymies. Zhu (2021), when studying a specific mapping (PRODUCER FOR PRODUCT), states that 4- and 5-year-olds are able to understand metonymies even when they do not know its reference, that is, the producer. Her data indicates that a specific cultural-linguistic knowledge is not necessary for the acquisition of an abstract PRODUCER FOR PRODUCT generalization.

Other studies address metonymy comprehension in clinical populations (mainly Autism Spectrum Disorder and Williams Syndrome). Rundblad and colleagues, for example, found that performance of such groups in metonymy comprehension is delayed in comparison to typically developing children (Rundblad & Annaz, 2010a; Van Herwegen, Dimitriou & Rundblad, 2013; Van Herwegen & Rundblad, 2018). Moreover, they found that for Williams Syndrome chronological age was not the best predictor for performance of novel metonymy; semantic knowledge instead was considered to be a better predictor for metonymy comprehension. Zheng, Jia, and Liang (2015) results corroborate the idea that metonymy comprehension is poorer in high-functioning autistic Chinese-speaking children than in typically developing ones. Unlike other studies mentioned herein, Melogno and Pinto (2014) outlined intervention activities that addressed difficulties in understanding metonymy usages in children with high-functioning Autism Spectrum Disorder. According to the authors,

interventions of this kind can improve their communicative competencies. In addition, Theodoropoulou (2021) analyzed the metonymy BODY PART FOR ITS FUNCTION in a case study with a 4-year-old boy with a speech disorder. She verified that a child at this age can produce metonymies as motivated by his interpersonal experience (i.e., his relationship with his mother).

Using different methods, most of the aforementioned studies compare metonymy to metaphor comprehension. With eyes on the possibility of having a standardized way to test metonymy individually as well as comparing it to other major figurative language phenomena in Brazilian Portuguese, we have devised the task presented here. In Brazil, even respected language assessment tools (e.g., MAC Battery[2], Montreal-Tolouse Language Assessment Battery[3], and Neupsilin[4]) which, in some manner, address figurative language comprehension, do not have specific ways or the goal of assessing metonymy.

A Psychometric test that intends to investigate the typical age of acquisition of each phenomenon separately would be useful both for theoretical matters or clinical and educational settings. By establishing a typical age of acquisition and age parameters, we would ultimately be able to assess comprehension in different populations (ages, backgrounds, clinical conditions) with the resulting instrument. Surprisingly, to date, there was no known standardized test in any language that assesses the comprehension of the main figurative language phenomena.

The objective of this article is to present the elaboration, sample items, and pilot data from the final product of a metonymy comprehension task. Said task is part of a comprehensive figurative language comprehension test (titled COMFIGURA), under development in Brazilian Portuguese and soon to be translated and adapted to English. COMFIGURA will

---

[2] Known in Brazil as 'Bateria MAC' (Bateria Montreal de Avaliação da Comunicação), it is an adapted Brazilian version of the original Canadian instrument Protocole Montréal d'Évaluation de la Communication and can be translated as Montreal Communication Evaluation Battery.

[3] Known in Brazil as 'MTL - Brasil' (Bateria Montreal-Tolouse de Avaliação da Linguagem), it is an adapted Brazilian version of the original Canadian instrument Protocole Montréal-Toulouse d'examen linguistique de l'aphasie.

[4] Brazilian instrument for neuropsychological assessment (Instrumento de Avaliação Neuropsicológica Breve).

feature tasks on metaphor, idioms, proverbs, and irony, in addition to metonymy, which is presented here[5].

2. Methods

Preparation of the Instrument

The elaboration of the task described here followed the psychometric model proposed by Pasquali (2010). This model is composed of the following procedures: (1) theoretical, (2) empirical or experimental, and (3) analytical or statistical. The first step, the theoretical one, consists of reviewing the state of the art regarding the construct (i.e., metonymy) in a specific theoretical approach (i.e., cognitive linguistics), providing operational definitions, and designing the instrument up to the pilot study. Following the pilot study, the second step is the empirical procedure. This is the moment when the experimental procedures per se are conducted, when the data are obtained. The final procedure regards data analysis. Considering that the objective of this article is to present only theoretical procedures, we will describe the method up to, and including, the pilot stage, since, for Pasquali (2010), the first procedure ends at this point. Of note, the theoretical procedures shall include statistical analyses regarding data from the pilot.

To carry out the theoretical procedures, the first step was to define the object of study, its attributes, and the attributes' dimensions and conceptualization, achieved through literature research. Considering the literature (Panther & Thornburg, 2007; Barcelona, 2000), metonymy is composed of two facets: conceptual and linguistic, as previously mentioned. It is possible to attribute a dimension of conventionality to its conceptual facet, and a dimension of familiarity to its linguistic facet. Thus, a conceptual metonymy (such as PRODUCER FOR PRODUCT) can be more or less conventional in a given linguistic community. Conventionality in this case refers to the strength of the association between the two domains involved (PRODUCER and PRODUCT, in this example). On the other hand, linguistic metonymies that manifest the same conceptual metonymy can be more or less familiar (i.e., frequent). We understand that the degree of familiarity is the result of the frequency in use of

---

[5] Since the whole test is going to be commercialized, the full version of the metonymy task will not be presented in this article. The reader who wants full access to the task may get in touch with the authors.

an expression at a certain time in a specific community an individual belongs to. In England, for example, *'I love reading Shakespeare''* is probably more familiar to an adult than *'I'll read Nikolai Leskov'*, even though both express the very same conceptual metonymy. In this sense, although metonymies are constituted from conceptual mappings, we can say that only its linguistic aspect is verifiable.

Even though metonymy is a verifiable phenomenon in linguistic expressions, its frequency is not easily measured, as opposed to other phenomena such as idioms and proverbs. This difficulty is due to it allowing a great number of references to and combinations between linguistic elements. Some metonymic uses are quite frequent because of our daily activities, but it is virtually impossible to verify their frequency. Take the conceptual metonymy OUTFIT FOR PERSON as an example. It can generate linguistic expressions such as *'The white shirt is waiting'*, *'The yellow shirt is waiting'* — and so on — and many other combinations between clothes and any other attributes (*'The ripped shirt is waiting', 'The silver tie is dancing', 'The white shirt is here'* etc.). On the other hand, the idiom *paint the town red* and the proverb *Never judge a book by its cover* have more stable structures and are more easily identifiable.

It is important to emphasize that, unlike other phenomena, metonymy cannot be identified in isolation. The phrase 'white shirt', for instance, does not constitute a metonymy by itself. To be identified as such, it needs to be in a sentence, as in *'the white shirt is waiting'*, or in the context of an utterance, such as when someone calls someone else '*'hey, white shirt'*. The expression *paint the town red* alone, in turn, constitutes an idiom by itself. Because of this characteristic, frequency (and therefore the dimension of familiarity) cannot be objectively verified. We have overcome this difficulty by brainstorming in the research group with the purpose of selecting linguistic metonymies that are presumably familiar.

After selecting familiar expressions, we have developed a psycholinguistic conventionality task about their conceptual mappings. We used the same mappings we had on the comprehension task and presented examples of other sentences in which they may appear. Through 5-point Likert scales, in which 1 meant not conventional and 5 meant totally

conventional, 125 participants[6] answered how conventional they thought the mappings were. Our results allowed us to rank the expressions by its conventionality rate, confirming our intuition and what we see in the literature (Panther & Radden, 1999). Rates on the conventional mappings for points 4 (very conventional) and 5 (totally conventional) on the scale accounted for 49.6% (OUTFIT FOR PERSON), 68.8% (INSTRUMENT FOR ACTION), 69.6% (PRODUCER FOR PRODUCT), 79.2% (PART FOR WHOLE), and 80.8% (CONTAINER FOR CONTENT[7]). This way, besides being familiar, linguistic metonymies chosen for the task also expressed conventional conceptual mappings in Brazil today.

In addition to these dimensions of conventionality and familiarity, the following criteria were considered during the elaboration of the task items: (1) lexical complexity[8] (semantics), (2) syntactic structure, (3) sentence length, (44) sorts of topics, (5) absence of context, and (6) absence of other figurative language phenomena. With respect to lexical complexity, we decided not to include words that could be difficult especially for kids to understand, such as *ethics* and *acumen*. Regarding syntactic structure, we created monoclausal active voice sentences, with simple subjects, following the canonical subject-verb-object order of Portuguese. As to sentence size, all sentences contain seven words. In relation to the sorts of topics, we tried to address timeless subjects (a birthday party and not a trendy pop singer), non-controversial subjects (classic literature and not abortion), and subjects accessible to the young public (sports and not politics). Moreover, we did not include trademarks. In terms of context, we chose to present decontextualized sentences so that the pattern of the other tasks in COMFIGURA was followed. Considering context can be a facilitator for figurative language comprehension, a test that does not provide contextual clues requires the participants to know the phenomena themselves in each case. Regarding the last criterion, which is more specific to figurative language, we also tried to isolate the phenomena since the main purpose

---

[6] Note that participants were not experts in linguistics, which might have prevented a better understanding of what a conventional association between domains is. Also note that the so-called synecdoche, the prototypical metonymic mapping, is amongst the mappings considered more conventional.

[7] We presented five instead of six mappings in the conventionality task because CONTAINER FOR CONTENT was instantiated twice in the instrument.

[8] Because we do not have norms in Brazilian Portuguese to establish semantic complexity, we made sure that most of the words used in the task were concrete nouns (camisa [shirt], prato [plate], teto [roof], garrafa [bottle], mão [hand]). Those which were not concrete, were very well known words in Portuguese, presumably heard from early on, such as jogo [game], trabalho [work] and amanhã [tomorrow].

of the test is to discriminate between the comprehension of different tropes. Metaphor was the phenomenon that required the most attention, due to its being the most pervasive one.

Once the criteria for elaborating the sentences that would make up the stimuli were defined, we turned to the structuring of the task. To standardize COMFIGURA, we opted to reproduce the structure that was previously used on metaphor (Siqueira, 2004) and idiom (Siqueira & Marques, 2018) tasks. First of all, the number of items was maintained (six). It is important to emphasize that this number of items was directly determined by the total number of items on COMFIGURA[9]. Our concern, in this case, was twofold: the chance of participant fatigue and the capacity of generalization of the item. More specifically, the higher the number of questions, the higher the chances of fatigued participants. On the other hand, the lower the number of items, the lower the capacity of generalization over the task results. Thus, we have decided to include a number of items that, though small, could provide evidence of construct validity. Construct validity is here defined as the extent to which the instrument measures the theoretical construct at hand (Anastasi & Urbina, 1997).

In addition to the number of items maintained from the previous tasks, we have also kept the number of questions asked (two), the structure of the questions (one open-ended and one dichotomous for each sentence) and the quality (the criteria (1) to (6) above adopted for elaborating the stimuli). It is worth mentioning that the metaphor task has already been validated with clinical (De Leon, Siqueira, Parente & Bosa, 2007) and non-clinical populations (Siqueira & Lamprecht, 2007), in Brazil and in the United States of America (Siqueira & Gibbs, 2007), with participants of different age levels (children, teenagers and adults) and cultures (Siqueira, Parente & Gil, 2009). Thus, we consider this test structure to be adequate to assess the comprehension of each of the phenomena in question.

As we have seen, the items in the task are composed of one sentence and two questions, one open-ended and one closed-ended, presented in this order. The open-ended question allows the participants to express their answers freely considering the given stimulus and, thus, is more interesting for qualitative analyses. Even though this open-ended question is richer in

---

[9] Considering the total amount of items in all tasks featured in COMFIGURA - (1) a metonymy comprehension task, (2) two metaphor tasks, (3) an idiom task, (4) a proverb task, and (5) an irony task - and taking into account that there are two questions per item, we have a total of 72 questions. Note that the increase of a single item per task would result in a total of 84 questions.

terms of analysis, it could be too broad for small children or some clinical populations, which could have difficulties in expressing their ideas. Therefore, closed-ended questions were included for us to gain some access to these speakers' intuitions. Regarding open-ended questions, we were concerned with creating each one so that they would not lead the participant to merely repeat the stimulus. For instance, for the sentence *'He used a whole bottle for the recipe'*, the open-ended question is *'What did he use?'*. In the case of closed-ended questions, we present two options; one contemplating a metonymic answer and another with the literal interpretation. For the example above, the closed-ended question is *'Did he use the bottle'* or *'What was in the bottle?'*. We randomized the order of the literal and metonymic options in such questions.

As for the task scores, we established that each question would be worth one point, totaling 12 points maximum (two for each of the six items). A participant, when responding to an open-ended question as expected, earns two points with no need for a second, dichotomous question. Closed-ended questions, therefore, are only presented to those who do not respond or do not answer the open-ended questions as expected. If a participant only gets the second question right, they only score one point.

The criteria for scoring the answers were established considering research group discussions and participants' answers to the pilot study. With the pilot, patterns were identified. For example, in the aforementioned item *'He used a whole bottle for the recipe'*, answers to the open-ended question, which asked *'What did he use?'* were usually very specific, and participants tended to cite various kinds of beverages. Answers of this kind (milk, water, beer, soda, for instance) were considered to be satisfactory, since they can be included in the superordinate category of *beverages*. Once the item elaboration stage was completed, the following step was to carry out the pilot studies, described below.

Participants

Considering all the procedures described here, 290 participants were tested in three different studies. The sample for the first study was 21 adults, students, and staff from a public university in Porto Alegre, Brazil, selected by convenience. Participants were between 19 and 54 years old (M=28; SD=11.18). The sample of the second study was made of regularly

enrolled students at a private elementary school, a public high school, and a public university in the greater Porto Alegre area. The sample was stratified into three age groups (103 children, 30 teenagers, and 30 adults). The children were between 5 years and one month old and 11 years and 11 months old (M=9.04; SD=1.5), the teenagers were between 12 years and one month old to 17 years and 11 months old (M=15.3; SD=1.7), and the adults were older than 18 (M=28.3; SD=12.3). As for the third study, the sample was composed of 106 children aging from two years and two months old to 5 years and 10 months old (M=3.1; SD=11.5 months). The sample was stratified into four age groups: 21 children were two, 34 were three, 31 were four, and 20 were five years old.

Age classification followed the guidelines of the *American Academy of Pediatrics*. As an inclusion criterion, it was established that participants should be speakers of Brazilian Portuguese. The exclusion criteria were the presence of cognitive and/or sensorial alterations and learning difficulties reported by the schools in the cases of children and teenagers. After the interviews, data from 9 participants with reports of any of these alterations were discarded; we therefore remained with 290 participants out of 299.

Procedures for Data Collection

After the approval of the project by the university's ethical review board (approval n. 2.469.701, 18/01/18, issued by the Federal University of Rio Grande do Sul), the schools were contacted. Once we presented the research goals to them and obtained consent from the staff, participants were approached. Data collection was carried out individually, in centers for early childhood education (such as nursery and preschool), in a private elementary school, in a public high school, and in a public university. Once consent from participants and/or legal guardians was obtained, the task was carried out. Data collection from each participant for the metonymy task lasted from 3 to 5 minutes.

During rapport, the experimenter conducted a spontaneous conversation with participants, which also served the purpose of subjectively estimating young children's linguistic abilities. During instructions, we made it clear to the participants that they would answer a few questions for which there was not necessarily one expected answer and that there was no relation between the task items. We also made it clear that they had no time limit to answer

each question. A practice item was included so that the participants could become familiar with the test format and would have no questions. During this practice, in case of a non-metonymic answer, some of the possible answers were provided. After the instructions were given, the test items were presented verbally. These procedures refer to all stages described here. In the case of younger children (preschoolers), we have conducted ludic activities prior to the task itself as part of the rapport to increase their engagement throughout the task.

Instrument elaboration

To develop the metonymy comprehension task, three pilot studies were conducted. The first two called for different subsequent adjustments, as detailed below. Having followed the guidelines which were previously discussed, the first version of the task was created. Ten items were elaborated for this first pilot so that we could have room to spot and exclude problematic ones. The practice and a sample of the test items at this first pilot are presented on Table 1.

Table 1 – Examples of task items on the first study

Sentences	Questions
Practice: O prédio todo foi à festa de aniversário. [The entire building went to the birthday party.]	a) Quem foi à festa de aniversário? [Who went to the birthday party?] a') Quem foi à festa foram as pessoas que moram no prédio ou o prédio? [Was it the building or the people who live in it who went to the birthday party?]
Test: Ele bebe duas garrafas inteiras sozinho. [He drinks two whole bottles by himself.]	a) O que será que ele está bebendo? [What is he drinking?] a') Será que ele está bebendo a garrafa ou o que está na garrafa? [Is he

drinking the bottle or what is in the

bottle?]

Those ten items were then submitted to the judgment of five experts in figurative language, former participants of our research group. They were asked to check whether the items were in fact assessing metonymy comprehension and to analyze items in terms of the criteria previously mentioned. For example, the sentence item '*Ela é só mais um rostinho bonito'* [she is just a pretty face] was discarded once judges considered it idiomatic as well as metonymic. Thus, four items that were considered less appropriate were discarded, such as the example detailed in Table 2.

Table 2 – Example of discarded item

Sentence	Questions
Test item: Ele comeu um frango no almoço. [He ate chicken for lunch.]	a) O que exatamente ele comeu? [What exactly did he eat?] a') Será que ele comeu todas as partes do frango ou só a carne do frango? [Did he eat the whole chicken or only parts of the chicken?]

The item in the example above was discarded because, even though the sentence was considered adequate, this metonymy prompts answers that do not allow us to assess if the participants actually understand the phenomenon. Participants invariably answered 'the chicken' for the open-ended question, which is a very conventional way to talk about that particular food and is metonymic in the sentence above.

After collecting data for this first stage, answers from participants (n=21) were checked. With these answers, we identified a few problems with some of the items. Subjects' answers were

corrected individually. These answers also helped in selecting the best six items to compose the task. Once six items were selected, the second study was carried out. Recall that this was a more extensive study, involving not only more participants (n=163) but a broader age range (children, teenagers and adults). Considering the good progress made on task execution, the research group decided to keep the instructions as they were. At that moment, items seemed adequate.

Our next step was to conduct a qualitative analysis, focusing on children's answers. Our aim at that point was to try to identify any difficulty items may have prompted and/or possible patterns children's answers may exhibit. It was only after observing these data that we realized some questions were self-explanatory. This means that children could often answer some of the questions adequately even if they did not understand the metonymy at stake. For example, a question such as the one presented on Table 1 (*What is it that he is drinking?*) naturally prompts drinks as answers. That is to say, the specific verb *to drink* calls for the expected answer, preventing us from knowing if the participant really understood the metonymy. The same happened to other items, such as in '*He likes to read Monteiro Lobato*', in which we previously asked '*What does he like to read?*'. Our solution was to substitute specific verbs in the items for superordinate, more generic ones. This procedure was made for these and other items.

Table 3 – Example of two test items in the third version

Sentence	Questions
Test item 1: Ele usou uma garrafa inteira na receita. [He used a whole bottle for the recipe.]	a) O que será que ele usou? [a) What did he use?] a') Ele usou a garrafa ou o líquido que está na garrafa? [a') Did he use the bottle or what was in the bottle?]
Test item 2: Ele pegou um Monteiro Lobato emprestado hoje.	a) O que será que ele pegou emprestado? [a) What did he borrow?]

[He borrowed a Shakespeare yesterday.]	a') Ele pegou uma pessoa ou algo escrito pela pessoa? [a') Did he borrow a person or something that was written by the person?]

After having modified the items, a third study was conducted, this time with younger children (from two to five years old, n=106). Our aim then was not only to check if the items were well-adjusted, but also to verify the adequacy of such a task for children that could barely speak the language; as young as two years old.

As for consistency of scoring criteria, six expert judges, members of the research group, independently analyzed open-ended questions through another Kappa test, using the software RStudio (v. 3.5.0). The evaluation of the first 50 participants' answers reached an inter-rater reliability score from 0.709 to 1 (Cohen's kappa; $p<0.001$), indicating a great level of coding stability. For every answer for which judges disagreed, between one or several evaluators, answers were discussed until a consensus was reached. Such discussions resulted in adjustments in scoring criteria.

Once we considered the task satisfactory as per data analysis of the third pilot, theoretical procedures were completed. We present below those statistical analyses. The level of significance used was 5% ($p<0.05$) and analyses were run on SPSS (v. 21).

3. Pilot data analysis

Analyses presented here take into account the third pilot of the task and participants' performance on it. Quantitative variables were described by mean and standard deviation or by median and interquartile range. Categorical variables were described by relative and absolute frequencies. Associations between age, when taken continuously, and accuracy rate were analyzed using Spearman's correlation coefficient. As expected, items showed significant variation amongst themselves with open-ended questions being more difficult than closed-ended ones. Our data also suggests a linearity regarding the acquisition of metonymy comprehension, with older children performing better on the task as we also expected.

As shown on Table 4, results point to significant differences between items. The items that obtained the greater accuracy rate on open-ended questions were five and six, while item three had the least number of expected answers. Regarding closed-ended questions, item six was the one with the greatest accuracy rate and item four was the one with the least. The fact that items behaved differently among themselves is not a problem; on the contrary, it is convenient to have a wide-enough range of difficulty in the task to discriminate amongst different individuals, such as age or clinical groups (Urbina, 2014).

Table 4 – Accuracy rate per item

Items	Open-ended	Closed-ended	p^{**}
	n (%)	n (%)	
1	18 (17.0)[b]	69 (65.1)[bc]	<0.001
2	7 (6.6)[ab]	70 (66.0)[bc]	<0.001
3	0 (0.0)[a]	60 (56.6)[ab]	<0.001
4	3 (2.8)[ab]	47 (4.,3)[a]	<0.001
5	46 (4.4)[c]	59 (55.7)[ab]	0.015
6	44 (4.5)[c]	86 (81.1)[c]	<0.001
p*	<0.001	<0.001	

* Cochran test; ** McNemar test; [a,b,c] Same letters in each column mean they do not differ on the multiple comparison Cochran test at 5% significance.

From a statistical point of view alone, questions with only two options are more likely to be accurately answered than open-ended questions, since participants have a 50% chance to provide the expected answer. Therefore, it is not a surprise that, in every single item, there was a statistically significant difference between open and closed-ended questions.

On the other hand, it was a surprise that we had an item that proved to be so difficult that no preschooler could come up with a metonymic interpretation. On item three, not a single accurate answer was provided by participants to the open-ended question. This item instantiates the conceptual metonymy PERSON FOR PRODUCT, which could be difficult to know if the interlocutor is not familiar with the particular source domain. We believe that this was the case. The source domain we used (Monteiro Lobato) is a renowned Brazilian children's book author whose profession, though well known for adults, was not identified without prompt by the preschoolers interviewed. Still, we opted to keep this item because it is a good example of how metonymy happens in the wild regarding world knowledge.

For items five and six, on the other hand, many children as young as two years old could come up with accurate answers (see Table 5). Those items instantiate the metonymies OUTFIT FOR PERSON (SHIRT FOR SOCCER PLAYER) and CONTAINER FOR CONTENT (PLATE FOR FOOD), respectively. In either case, such task items involve very frequent topics and linguistic expressions, both because of common daily activities and Brazilian culture specifically. For item five, however, results did not evidence a gradual improvement with age. We attribute this both to a small number of participants aged 5 years old and to the fact that we did not establish intervals between children's age ranges in our sample. Thus, children who were almost 5 have outperformed kids who were recently 5.

Table 5 – Percentage of correct answers according to age

Questions	2 years old (n=21)	3 years old (n=34)	4 years old (n=31)	5 years old (n=20)	$p*$
	n (%)	n (%)	n (%)	n (%)	
1					
Open	1 (4.8)	6 (17.6)	8 (25.8)	3 (15.0)	0.251
Closed	7 (33.3)	24 (70.6)	22 (71.0)	16 (80.0)	0.004
2					

Open	0 (0.0)	1 (2.9)	1 (3.2)	5 (25.0)	0.003
Closed	8 (38.1)	19 (55.9)	24 (77.4)	19 (95.0)	<0.001
3					
Open	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	-
Closed	6 (28.6)	19 (55.9)	19 (61.3)	16 (80.0)	0.001
4					
Open	0 (0.0)	1 (2.9)	1 (3.2)	1 (5.0)	0.361
Closed	2 (9.5)	15 (44.1)	18 (58.1)	12 (60.0)	0.001
5					
Open	8 (38.1)	11 (32.4)	19 (61.3)	8 (40.0)	0.307
Closed	3 (14.3)	15 (44.1)	26 (83.9)	15 (75.0)	<0.001
6					
Open	3 (14.3)	14 (41.2)	14 (45.2)	13 (65.0)	0.002
Closed	13 (61.9)	30 (88.2)	24 (77.4)	19 (95.0)	0.039

*Chi square test for linear trend.

We also analyzed results taking age as a continuous variable. Figure 4 presents linear associations between age (in months) and the scoring rate on open-ended (r_s=0.361, $p<0.001$) and closed-ended questions (r_s=0.576, $p<0.001$). Both correlations were positive and statistically significant, albeit in less intensity for open-ended questions. Such weaker association mirrors the smaller score for open-ended questions in all ages. Specifically, regarding open-ended questions, the correlations for half of the items were not significant. Considering that this instrument is to be part of a more comprehensive task and that our ultimate goal is to investigate and/or attest a gradual line of acquisition between the phenomena, we believe that small differences within a phenomenon are not so crucial.

Figure 4 – Association between scoring rate and age in months
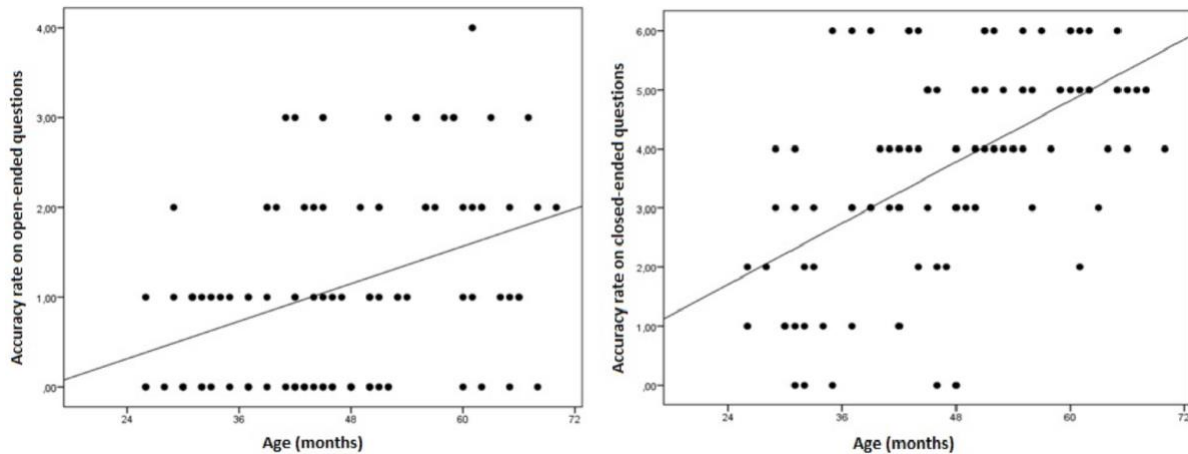
As shown above, some of the items were highly scored by our participants and others resulted in very poor accuracy rates, which suggests our items cover different aspects of metonymy. Therefore, items seem to be representative enough of the phenomenon to discriminate between age groups (and potentially groups of clinical conditions). Our findings, therefore, corroborate what the literature indicates: metonymy comprehension should be a gradually acquired skill. We conclude the task is an adequate tool for assessing metonymy comprehension to any age above two years old.

4. Final remarks

The objective of this paper was to present the elaboration of a metonymy comprehension task as well as pilot test it. As one can observe from the details on the principles and procedures that guided the creation of the task, this is a process that demands the definition of a theoretical orientation and technical rigor. The task here was developed considering the experientialist perspective of cognitive linguistics and psychometric guidelines for the elaboration of a psycholinguistic assessment instrument.

Central aspects of our theoretical model were considered since the beginning of the task design. CL's postulation of a conceptual facet (a mapping between domains) guided our selection of metonymic expressions that were not all linguistic instantiations of the same conceptual mapping. Similarly, we controlled the conventionality of such mappings, making sure that only highly conventional mappings were included in the task. Conventional

mappings regard mappings that potentially involve experiences that are likely to have already been experienced by the age of 4. This aspect is related to the central notion of embodiment in CL. Instantiations and mappings that describe early everyday experiences were used. This choice is related to the fact that we aim to propose an instrument that is able to test people at different ages, socioeconomic statuses, educational levels, and geographical regions (consequently, from different cultural backgrounds).

Developing a comprehensive test that could be used to assess individuals at different stages in life and from different backgrounds requires the adoption of theoretical concepts and specific methodological procedures, not only to elaborate the tasks but also to carry them out. Carrying out tasks involves situation managing (such as the experimenter's attitude during rapport or the arrangement of the material) and time monitoring (such as dividing the interview into two or more sessions). In our case, this included creating ludic activities for very young children to feel comfortable to engage in answering questions. This is the kind of knowledge that is not usually described in manuals. Even when variables are controlled and the test is elaborated according to the strictest standards in psychometrics, there can be some subjectivity involved in carrying out the task and in scoring the answers, and some degree of flexibility will be required from the researcher/experimenter.

To maintain the pattern of the figurative language comprehension test (COMFIGURA) as a whole, to minimize participants' potential fatigue from answering several items, and to ensure some power of generalization, we established the number of six items (and a maximum of twelve questions) in this instrument as well as in the other tasks. Three pilot studies and a conventionality task were necessary and useful for us to adjust items, and their results will be presented in the technical manual of the instrument. Alterations were made after the first and second pilots. Data collection from the second pilot, which included a higher number of subjects, was especially useful in terms of evidence of a variety of adequate answers to the questions and, thus, improved the scoring criteria. These criteria will also be presented in the technical manual.

During the development of this task, we have faced a few challenges. The first big challenge was to isolate metonymies from other phenomena, especially metaphor, but also idioms, on task items. In the present study, we have tried to avoid the occurrence of metaphtonymies (Goossens, 1990) in the elaboration of the items. Considering the hypothesis that different

figurative language phenomena are mastered at different developmental stages, an item that mixes both tropes could confound the assessment of a specific phenomenon. However, one of our items (number four) could be said to have both a metonymic and a metaphoric interpretation. We have decided not to exclude it because we believe that if participants answered the questions according to a metaphoric interpretation, they would be also equipped to understand it metonymically, once literature points to the fact that metaphor is a more complex phenomenon and consequently is of later acquisition. We believe that this has not affected our results, and it confers ecological validity to the task, in the sense that people experience mixed phenomena in the wild.

The second challenge was to control items for sentence size, syntactic structure, semantic complexity, and familiarity and, at the same time, keep it standardized within this task and in relation to the other tasks of the instrument. To a lesser degree in the case of familiarity and greater degree in the case of the other aspects, we believe we have been able to overcome such challenges. The last, but not least of them was to create a suitable instrument to assess individuals at different ages and conditions, clinical and non-clinical, which, even though we have not tested in this study, are still a target audience of the task.

A very specific issue involves one of the items – number three, to which no young children have accurately answered. This item involves world knowledge (a children's book author) and presents an obstacle for people who are not familiar with it. Most probably, adults will provide expected answers without hesitation for that item. Evidence of such is that older kids intuitively answer our open-ended questions as if the author is a painter or an artist of another kind. Other studies found similar results (e. g., Zhu, 2021). In items involving a popular sport or food, which are presumably learned earlier in life, in turn, even very young children perform well. We do not consider this a problem and have kept the item because we understand this is a more potentially discriminatory item in terms of age.

As with any psycholinguistic instrument, this task presents limitations. One of them regards the lack of diverse metonymic mappings. We do not contemplate different mappings that could be more or less complex. This limitation derives from another one: the fact that we only have six items, which we acknowledge is a small number.

Despite these limitations, the development of this task should help compensate for the lack of metonymy comprehension assessments in the Portuguese language and eventually in some other languages. As of today, there is no similar instrument for the assessment of metonymy comprehension that may help educators, educational and clinical psychologists, and speech therapists to diagnose learning disabilities and clinical conditions. We understand that the main contributions of the task itself and of COMFIGURA as a whole is to make figurative language assessments possible in different populations and to contribute to the progress of theoretical aspects of figurative language development. Ultimately, advances resulting from the test can reverberate in both theoretical, clinical, and educational contexts.

Acknowledgements

Conflict of interests

(X) The authors hereby declare that we do not have any potential conflict of interest in this study.

Authors contribution

Maity Siqueira was responsible for the conceptualization of the study. Maity Siqueira, Laura Baiocco, Carolina Girardi Ferrari, and Nichele Lopes have participated in data collection. All authors participated in study design, statistical data analysis, data curation and validation, and manuscript writing. All authors approve the final version of the manuscript and are responsible for all its aspects, including the guarantee of its veracity and integrity.

The Ethics Committee Approval Statement

This study was approved by the Institutional Review Board at UFRGS (CEP/UFRGS), approval number 5.816.521, CAEE 7 73089117.6.0000.5347.

References

Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Prentice Hall.

Barcelona, A. S. (2000). *Metaphor and Metonymy at the Crossroads*. De Gruyter Mouton.

De Leon, V., Siqueira, M., Parente, M., & Bosa, C. (2007). Especificidade da compreensão metafórica em crianças com autismo. *Psico, 38*(3), 269-277. https://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/1454 (Acessado 01 março, 2023).

Falkum, I. L., Recasens, M., & Clark, E. V. (2017). "The moustache sits down first": on the acquisition of metonymy. *Journal of Child Language*, *44*(1), 87-119. https://doi.org/10.1017/S0305000915000720

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology*: Learning, Memory, and Cognition, *25*(6), 1366–1383. https://doi.org/10.1037/0278-7393.25.6.1366

Frisson, S., & Pickering, M. J. (2007). The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes*, *22*(4), 595-613. https://doi.org/10.1080/01690960601017013

Goossens, L. (1990) Metaphtonymy: the interaction of metaphor and metonymy in expressions for linguistic action. *Cognitive Linguistics*, *1*(3), 323-342. https://doi.org/10.1515/cogl.1990.1.3.323

Köder, F., & Falkum, I. L. (2020). Children's metonymy comprehension: Evidence from eye-tracking and picture selection. *Journal of* Pragmatics, *156*, 191-205. https://doi.org/10.1016/j.pragma.2019.07.007

Kövecses, Z. (2010). *Metaphor*: A practical introduction. Oxford University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things*. University of Chicago Press.

Melogno, S., & Pinto, M. A. (2014). Enhancing metaphor and metonymy comprehension in children with high-functioning autism spectrum disorder. *Psychology*, *5*(11), 1375-1383. https://doi.org/10.4236/psych.2014.511148

Nerlich, B., Clarke, D. D., & Todd, Z. (1999). Mummy, I like being a sandwich. Metonymy in Language Acquisition. In K. U. Panther, & G. Radden (Eds.), *Metonymy in Language and Thought* (pp. 361-383). John Benjamins.

Panther, K. U., & Radden, G. (Eds.). (1999). *Metonymy in Language and Thought*. John Benjamins.

Panther, K. U., & Thornburg, L. L. (2007). Metonymy. In D. Geeraerts, & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics* (pp. 236-263). Oxford University Press.

Pasquali, L. (2010). *Instrumentação psicológica*: fundamentos e práticas. Artmed.

Rundblad, G., & Annaz, D. (2010a). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism*, *14*(1), 29-46. https://doi.org/10.1177/1362361309340667

Rundblad, G., & Annaz, D. (2010b). Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, *28*(3), 547-563. https://doi.org/10.1348/026151009X454373

Siqueira, M. (2004). *As metáforas primárias na aquisição da linguagem: um estudo interlinguístico* [Unpublished doctoral dissertation]. Pontifical Catholic University of Rio Grande Sul, Porto Alegre. http://www.leffa.pro.br/tela4/Textos/Textos/Anais/Textos_Em_Psicolin/TESES%20DOUTO RADO/As%20Met%C3%A1foras%20Prim%C3%A1rias%20Na%20Aquisi%C3%A7%C3% A3o%20Da%20Linguagem-%20Um%20Estudo%20Interling%C3%BC%C3%ADstico.pdf (Acessado 01 março, 2023).

Siqueira, M., & Gibbs Jr, R. (2007). Children's acquisition of primary metaphors: a crosslinguistic study. *Organon*, *21*(43), 161-179. https://doi.org/10.22456/2238-8915.39590

Siqueira, M., & Lamprecht, R. (2007). As metáforas primárias na aquisição da linguagem: um estudo interlinguístico. *Delta, 23*(2), 245-272. https://doi.org/10.1590/S0102-44502007000200004

Siqueira, M., Parente, M., & Gil, M. (2009). Metáfora e cultura: uma interface entre a Linguística e a Antropologia. *Antares: Letras e Humanidades, 2*, 99-111. https://professor.ufrgs.br/maity/files/2009._siqueira_parente_gil.pdf (Acessado 01 março, 2023).

Siqueira, M., & Marques, D. (2018). Desenvolvimento e validação do instrumento de compreensão de expressões idiomáticas. *Revista de Estudos da Linguagem, 26*(2), 571-591. http://dx.doi.org/10.17851/2237-2083.26.2.571-591

Theodoropoulou, M. (2021). "Has she got a mouth?" Metonymy, salience, and experience in a child's speech. *14ο Διεθνές Συνέδριο Ελληνικής Γλωσσολογίας*, *1*(1), 1269-1277. https://doi.org/10.26220/icgl.v1i1.3759

Van Herwegen, J., Dimitriou, D., & Rundblad, G. (2013). Development of novel metaphor and metonymy comprehension in typically developing children and Williams syndrome. *Research in Developmental Disabilities*, *34*(4), 1300-1311. https://doi.org/10.1016/j.ridd.2013.01.017

Urbina, S. (2014). *Essentials of psychological testing*. John Wiley & Sons.

Van Herwegen, J., & Rundblad, G. (2018). A cross-sectional and longitudinal study of novel metaphor and metonymy comprehension in children, adolescents, and adults with Autism Spectrum Disorder. *Frontiers in psychology*, *9*, 945. https://doi.org/10.3389/fpsyg.2018.00945

Xie S (2022) The developmental trajectory of mandarin Chinese-speaking children's pure metonymy comprehension ability. Forum for Linguistic Studies *4*(1): 78–99. https://doi.org/10.18063/fls.v4i1.1457

Zheng, Q., Jia, Z., & Liang, D. (2015). Metaphor and metonymy comprehension in Chinese-speaking children with high-functioning autism. *Research in Autism Spectrum Disorders*, *10*, 51-58. https://doi.org/10.1016/j.rasd.2014.11.007

Zhu, R. (2021). Preschoolers' acquisition of producer-product metonymy. *Cognitive Development*, *59*, 101075. https://doi.org/10.1016/j.cogdev.2021.101075

This preprint was submitted under the following conditions: