Applied Psycholinguistics (2026), **47**, e48, 1–36
doi:10.1017/S0142716426100897

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Development of language-mediated abstraction: Insights from word ladders task

Caterina Villani, Adele Loia and Marianna M. Bolognesi

Department of Modern Languages, Literatures, and Cultures, University of Bologna, Bologna, Italy
Corresponding author: Marianna M. Bolognesi; Email: m.bolognesi@unibo.it

(Received 19 December 2025 UTC; revised 28 May 2026 UTC; accepted 14 July 2026 UTC)

Abstract

Abstraction is a fundamental cognitive capacity underlying the organization of semantic knowledge. This study examines the development of language-mediated abstraction in school children, defined as the ability to generate increasingly more general lexical items from a given prompt, and investigates how this ability is shaped by word semantics. Children aged 9–13 years completed a longitudinal, generative word-production task in which they constructed taxonomic hierarchies by linking words through categorical inclusion relations (*IS-A, e.g., dog IS-A animal*), starting from prompts that varied in concreteness. Their productions were analyzed in terms of hierarchical *productivity* (ladder length) and *validity* (accuracy of taxonomic relations). Results revealed that language-mediated abstraction improves over time, particularly among younger children. Moreover, these longitudinal changes could not be explained by chronological age alone. Improvements occurred across semantic domains, although abstract prompts consistently elicited shorter taxonomic hierarchies than other concepts. Developmental variations also emerged in abstraction pathways, with nontaxonomic semantic relations occurring most frequently in children of intermediate age and in response to abstract and emotional concepts. Together, the findings provide evidence that concreteness and specificity constitute dissociable dimensions of abstraction, each contributing uniquely to the development and organization of semantic categories. By adopting a generative, longitudinal approach, this study offers novel insight into how children progressively construct and refine hierarchical conceptual knowledge through language.

Keywords: Abstraction; concreteness; language development; specificity; word ladders task

Introduction

Abstraction, a hallmark of human cognition, is the ability to form general concepts or ideas by extracting similarities and general tendencies from direct experience, language, or other concepts (Reilly et al., 2025). Although concepts can be represented independently of words, linguistic labels often act as cues (Lupyan, 2012;

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Lupyan & Lewis, 2019) that help create and organize our knowledge into coherent categories. Crucially, words can define different types of categories, ranging from very specific (e.g., "electric bike") to very general (e.g., "vehicle"). This hierarchical organization applies to varying degrees across different types of natural categories, notably Linnaean categories (e.g., "plants" and "animals") and inanimate entities (e.g., "tools," "modes of transportation"). It also seems applicable to some abstract categories, such as social reality—for instance, "religion" is a broad category, whereas "Hinduism" represents a more specific subtype. In this sense, hierarchical semantic categorizations are applicable to both categories defining concrete and abstract concepts.

Word **specificity** ("electric bike" vs. "vehicle") and word **concreteness** ("bike" vs. "religion") are two distinct variables that are involved in the general mechanisms of abstraction (Villani & Bolognesi, 2026). The *APA Dictionary of Psychology* (VandenBos, 2015) provides two primary meanings for the notion of "abstraction": (1) the process of deriving general concepts by recognizing shared features among specific instances, and (2) an intangible concept, such as "goodness" or "truth." While the first definition pertains to categorization, the second relates to the perceptibility of a concept's referent.

Research on abstraction mechanisms has often focused solely on specificity or solely on concreteness, with little attention given to their interplay in language development, conceptual processing, and semantic categorization. Classic works on specificity, or better, on the architecture of semantic categories in the human mind, have primarily focused on natural categories (Rosch, 1973), which define referents in the world that can be directly experienced through perception. These studies show that words that denote basic-level categories, as opposed to highly general and highly specific, are easier to understand and process (for a review, see Hajibayova, 2013; Rosch et al., 1976). Recent developmental studies have taken the perspective of language acquisition, showing for instance that richer linguistic knowledge of superordinate (i.e., very general and broad) categories is linked to faster vocabulary growth (Lewis et al., 2021). This developmental trajectory may, in turn, facilitate knowledge organization and inferential reasoning (Suffill et al., 2022). However, as described in further detail in the next section, previous studies did not investigate the evolving structure and construction of the hierarchical relations between words in speakers' minds, during language development. Rather, they based their analyses on specificity/generality judgments elicited from speakers and corpus data. Therefore, an important question remains: how do children use language to organize and reorganize their conceptual knowledge hierarchically, into categories of increasing inclusiveness, during language production? Examining this process is essential because it moves beyond passive vocabulary knowledge—simply recognizing and judging provided words—to uncover how children actively structure linguistic and conceptual information. By assessing their evolving ability to generate hierarchies of semantic categories through words, or "word ladders," we can gain deeper insights into the cognitive mechanisms underlying category formation and organization.

In addition, it remains an open question whether and to what extent the mechanisms of semantic categorization function equally well in language development for both concrete and abstract concepts. In fact, abstract concepts have been shown to be more difficult to learn (Gleitman et al., 2005). Overall, abstract words

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

are processed more slowly and less accurately than concrete ones (Kroll & Merves, 1986; Paivio, 1991; Schwanenflugel, 1991; Schwanenflugel et al., 1988) and are acquired later (Bellagamba et al., 2022; Della Rosa et al., 2010; Gilhooly & Logie, 1980; Villani et al., 2019). The challenge associated with abstract concepts perhaps lies in the fact that often multiple situations and experiences are grouped under a single label rather than pointing to a single clear referent (Davis et al. 2020). From the perspective of semantic categorization, some scholars suggest that abstract concepts may lack taxonomic organization or that such organization may be much more limited compared with concrete concepts (Borghi & Binkofski, 2014; Crutch, 2005; Desai et al., 2018; Troyer & McRae, 2022). However, empirical evidence in this field remains scarce.

Finally, sources of variation in language development should not be underestimated, particularly when it comes to precise semantic relations such as the hypernym/hyponym relation, which is typically formalized in lexical semantics and ontology through the IS-A relation (e.g., dog IS-A animal). In many cases, maintaining strict adherence to this relation during abstraction proves challenging. For example, while a monkey IS-A mammal, is it truly the case that a table IS-A furniture? Or is it the case that a table can more accurately be described as *a piece of* furniture—one that contributes to forming the broader category of furniture only when considered alongside items like a *couch* and a *lamp*? In other words, is "table" a hyponym or a meronym of "furniture"? These ambiguities highlight the complexities of taxonomic organization and raise questions about whether abstraction follows a shared pathway across speakers or, instead, unfolds through a web of intricate semantic relations that are not necessarily formalizable by IS-A relations[1]. In this process, individuals navigate a winding path to abstraction through language, like the trails in a forest, where each person carves out their own route, more or less well-worn.

Based on the arguments outlined above, the general aim of the present study is to investigate children's developing ability to construct taxonomic lexical hierarchies through the Word Ladders task across different ages and semantic domains (see the specific aims in the following section). More specifically, the study focuses on children's capacity to organize lexical items into hierarchical category-based relations, and on how this competence develops over time and varies as a function of word type. In this sense, the present work does not aim to address abstraction in its broadest cognitive formulation. Rather, it examines the development of taxonomic lexical organization as reflected in the production of hierarchical word chains. This ability is particularly relevant in educational contexts, where the acquisition and manipulation of taxonomically structured knowledge play an important role in vocabulary development and in the learning of specialized terminology.

The research questions addressed in the present study are the following:

RQ1: Does language-mediated abstraction improve with age (between participants) and over the span of one school year (within participants)?

RQ2: How does semantics (in particular, concrete vs. abstract) influence children's ability to construct ladders and therefore to perform language-mediated abstractions over time/age?

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

RQ3: To what extent is the abstraction pathway shared by speakers during development?

The remainder of the paper is structured around the three research questions. After presenting the theoretical background and the methodological framework, we discuss each research question separately, reporting the corresponding analyses, results, and brief interpretations in sequence. This organization is intended to facilitate readability and allow the reader to follow one line of inquiry at a time, while progressively building the overall argument of the paper. A final General Discussion section then integrates the findings across the three analyses and discusses their broader theoretical implications.

Theoretical background

The hierarchical organization of semantic categories in linguistic development

A key finding of Rosch et al.'s (1976) classic work on the organization of the mental lexicon in the human mind is that categories can be grouped into three main levels: superordinate, basic, and subordinate. For instance, "animals" represents the superordinate level, "dog" is at the basic level, and "beagle" is the subordinate level. These levels reflect a hierarchical structure, where categories are connected through inclusion relations, with specific categories nested in the more general ones: Beagles are dogs and dogs are animals. Categories differ in the amount and type of information they convey, with basic-level categories being the most frequently used because they offer the optimal balance between informativeness and distinctiveness perceived among category members, allowing for their efficient differentiation. Superordinate categories are more general and less informative, with members sharing only a few attributes, whereas subordinate categories contain highly detailed and specific information, but their attributes often overlap with those of other categories at the same level (for a review, see Cohen & Lefebvre, 2005).

Studies on child language development have shown that categories at different levels of abstraction are acquired at different developmental stages. Early word learning is characterized by a preference for basic-level categories over superordinate terms (Markman, 1990). Consequently, children tend to acquire superordinate terms later, as their understanding of hierarchical category relations matures, and they begin to recognize the broader structural organization of lexical categories (Clark, 1993; see also Mervis & Crisafi, 1982).

The ability to use conventional forms of taxonomic knowledge—superordinate and subordinate categories—becomes more prominent as children reach school age, around 7 or 8 years. Early on, children tend to use slot-filler categories based on shared functions within events (Horton & Markman, 1980; Lucariello et al., 1992; Lucariello & Nelson, 1985; Nelson, 1988). For example, in the event of eating, "breakfast" may initially be linked to a single item like "cereal," but with further experience, it can expand to include other foods such as "toast" or "eggs." As children develop more slot-filler categories, like "food for breakfast" or "food for lunch," they begin to understand the general superordinate category of "food," which encompasses a range of items with shared functional features, like edibility.

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

However, there is also evidence that when the task provides cues to direct children's attention to taxonomic relations (e.g., novel labels, linguistic descriptions), even preschool-aged children (3-year-olds) can learn and use superordinate categories (e.g., Markman, 1989; Waxman & Gelman, 1986). Taken together, these findings suggest that language has a facilitatory effect on abstraction processes, which become more refined and consolidated with age and formal education. Recently, Lewis et al. (2021) demonstrated that early development of words at superordinate level facilitates subsequent vocabulary growth. In other words, given the same vocabulary size, children who know more superordinate words, such as "color," "animal," and "move," are faster at learning new words associated with those categories. However, a key concern in this study—acknowledged by the authors themselves—is the method used to identify semantically related words. The authors extracted these relations from WordNet (Miller et al., 1990), a structured database of English, organized by hierarchical semantic relations. The problem arises because the knowledge encoded in WordNet may only loosely correspond to the knowledge of average speakers (for a discussion, see Bolognesi et al., 2020; Villani & Bolognesi, 2026).

Although previous research indicates a positive correlation between abstraction skills and vocabulary acquisition, it remains therefore unclear how children actively apply the ability to organize categories into hierarchical taxonomies, in production tasks. Rather than relying on indirect measures such as speaker judgments, corpus data, or passive classification tasks, we therefore investigate how children progress along the scale of language-mediated abstraction in a production task in which they are asked to generate *paradigmatic* relations. In lexical semantics, *syntagmatic* and *paradigmatic* relations capture complementary dimensions of semantic structure. Syntagmatic relations link words that co-occur within events or sequential contexts and are often thematic in nature (e.g., *dog–bark*, *spoon–eat*), whereas paradigmatic relations connect words that occupy the same structural position and are related by shared semantic features or category membership, such as synonymy, antonymy, or hypernymy (e.g., *dog–animal*, *dog–cat*). Developmental research has shown that children's earliest lexical associations are predominantly syntagmatic, reflecting event-based knowledge grounded in everyday experience (thematic relations), while paradigmatic relations emerge later as cognitive and linguistic abstraction increases. A seminal study by Nelson (1977) demonstrated that young children preferentially produce syntagmatic associations in word-association tasks, with a gradual developmental shift toward paradigmatic responses in later childhood. This shift is theoretically significant because paradigmatic relations presuppose the ability to represent stable categories and hierarchical semantic structures independent of immediate context. Converging evidence from studies on word learning further indicates that sensitivity to taxonomic organization, a core paradigmatic relation, develops later than sensitivity to thematic relations (Markman & Hutchinson, 1984). The production of paradigmatic relations thus constitutes a major milestone in lexical development, signaling the reorganization of the lexicon into a more abstract, category-based system that supports efficient vocabulary growth and semantic generalization. We argue that language-mediated abstraction is a dynamic skill that evolves over time, potentially not following a linear path for every individual or category. **We hypothesize that language-mediated abstraction skills improve**

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

overall with age and through formal schooling, as shown in previous studies (Horton & Markman, 1980; Lucariello & Nelson, 1985; Lucariello et al., 1992; Nelson, 1977; Nelson, 1988). Specifically, as children's reading and writing skills advance, their linguistic competence will also expand to include less frequent words, both more specific and more general, enabling them to identify and produce taxonomic relations between words.

The developmental trajectory of different semantic categories

Abstract concepts are often considered more challenging to learn than concrete ones (Gleitman et al., 2005), an effect that can be explained at least in part by the fact that concrete words refer to readily perceivable objects and are therefore easier to map onto referents (Borghi & Binkofski, 2014; Buccino et al., 2019). Classic research by Brown (1957), cited in Schwanenflugel (1991), found that 75% of the words most frequently produced by school-age children (Grades 1–8) were concrete, compared with just 28% in adults' lexicons—a pattern recently confirmed by Lupyan and Winter (2018). Age-of-acquisition (AoA) ratings, which estimate the age at which words are typically learned, further support this trend; abstract words are, on average, acquired later than concrete ones, with the proportion of abstract words increasing rapidly during middle childhood (ages 7–10) and reaching around 40% of vocabulary by age 12 (Kuperman et al., 2012; Ponari et al., 2018).

Within the broad category of abstract concepts, emotional concepts represent a particularly interesting case. Although rated as more abstract than concrete words, they are consistently judged as less abstract than other abstract concepts (Della Rosa et al., 2010; Kousta et al., 2011; Villani et al., 2019). This intermediate status is argued to act as a developmental bridge, facilitating the transition from concrete to more abstract concepts (Abstract-via-Emotion hyphothesis, Kousta et al., 2011; Newcombe et al., 2012; Ponari et al., 2018; Vigliocco et al., 2009, 2018; see Winter, 2022 for a critical discussion). Recent research by Nook et al. (2020) tracked the development of emotion concepts and found that younger children define emotions through external, situational contexts, whereas older children increasingly rely on internalized, abstract definitions (e.g., using synonyms). Interestingly, the age at which emotions are understood varies widely: basic emotions like "love" are understood by age 4, whereas more complex emotions, like "pride" and "amazed," reach full comprehension around age 11. However, the ability to define emotions, provide synonyms, and offer situational examples—requiring a higher level of abstraction—continues to develop into early adulthood. A developmental shift toward abstraction was found by Muraki et al. (2025) in a large-scale vocabulary study investigating words with multiple meanings. Using meaning-specific concreteness ratings and AoA norms spanning Grade 2 to college age, the authors showed that earlier-acquired meanings tend to be more concrete, whereas later-acquired meanings become increasingly abstract. Importantly, this effect was stronger for words typically classified as concrete, suggesting that initial sensorimotor-based meanings may scaffold the subsequent acquisition of more abstract senses of the same lexical item, presumably drawing on language experience.

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Along the language developmental trajectory, concepts denoting social roles (e.g., "doctor," "teacher") represent a distinct category that has received comparatively less attention. Social roles categories combine both concrete and abstract features; they refer to individual people, yet their meaning is shaped by culturally defined roles, expectations, and relational norms rather than purely perceptual features. Social role concepts also tend to exhibit more evident gradations of specificity (doctor → surgeon) compared with purely abstract concepts, making them more comparable to concrete categories. Behavioral studies show that even preschoolers represent these categories by attending to both labels and role-relevant traits (e.g., Foster-Hanson & Rhodes, 2019), indicating that social role categories are psychologically real from early development.

Together, these studies suggest that conceptual representations evolve with development and experience, following a trajectory from early acquisition of concrete and emotional words to the gradual incorporation of more abstract words. However, this developmental pattern may be shaped not only by the concreteness or complexity of concepts but also by their position within conceptual hierarchies (superordinate, basic, and subordinate levels). Notably, some highly frequent abstract words (e.g., "freedom," "idea") are acquired earlier than less frequent general concrete words (e.g., "vehicles," "furniture") or highly specific terms (e.g., "submarine," "chaise lounge"). This suggests that both concreteness and specificity may jointly shape word learning.

A key limitation in much of the existing research is that studies rarely control for word specificity while selecting stimuli. Instead, experimental stimuli are typically selected and balanced for frequency, often favoring basic-level words—leaving superordinate and subordinate terms underrepresented. This gap is partly due to the lack of standardized metrics and resources for measuring word specificity, a methodological challenge recently highlighted by Villani and Bolognesi (2026).

Recent studies have begun to address these gaps by collecting specificity ratings and examining how specificity relates to other semantic properties. Overall, these studies report a mild positive correlation between concreteness and specificity (Bolognesi & Caselli, 2023; Bolognesi et al., 2020; Ravelli et al., 2024). Crucially, specificity also influences semantic decision latencies, with more specific words processed faster than less specific ones—an effect independent of word concreteness (Lamarra et al., 2026). This further suggests that concreteness and specificity engage distinct mechanisms in conceptual processing.

Recent insights into how specificity and abstraction jointly structure the lexicon come from Villani et al. (2024), who conducted a large-scale semantic content analysis of roughly 1,000 Italian words and revealed that both concrete and abstract categories distribute along a continuous general-to-specific scale. Some categories (e.g., "the animal kingdom," "social entities") span long taxonomic hierarchies, linking highly general to highly specific terms, and others (e.g., "natural elements," "meta-logical relations") form shorter taxonomies concentrated at either the general or the specific end of the continuum.

Importantly, these findings challenge the prevailing assumption that only concrete concepts are organized into taxonomic hierarchies, whereas abstract concepts are thought to rely on relational and thematic links (Gentner & Asmuth, 2019; Katja Wiemer-Hastings & Xu, 2005; see also Troyer & McRae, 2022).

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

However, it is worth noting that Villani et al.'s analysis is based on indirect semantic judgments obtained through Likert-scale ratings rather than on speakers' actual linguistic production. This leaves open the empirical question of whether speakers spontaneously produce taxonomic associations for abstract concepts, as they do for concrete ones. Most prior evidence on this topic comes from free-production tasks, where participants can generate associations across different types of semantic relations without constraints. Thus, direct empirical investigation is still needed to determine whether both concrete and abstract concepts are equally accessible through taxonomic associations in speakers' spontaneous language use.

The second aim of this study is to examine how children's ability to form hierarchical relations between words—and thus engage in linguistic abstraction—develops across different semantic categories (i.e., concrete, abstract, emotional, and social role concepts). Specifically, we investigate whether the concreteness of target words influences the ease and precision with which children establish taxonomic hierarchical structures. Based on the research reviewed above, **we hypothesize that concrete categories (e.g., animals) elicit more structured and accurate hierarchies than abstract categories (e.g., values), with younger children showing particular difficulty forming abstract and emotional hierarchies. We also predict developmental gains over time, with younger children showing steeper improvements as their linguistic competence matures.**

Taxonomic and functional abstraction in language

One of the classic principles identified in children's vocabulary development is the so-called taxonomic bias, which refers to the tendency to extend newly learned words based on shared category membership rather than thematic relationships (Markman & Hutchinson, 1984). When learning a new noun, young children are more likely to generalize it to other objects of the same kind (e.g., calling both a Labrador and a Poodle "dog") rather than to thematically related objects (e.g., calling a leash "dog"). This bias supports the organization of concepts into hierarchical taxonomic structures, facilitating efficient category learning and lexical development in the subsequent years, particularly during schooling. However, taxonomic relations in language-mediated abstraction are not always straightforward and can take different routes, influencing how knowledge is structured and labeled with words. A key distinction in lexical semantics is between hypernymy—which organizes concepts into taxonomic hierarchies (e.g., animal → dog)—and meronymy, which reflects part–whole relationships (e.g., wheel → car). While hypernyms facilitate abstraction by grouping entities based on shared properties, meronyms capture compositional and functional dependencies, which are equally relevant for conceptual organization (Murphy, 2002). Meronymic and taxonomic relations may sometimes become cognitively intertwined, a well-known problem for ontology construction that makes it difficult to disentangle hierarchical abstraction from compositional or situational structure (e.g., Ferrucci, 2022; Guarino, 1998; Gangemi et al., 2001, 2010; Vossen, 1998). While some category relations clearly instantiate an IS-A hierarchy (e.g., monkey IS-A mammal), other cases are less straightforward and may recruit associative or compositional reasoning. For instance, objects such as table, lamp, and couch are often grouped together under the broader notion of furniture, not only because they belong to the same semantic

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

category, but also because they co-occur within a functional and situational whole (e.g., a furnished room). In such cases, categorization may partially rely on part–whole or scene-based organization rather than on a purely taxonomic hierarchy. Beyond meronymy, other semantic relations can also interfere with strict taxonomic abstraction. For example, one could argue that some thematic associations may still play a crucial role in how concepts are abstracted and related to one another. For instance, "clock" may be considered as an instantiation of "time," and therefore one may suggest that the relation "clock"→"time" is based on a mechanism of abstraction. Moreover, even a concrete concept like "cat" may be related to more than one superordinate category. From a strictly taxonomic perspective, a cat is a feline, which is a type of mammal within a Linnaean hierarchical structure. However, in everyday language, a cat might also be categorized as a pet, and within this context, it can even be seen as a friend or companion. This dual categorization, or double inheritance, illustrates how a single concept can belong to different semantic networks depending on the relational context. The problem of double inheritance is well known in lexical semantics and ontology construction (Guarino & Welty, 2000), as the same object can fit into multiple hierarchies, making abstraction a flexible and context-dependent process. In some cases, abstraction may even give rise to ad-hoc categories—temporary or context-specific groupings that emerge based on immediate needs or communicative goals. For instance, a cat might be labeled as a "device to catch mice" in one conversation and as a "friend" in another, depending on the context in which the term is used. These ad-hoc categories are a crucial aspect of how speakers dynamically adjust their lexical categories to fit particular situations, further demonstrating the fluidity and contextuality of meaning-making in language.

Although different routes to abstraction are valid pathways for organizing knowledge, there might be sources of variation in language-mediated abstraction pathways during development. In particular, the challenge of disentangling the different semantic relations extends beyond the scope of the present study. However, from a developmental perspective we argue that children may display a variety of routes in language-mediated abstraction processes. This may result in a variety of semantic relations beyond the strict IS-A relation, addressed in the classic language-mediated hierarchical categorization process. We argue that such "noise" (see footnote in the Introduction section)—the presence of non-IS-A relations produced in language-mediated abstractions—may be interpreted as a perception of greater complexity by children, who may therefore resort to a broader range of semantic relationships, beyond taxonomic ones, during language-mediated abstraction tasks. Hence, in a longitudinal perspective, we expect more noise in word ladders produced during the first period of data collection compared with the second. Given the complexity of abstract words, **we hypothesize that abstract concepts may produce more noise than other types of words. Additionally, we expect noise levels to vary based on children's ages**. Here, our standpoint remains quite explorative, as the effect could go in either direction: younger children might generate more noise due to their creative but less conventional vocabulary, or their more limited vocabulary and reliance on rigid taxonomic relations learned in school may lead to less noise overall.

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Method

This study uses a generative word-production approach in which children of different ages (9–13 years old) are asked to construct ladders of words linked by categorical inclusion relations, to investigate how the active vocabulary of words varying in generality/specificity mastered by school-aged children evolves over the span of one school year. We do this by using a hard-copy version of the mobile app Word Ladders, which was developed by the Abstraction research group to collect linguistic data on the semantic relation of categorical inclusion in a gamified environment, also referred to as hypernymy/hyponymy in linguistics, or IS-A relation. For these studies, which involved in-person school interventions and protocols repeated at two different time points during the school year, we chose an analog, paper-based version of the task.

The study was not preregistered. The dataset generated in the current study and analysis scripts are available in the online repository the Open Science Framework (OSF): https://osf.io/a3beq/.

Participants

A total of 195 children, aged 9–13, were recruited from a primary and middle school in Imola (BO), Italy. 37 children with special educational needs or learning disorders were excluded, leaving a final sample of 158. This included two fifth-grade classes ($n = 41$, 20 female, aged 9–10), two sixth-grade classes ($n = 37$, 22 female, age 10–11), two seventh-grade classes ($n = 41$, 22 female, age 11–12), and two eighth-grade classes[2] ($n = 42$, 30 female, age 12–13). The study was approved by the Ethics Committee of the University of Bologna with informed consent and privacy policy signed by the school principal with the approval of the children's parents or caregivers. The school interventions were officially included in the plan of additional educational activities and workshops provided by the school for the school year 2023–2024 (POF 2023–2024, protocol number 0006311, dated 12/06/2023 and classified as "IV.5—Progetti e materiali didattici"), as a formal collaboration between the public institution IC7 public school in Imola (BO), Italy and the research group Abstraction.

Materials

The stimuli used in the study consisted of 16 Italian basic-level nouns:

- four concrete words:
  - "torta," in English "cake";
  - "giocattolo," Eng. "toy";
  - "automobile," Eng. "car";
  - "cavallo," Eng. "horse"
- four abstract words:
  - "libertà," in English "freedom";
  - "tempo," Eng. "time";
  - "mente," Eng. "mind";
  - "sogni," Eng. "dream"

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

- four words denoting emotions:
  - "gioia," in English "joy";
  - "affetto," Eng. "affection";
  - "paura," Eng. "fear";
  - "rabbia," Eng. "anger"
- four neutral social role words:
  - "insegnante," in English "teacher";
  - "medico," Eng. "doctor";
  - "autista," Eng. "driver";
  - "badante," Eng. "caregiver"

Concrete, abstract, and emotional words were selected from the ANEW Italian database (Montefinese et al., 2014), based on familiarity ratings > 7. Social role words, not included in the ANEW database, were selected by the experimenters from grammatically neutral Italian occupation terms because they denote human referents with abstract, socially defined identities. Table 1 presents descriptive statistics for stimuli selected from the ANEW database, along with the results of ANOVA and post hoc comparisons for psycholinguistic (familiarity, frequency) and semantic variables (imageability, concreteness, valence, arousal).

Procedure

Each class was tested individually at the beginning (T1: November 2023) and at the end of the school year (T2: May 2024). After a brief welcome, the experimenters verbally introduced the task and explained the semantic relation IS-A using a graphic tutorial of the Word Ladders app (Figure 1A). Pupils were informed that they would participate in a game involving the creation of word ladders. Each child was provided with paper sheets containing 16 words, each embedded in a ladder (Figure 1B), presented in a randomized order unique to each participant. The aim was to construct the longest word ladders by adding steps above and below an initial prompt. For example, given the word "dog," they should add on the steps above increasingly general words such as "mammal," "animal," and "living being," and more specific ones on the steps below, like "Labrador." Instructions highlighted that the words added should be more general than the previous ones and should describe "a type of." Examples with correct and incorrect semantic relations were provided. For instance, "bread" is a type of "food," and "baguette" is a specific type of bread, but "flour" is not. Similarly, "number" is a type of "quantity," and "negative numbers" are a specific type of number, but a "calculator" is not. The same target words and instructions were used in both experimental sessions, with word order randomized for each child. Each child performed the task individually in a group setting, with the experimenters and teacher supervising, without intervening. Instructions highlighted that longer ladders could earn more points, but only words linked by the "is a type of" relation would score. Pupils were also informed that some ladders might be shorter, and completing all steps was not necessary. Each session lasted about one hour.

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Table 1. Descriptive statistics for semantic types, with ANOVA results and post hoc comparisons on psycholinguistic and semantic variables, i.e., frequency: CoLFIS (Bertinetto et al., 2005), Repubblica (Baroni et al., 2004), familiarity, imageability, concreteness, valence, and arousal (Montefinese et al., 2014)

Variable		Abstract words	Concrete words	Emotions words	*ANOVA* (*p*)	*Post hoc comparisons*
Ln_Colfis	*Mean*	6.62	5.17	5.57	$p = .16$	
	SD	1.06	1.18	.67		
Ln_FreqRep	*Mean*	11.01	8.95	9.73	$p = .05$	
	SD	1.24	.94	.8		
Familiarity	*Mean*	7.25	7.59	7.32	$p = .165$	
	SD	.21	.19	.3		
Imageability	*Mean*	6.04	8.21	7.11	$p < .01$	Concrete > Abstract ($p = .04$)
	SD	1.09	.2	.53		
						Concrete > Emotions ($p < .001$)
						Emotions vs. Abstract ($p = .13$)
Concreteness	*Mean*	4.09	8.05	5.26	$p < .01$	Concrete > Abstract ($p < .001$)
	SD	.8	.37	.25		Concrete > Emotions ($p < .001$)
						Emotions > Abstract ($p = .02$)
Valence	*Mean*	7.16	6.93	5.28	$p = .4$	
	SD	1.36	.55	3.2		
Arousal	*Mean*	5.81	5.63	6.73	$p = .02$	Emotions > Concrete ($p = .04$)
	SD	.38	.5	.5		Emotions > Abstract ($p = .02$)
						Abstract vs. Concrete ($p = .85$)

Coding

The responses were manually entered into a spreadsheet by one of the authors. The words above the target word were coded as hypernyms, whereas those below were coded as hyponyms. Demographic information on gender and age was also recorded. For each ladder, we calculated two measures: *raw length,* the sum of hypernyms and hyponyms produced in response to a word (including the target word), indicating lexical productivity, and *clean length* the sum of valid words in the ladder (including the target word), indicating the validity of the taxonomy. For instance, given the target word "cake," the words "food," "nourishment," "physical energy," and "help" were produced as hypernyms, whereas "tiramisú" and "sugar" were produced as hyponyms. The raw length is 7, and the clean length is 4, corresponding to the valid words in the taxonomy: "nourishment," "food," "cake," "tiramisú" (for other examples, see Table 2). Validity has been evaluated at each step of the ladder, and if an invalid word appeared in the middle of a ladder, subsequent valid-looking words were still counted as valid (see, e.g., the ladder for "freedom" in Table 2). The coding of valid words (necessary to obtain *clean length*)

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

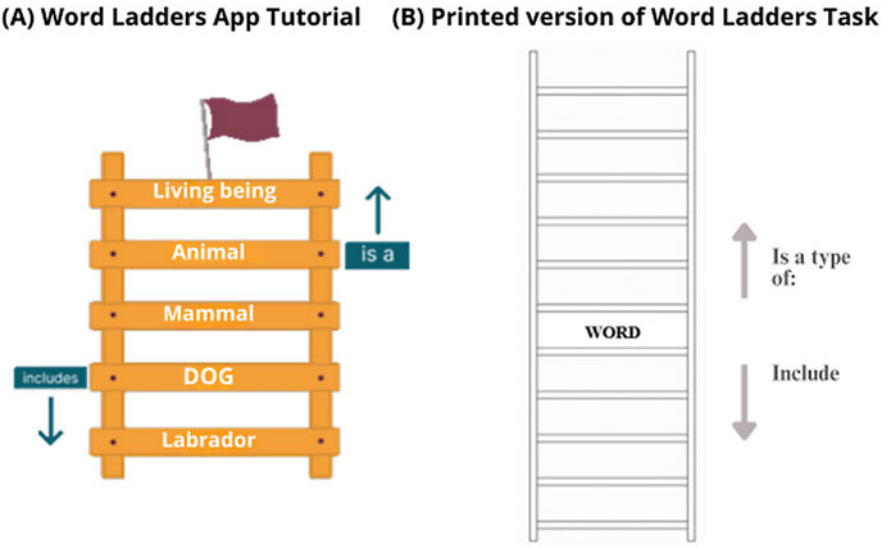


Figure 1. Illustration of the Word Ladders task in a mobile app (A) and a printed (B) version of the experimental protocol.

was done separately for the first and second data collections. In the first data collection, a primary coder initially coded all data using WordNet taxonomies (Fellbaum, 1998; Miller et al., 1990), and common-sense classifications (see Coding Scheme in the OSF repository), with a second coder, naïve to the study aim, independently coding all the data for inter-rater reliability. Any disagreements were resolved by consensus. Cohen's Kappa indicated high reliability (unweighted κ = .84, weighted κ = .95) (Cohen, 1960). For the second data collection, the same two coders split the dataset, with each coding half. Because items were single-coded in this round, inter-rater reliability was not computed. However, coders periodically cross-checked and discussed examples to ensure consistent application of the coding scheme.

Data analysis

Data analysis and visualizations were carried out with R (RCoreTeam, 2019) and RStudio (v4.4.3). We analyzed the *raw length* (word ladder productivity) and *clean length* (word ladder validity) for each ladder separately using generalized linear mixed-effects models with a Poisson distribution and a log link. Models were fitted using the glmer() function from the "lmer4" R's package (Bates et al., 2015). Significance of the main effects and interactions was assessed by conducting Wald Chi-Square tests using Anova () function from "car" R's package (Fox & Weisberg, 2011). Pairwise comparisons were carried out with the "emmeans" R's package (Lenth et al., 2024), using Tukey's adjustment for multiple comparisons. In the following sections, we present three distinct analyses, each addressing one of the research questions, and discuss their results separately.

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Table 2. Example of clean and raw ladders collected with the Word ladders task for each target word. For improved readability, we report the target word in its English translation, and the ladder produced in the original language (Italian) and in its English translation. Note that the interpretation of IS-A relations may be affected by linguistic nuances, as some Italian terms may not have precise English equivalents and can appear uncommon or archaic in translation

	Target word		Clean ladder	Raw ladder
Abstract	Freedom	IT	sensazione, LIBERTA', di parola	sensazione, benessere, salute, LIBERTA', di parola
		ENG	sensation, FREEDOM, of speech	sensation, well-being, health, FREEDOM, of speech
	Mind	IT	entità, cosa astratta, MENTE, mente fantasiosa	entità, cosa, cosa astratta, senso, MENTE, mente fantasiosa, mente catastrofica
		ENG	entity, abstract thing, MIND, imaginative mind	entity, thing, abstract thing, sense, MIND, imaginative mind, catastrophic mind
	Dream	IT	cosa astratta, cosa derivante dal subconscio, SOGNO, sogno felice	cosa astratta, cosa derivante dal subconscio, SOGNO, sogno felice, di un arcobaleno, colori
		ENG	abstract thing, thing derived from the subconscious, DREAM, happy dream	abstract thing, thing derived from the subconscious, DREAM, happy dream, of a rainbow, colors
	Time	IT	entità, cosa astratta, misura, grandezza, TEMPO	entità, cosa astratta, misura, grandezza, TEMPO, cronometro, gara, atletica
		ENG	entity, abstract thing, measure, quantity, TIME	entity, abstract thing, measure, quantity, TIME, stopwatch, race, athletics
Concrete	Car	IT	mezzo di trasporto, AUTOMOBILE, Citroën	mezzo di trasporto, pezzi di macchina, motore, AUTOMOBILE, Citroën, Audi, Lamborghini
		ENG	means of transportation, CAR, Citroën	means of transportation, machine parts, engine, CAR, Citroën, Audi, Lamborghini
	Horse	IT	essere vivente, animale, mammifero, equino, CAVALLO, purosangue, purosangue arabo, purosangue arabo nero	composizione di atomi, insieme di tessuti e organi, essere vivente, animale, mammifero, equino, CAVALLO, purosangue, purosangue arabo, purosangue arabo nero
		ENG	living being, animal, mammal, equine, HORSE, thoroughbred, Arabian thoroughbred, black Arabian thoroughbred	composition of atoms, set of tissues and organs, living being, animal, mammal, equine, HORSE, thoroughbred, Arabian thoroughbred, black Arabian thoroughbred

(*Continued*)

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Table 2. (*Continued*)

	Target word		Clean ladder	Raw ladder
	Toy	IT	gioco, GIOCATTOLO, pupazzo	stato d'animo, emozione, divertimento, gioco, GIOCATTOLO, pupazzo, peluche
		ENG	game, TOY, puppet	state of mind, emotion, fun, game, TOY, puppet, soft toy
	Cake	IT	risorsa, cibo, alimento, dolce, TORTA, al cioccolato	inquinamento, energia, risorsa, cibo, alimento, dolce, TORTA, al cioccolato, alla vaniglia, alle carote
		ENG	resource, food, nourishment, dessert, CAKE, chocolate cake	pollution, energy, resource, food, nourishment, dessert, CAKE, chocolate cake, vanilla cake, carrots cake
Emotions	Affection	IT	emozione, AFFETTO, amore	emozione, pensiero, AFFETTO, dolcezza, amore
		ENG	emotion, AFFECTION, love	emotion, thought, AFFECTION, sweetness, love
	Joy	IT	sensazione, emozione, GIOIA	pensiero, sensazione, emozione, GIOIA, felicità
		ENG	sensation, emotion, JOY	thought, sensation, emotion, JOY, happiness
	Fear	IT	cosa astratta, sensazione, PAURA, dei ragni	cosa astratta, idea, pensiero, emozione, sensazione, PAURA, dei ragni, aracnofobia
		ENG	abstract thing, sensation, FEAR, of spiders	abstract thing, idea, thought, emotion, sensation, FEAR, of spiders, arachnophobia
	Anger	IT	cosa astratta, emozione, RABBIA	cosa astratta, mentalità, sentimento, emozione, RABBIA, rabbia perché si è rotta la penna
		ENG	abstract thing, emotion, ANGER	abstract thing, mentality, feeling, emotion, ANGER, anger because the pen broke
Social role	Driver	IT	entità, mestiere, AUTISTA, autista privato	entità, essere vivente, mammifero, persona, mestiere, AUTISTA, autista privato
		ENG	entity, profession, DRIVER, private driver	entity, living being, mammal, person, profession, DRIVER, private driver
	caregiver	IT	assistente, BADANTE, degli anziani	assistente, aiuto, BADANTE, degli anziani, di bambini
		ENG	assistant, CAREGIVER, of elders	assistant, help, CAREGIVER, of elders, of children

(*Continued*)

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Table 2. (*Continued*)

Target word		Clean ladder	Raw ladder
Teacher	IT	dovere, lavoro, ruolo, INSEGNANTE, professore	dovere, lavoro, ruolo, INSEGNANTE, insegnare, professore, maestra
	ENG	duty, job, role, TEACHER, professor	duty, job, role, TEACHER, teach, professor, primary school teacher
Doctor	IT	essere vivente, umano, persona, lavoratore, MEDICO, chirurgo, chirurgo plastico, chirurgo plastico facciale	essere che vive grazie a qualcosa, essere vivente, essere, umano, persona, lavoratore, MEDICO, chirurgo, chirurgo plastico, chirurgo plastico facciale
	ENG	living being, human, person, worker, DOCTOR, surgeon, plastic surgeon, facial plastic surgeon	being who lives thanks to something, living being, being, human, person, worker, DOCTOR, surgeon, plastic surgeon, facial plastic surgeon

RQ1: Does language-mediated abstraction improve with age (between participants) and over the span of one school year (within participants)?

Analysis 1 tests whether abstraction skills improve over time and if this improvement correlates with children's ages. Given that conceptual representation is flexible and develops over time, we hypothesized variations in word ladders created at the beginning and the end of the school year, reflecting performance improvement. Additionally, we expected a positive correlation between age and the productivity and validity of word ladders. Specifically, as children's age increases, the length of their word ladders will become longer and will include more valid words. We analyzed *raw length* and *clean length* in two separate models, with Time of data collection (T1 vs. T2) as a fixed effect and participants and words as random effects. Spearman correlations were then conducted to explore the relationship between Age (9–10, 10–11, 11–12, and 12–13) and each dependent variable, first on the overall data and then separately for T1 and T2.

Results

Time. We found a significant main effect of time on both the productivity and validity of word ladders. Raw length was significantly higher at T2 ($M = 5.06$, $SD = 1.89$) than at T1 ($M = 4.43$, $SD = 1.89$), $\chi^2(1) = 69.25$, $p < .001$. Similarly, clean length was higher at T2 ($M = 3.89$, $SD = 1.75$) than at T1 ($M = 2.75$, $SD = 1.35$), $\chi^2(1) = 415.22$, $p < .001$.

Age. Overall, we found a very small positive correlation between age and raw length ($rs = .04$, $p = .005$), and no significant correlation between age and clean length ($rs = -.01$, $p = .731$). At T1, we found a slightly stronger positive correlation between age and raw length ($rs = .08$, $p < .001$), and a weak positive correlation with clean length ($rs = .04$, $p = .032$). At T2, the positive

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

correlation between age and raw length decreased to a very small, nonsignificant value (rs = .02, p = .318), and the correlation between age and clean length approached zero, with no significance (rs = −.01, p = .658).

Discussion

In line with our first hypothesis, we found a clear improvement in both productivity and validity of the word ladders created by children from the beginning to the end of the school year (within participants). Our second hypothesis was not supported. Results showed very weak correlations between children's age and the productivity and validity of word ladders across both time points (between participants). At T1, correlations were slightly positive for each measure, and at T2, the correlations were not meaningful. Taken together, these findings suggest that developmental progression over time, rather than chronological age, drives improvements in children's ability to produce consistent categorizations. Both younger and older children appear to develop abstraction abilities throughout the school year, largely independent of their initial performance level.

RQ2: How does semantics (in particular concrete vs. abstract) influence children's ability to construct ladders and therefore to perform language-mediated abstractions over time/age?

Analysis 2 tests whether word semantics affect children's performance on the word ladders task across ages and over time. Building on previous research showing an early acquisition of concrete over abstract words, we hypothesized that ladders based on concrete words will be longer and overall will contain more valid words than those based on abstract words. Emotion words are learned earlier and perceived as less abstract than neutral abstract words. Similarly, social role-related words combine both concrete and abstract elements, representing human beings who embody various social phenomena. Considering these factors, it may be easier for children to construct ladders for emotion and social role words than for more abstract words. Additionally, these effects may vary with age and over time: younger children might initially find it more challenging to create long and valid ladders for abstract words than for concrete words. However, their improvement over time may be more pronounced than that of older children.

To test these hypotheses, we expanded our analysis by modeling the dependent variable *raw length* and *clean length* separately, using Age (9–10, 10–11, 11–12, and 12–13), Semantic Type (abstract, concrete, emotions, and social role), Time (T1 vs. T2), and their interaction as fixed effects, with participants and words as random effects.

Results

Productivity: raw length

We found significant main effects of Semantic Type, $\chi^2(3) = 80.52$, $p < .001$, and Time, $\chi^2(1) = 68.09$, $p < .001$, and significant interactions between Age and Semantic Type, $\chi^2(9) = 24.19$, $p = .004$, Age and Time, $\chi^2(3) = 29.33$, $p < .001$, and

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Semantic Type and Time, $\chi^2(3) = 10.63$, $p = .013$. No other main effects or interactions reached significance; all $ps > .06$.

Pairwise comparison: Semantic Types. Children produced shorter ladders for abstract words than for concrete ($z = -7.92$, $p < .001$) and social role words ($z = -6.74$, $p < .001$). They also generated shorter ladders for emotions compared with social role words ($z = -4.62$, $p < .001$). In contrast, they produced longer ladders for concrete words than for emotional words ($z = 5.80$, $p < .001$). No difference was found between abstract and emotional words ($z = -2.13$, $p = .14$), and between concrete and social roles words ($z = 1.18$, $p = .64$) (see Figure 2).

Pairwise comparison: Semantic Type and Age. Children aged 9–10 years generated shorter ladders for abstract words compared with concrete ($z = -7.71$, $p < .001$), emotions ($z = -2.62$, $p = .043$), and social role words ($z = -5.43$, $p < .001$). They also produced shorter ladders for emotions compared with social role words ($z = -2.82$, $p = .025$). In contrast, they produced longer ladders for concrete words compared with emotional ones ($z = 5.12$, $p < .001$). There was no difference in raw length between ladders for concrete and social role words ($p = .09$). For children aged 10–11, 11–12, and 12–13 years, a consistent pattern emerged. In all three groups, ladders for abstract words were shorter than those for concrete words (10–11-year-olds: $z = -4.39$, $p < .001$; 11–12-year-olds $z = -4.97$, $p < .001$; 12–13-year-olds: $z = -7.27$, $p < .001$) and social role words (10–11-year-olds: $z = -4.24$, $p < .001$; 11–12-year-olds: $z = -4.70$, $p < .001$; 12–13-year-olds: $z = -6.29$, $p < .001$). Ladders for concrete words were longer than those for emotional words (10–11-year-olds: $z = 3.10$, $p = .011$; 11−12-year-olds: $z = 4.55$, $p < .001$; 12–13-year-olds: $z = 5.26$, $p < .001$), whereas ladders for emotional words were shorter than for social role ones (10-11-year-olds: $z = -2.95$, $p = .017$; 11–12-year-olds: $z = -4.28$, $p = .001$; 12–13-year-olds: $z = -4.27$, $p = .001$). Across these groups, there were no differences in raw length between ladders for abstract and emotional words (10–11-year-olds $p = .56$; 11–12-year-olds $p = .97$; 12–13-year-olds $p = .17$) or between ladders for concrete and social role words (10–11-year-olds $p = .10$; 11–12-year-olds $p = .99$; 12–13-year-olds $p = .75$). We also found that ladders for abstract words were significantly shorter in 9–10-year-olds compared with 10-11-year-olds ($z = -3.37$, $p = .004$) and 11–12-year-olds ($z = -3.24$, $p = .007$). Conversely, ladders for abstract words were longer in 10–11-year-olds ($z = 2.94$, $p = .017$) and 11–12-year-olds ($z = 2.79$, $p = .027$) compared with 12–13-year-olds. There were no differences between age groups in raw length of ladders for concrete ($ps > .95$), social roles ($ps > .06$), and emotional words ($ps > .13$).

Pairwise comparison: Age and Time. At T1, word ladders were significantly shorter in 9–10-year-olds compared with 10–11-year-olds ($z = -3.31$, $p = .005$) and 12–13-year-olds ($z = -2.57$, $p = .05$). No other comparisons reached significance, all $ps > .26$. At T2, word ladders were significantly shorter in 11–12-year-olds compared with 12–13-year-olds ($z = 3.36$, $p = .004$). No other comparisons reached significance, $ps > .07$. We also found that word ladders were longer at T2 compared with T1 in 9–10-year-olds ($z = -5.68$, $p < .001$), 10–11-year-olds ($z = -2.47$, $p = .014$), and 11–12-year-olds ($z = -7.25$, $p < .001$). No difference was found in 12–13-year-olds, $p = .51$.

Pairwise comparison: Semantic Type and Time. Children produced longer ladders at T2 than T1 for concrete ($z = -4.00$, $p = .001$), emotions ($z = -3.12$, $p = .0018$), and

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

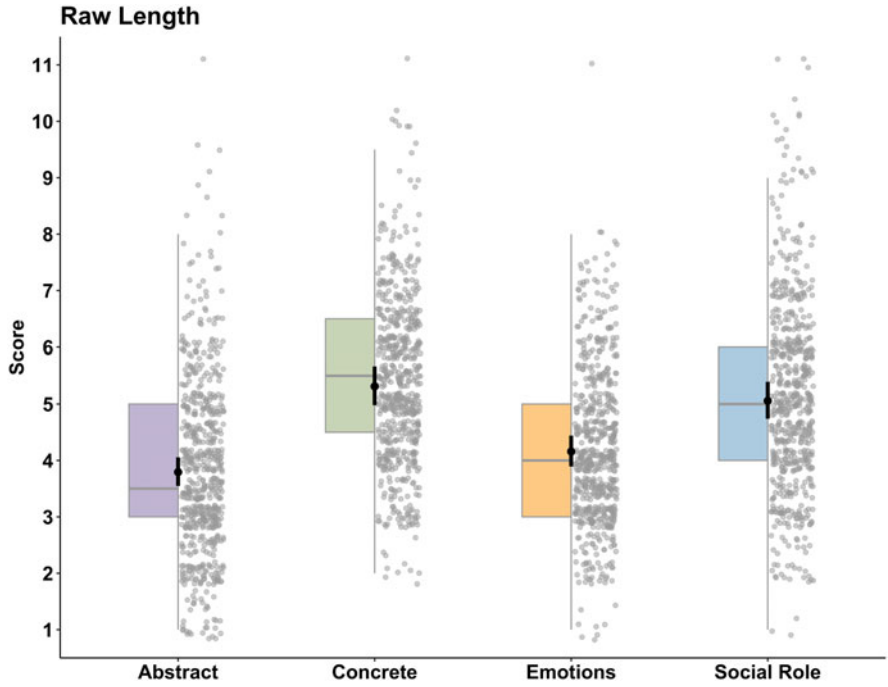


Figure 2. Raw Length scores (word ladders productivity) as a function of semantic type. Raw data in the background are aggregated over participants.

social role words ($z = -6.14$, $p < .001$). No difference was found for abstract words, $p = .23$. At both times, children produced shorter ladders for abstract words than for concrete (T1: $z = -6.55$, $p < .001$; T2: $z = -7.73$, $p < .001$) and social role words (T1: $z = -4.80$, $p < .001$; T2: $z = -7.32$, $p < .001$). They also produced shorter word ladders for emotional words than for social role words, T1: $z = -3.44$, $p = .003$; T2: $z = -4.90$, $p < .001$, while they produced longer ladders for concrete words compared with emotional ones (T1: $z = 5.20$, $p < .001$; T2: $z = 5.31$, $p < .001$). No difference was found between concrete and social role words (T1: $p = .29$; T2: $p = .98$) and between abstract and emotional words (T1: $p = .53$; T2: $p = .07$) (Figure 3).

Validity: clean length

We found significant main effects of Semantic Type, $\chi^2(3) = 69.15$, $p < .001$, and Time, $\chi^2(1) = 410.26$, $p < .001$, and significant interactions between Age and Time, $\chi^2(3) = 11.12$, $p = .011$, and Semantic Type and Time, $\chi^2(3) = 9.01$, $p = .029$. No other main effects or interactions reached significance, all ps $> .52$.

Pairwise comparison: Semantic Type. Children produced fewer valid words in the ladders for abstract words than for concrete ($z = -7.46$, $p < .001$), emotional ($z = -2.60$, $p = .046$), and social role words ($z = -6.272$, $p < .001$). In contrast, they produced more valid words in the ladders for concrete words than for emotional ones ($z = 4.87$, $p < .001$), and fewer valid words in the ladders for emotional words

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

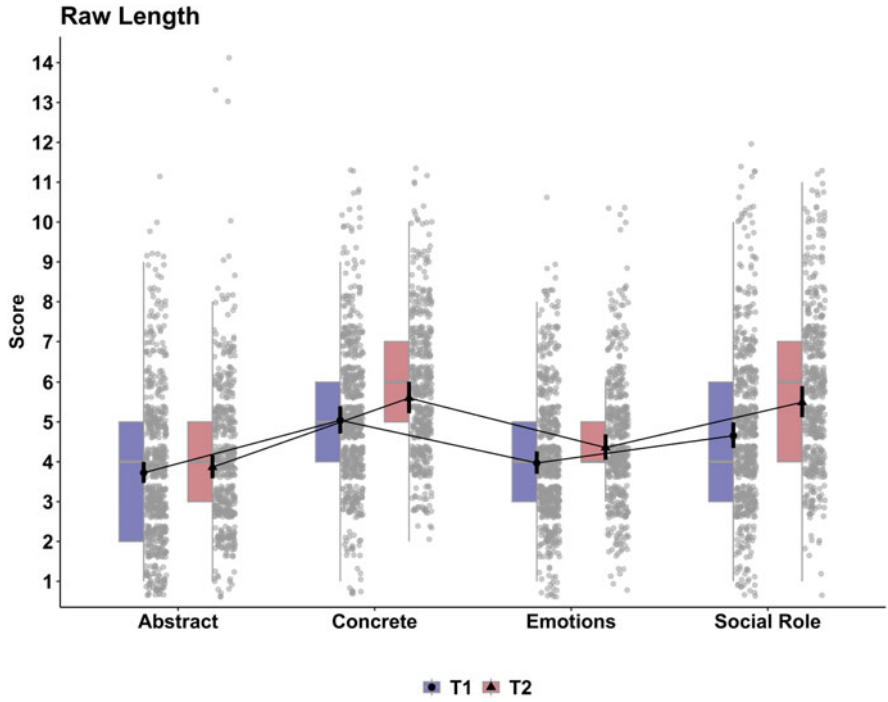


Figure 3. Raw Length scores (word ladder productivity) for each semantic type as a function of time. Raw data in the background are aggregated over participants.

than for social role ones ($z = -3.68$, $p = .001$). No difference was found in the validity of ladders between concrete and social role words ($p = .63$) (see Figure 4).

Pairwise comparison: Age and Time. While the interaction was significant in the overall model, we found no difference in word ladders validity between the age groups at either time point (T1 $ps > .26$ and T2 $ps > .21$). However, the number of valid words in the ladders significantly increased at T2 compared with T1 in all age groups (9–10-year-olds: $z = -10.14$, $p < .001$; 10–11-year-olds: $z = -11.20$, $p < .001$; 11–12-year-olds: $z = -12.14$, $p < .001$; 12–13-year-olds: $z = -5.72$, $p < .001$.

Pairwise comparison: Semantic Type and Time. Word Ladders validity significantly increased from T1 to T2 for each semantic type (all $ps < .001$). At both times, children produced fewer valid words in ladders for abstract words than for concrete (T1: $z = -7.71$, $p < .001$; T2: $z = -6.68$, $p < .001$), and social role words (T1: $z = -5.96$, $p < .001$; T2: $z = -6.12$, $p < .001$). They also produced more valid words in ladders for concrete words than for emotional words (T1: $z = 4.95$, $p < .001$; T2: $z = 4.48$, $p < .001$), and they produced fewer valid words in ladders for emotional words than for social role ones (T1: $z = -3.20$, $p = .007$; T2: $z = -3.92$, $p = .005$). No difference was found in ladders validity between concrete and social role words at either time point (T1: $p = .29$; T2: $p = .94$). However, a key difference emerged: at T1, children produced fewer valid words in ladders for abstract words compared with emotional ones ($z = -2.78$, $p = .028$), whereas at T2, such difference was no longer observed ($p = .12$) (Figure 5).

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

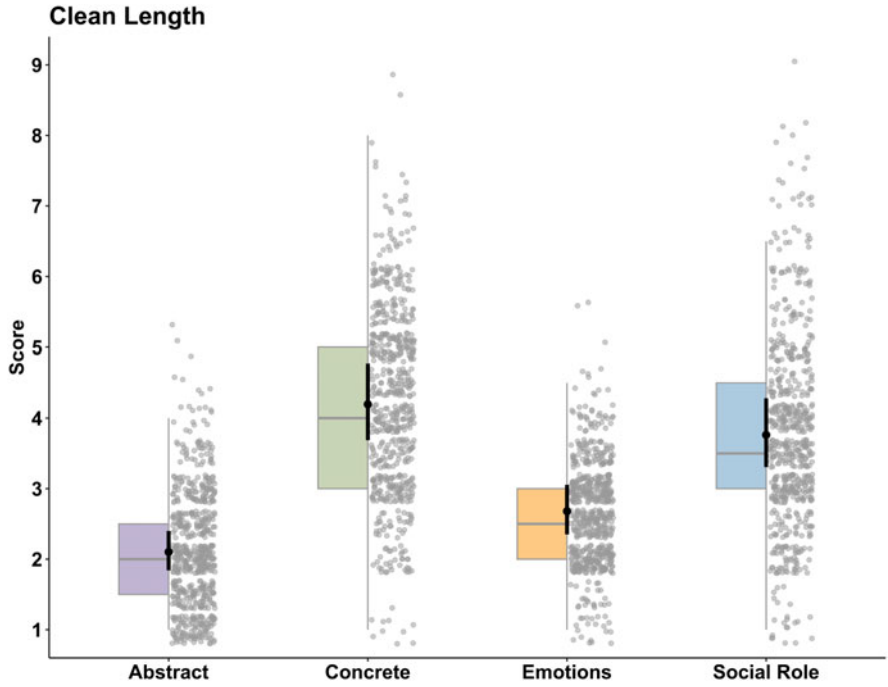


Figure 4. Clean Length scores (word ladder validity) as a function of semantic types. Raw data in the background are aggregated over participants.

Discussion

Overall, as predicted, children produced longer and more valid ladders for concrete concepts (e.g., "horse," "toy") than for abstract concepts (e.g., "time," "mind"). Ladders for concepts expressing social roles (e.g., "doctor," "driver") were similar in length and validity to those for concrete ones, but longer and more valid than those for abstract and emotional concepts. Ladders for emotional concepts (e.g., "fear," "joy") were shorter than those for concrete ones, with similar productivity but higher validity than ladders for abstract concepts. A similar pattern emerged across the two times of data collection. However, while ladders validity increased from T1 to T2 for all semantic types, ladders productivity increased only for concrete, emotional, and social role concepts, with abstract concepts consistently yielding shorter ladders. Interestingly, at T2 no difference was found between ladder validity for abstract and emotional concepts. Notably, we found age differences primarily in ladders based on abstract concepts. While there was no difference in ladders productivity for concrete, emotional, and social role concepts across age groups, the youngest children (9–10) created shorter ladders for abstract concepts than emotional ones, a pattern not seen in older groups (10–13). The youngest children (9–10) also produced shorter ladders overall than older children (10–11 and 11–12), with no difference compared with the oldest group (12–13). In contrast, 12–13-year-olds produced longer ladders overall than the 10–11 and 11–12 groups, with no significant difference between the latter two. Consistent with this, ladders productivity improves from T1 to T2, with significant increases in younger children, but no improvement in the oldest group

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

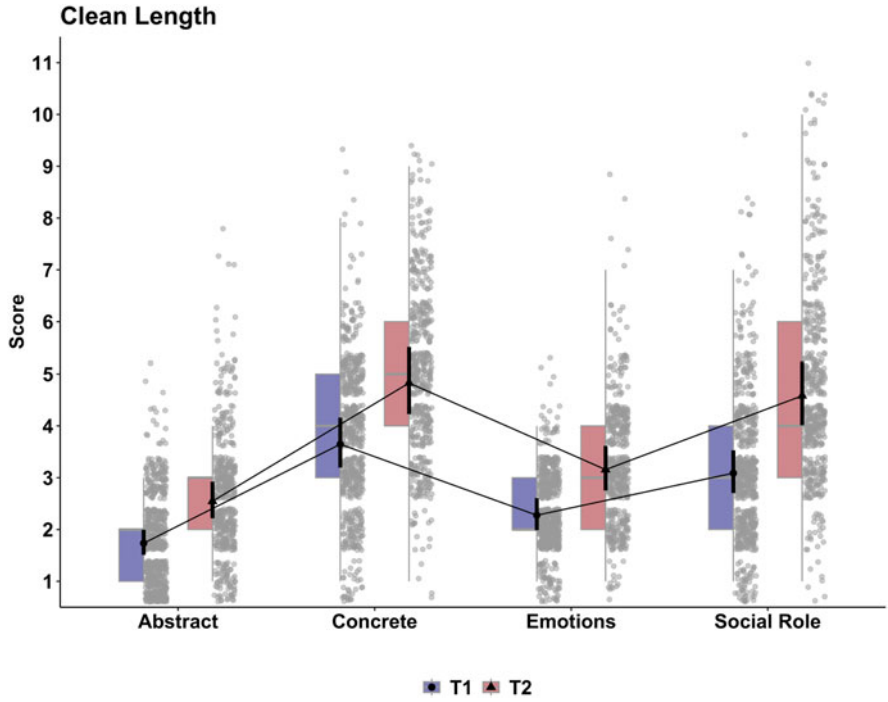


Figure 5. Clean Length scores (word ladder validity) for each semantic type as a function of time. Raw data in the background are aggregated over participants.

(12–13). Contrary to our predictions, we found no significant interaction between word semantics and ladders validity, suggesting a similar pattern in all age groups. Nevertheless, ladders validity increased from T1 to T2 for all age groups, with no significant differences between them.

Overall, these findings indicate that children's language-mediated abstraction abilities show robust developmental gains over time, but that these gains are constrained by the semantic nature of the concepts involved. Across the school year, children became increasingly capable of constructing longer and more valid semantic hierarchies for concrete concepts, emotional concepts, and concepts denoting social roles. In contrast, truly abstract concepts consistently posed greater challenges; although children improved the *validity* of the semantic relations they produced, they did not extend these hierarchies in length over time. This pattern suggests that while children gradually learn to organize abstract concepts more coherently, their capacity to elaborate such representations remains limited during this developmental window.

RQ3: To what extent is the abstraction pathway shared by speakers during development?

Analysis 3 explores variations in language-mediated abstraction pathways during development. Although the Word Ladders task required the production of

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

categorical inclusion relations, children often generated other semantic relations in response to a given prompt, such as free-associations (e.g., prompt: "medico"; response: "ospedale," in English: "doctor-hospital"; or "tempo-orologio," in English: "time-clock") or meronyms relations (e.g., "automobile-ruote," in English: "car-wheels"; or "affetto-abbraccio," in English: "affection-hug") (See Raw ladders in Table 2). While a fine-grained inspection of these semantic relations is currently addressed in a separate study, here we focused on a quantitative analysis of the distribution of non-IS-A relations in our data. This approach provides an initial understanding of the different linguistic abilities underlying abstraction, specifically the processes of generalization and semantic categorization, associated with hypernyms production, and the process of specification through linguistic creativity and compositionality, associated with hyponyms production.

To this end, we introduced a new variable, labeled "noise," defined as the ratio of *raw length* (the total number of words in the ladder) to *clean length* (the number of valid words in the ladder). We interpreted the *noise* variable as follows:

- Noise = 1: all produced words are valid within the taxonomy;
- Noise > 1: there are words that are not valid within the taxonomy, with higher values indicating greater noise.

The noise variable provides two key insights: first, it serves as an indicator of the complexity involved in generating taxonomic relationships for a given word, and second, it offers a deeper understanding of the broader impact of language on the abstraction process. Higher noise values reflect greater complexity, as children produce a wider range of semantic relationships that extend beyond taxonomic relations, thereby highlighting greater linguistic flexibility and creativity. Overall, we expected more noise in word ladders produced during the first period of data collection compared with the second one. Given the complexity of abstract words, we hypothesized more noise in word ladders for abstract words than for other concepts kind. Additionally, we hypothesized variations in word ladders noise as a function of children's ages (a bidirectional hypothesis). Younger children, who tend to have a more creative yet less conventional vocabulary, may produce word ladders with more noise. However, due to their more limited vocabulary, it is also possible that younger children might exhibit less noise.

Moreover, we expected noise to be distributed differently above and below the target words, with more noise for hypernyms and less for hyponyms. This can be understood by considering the structure of linguistic taxonomies, such as WordNet. For example, given a target word like "dog," its hypernyms (more general terms) tend to converge into a few increasingly broader categories, often based on linguistic conventions (e.g., dog → mammal → animal → living being → entity). These labels may not be fully established in school-age children, who might rely on more creative linguistic expressions when forming general categories. In contrast, possible valid hyponyms of "dog" (more specific terms) include a wide variety of co-hyponymic options (e.g., dalmatian, chihuahua, labrador, etc.), leaving less room for errors.

To test these hypotheses, we first modeled the *noise* variable on the entire dataset. Next, we conducted a Welch two-sample *t*-test to compare noise levels between

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

hypernyms and hyponyms, followed by separate models for each group. The structure of the models was the same as in Analysis 2.

Results

Nontaxonomic relation: noise

We found significant main effects of Age, $\chi^2(3) = 11.47$, $p = .009$, Semantic type, $\chi^2(3) = 28.21$, $p < .001$, and Time, $\chi^2(1) = 323.31$, $p < .001$, showing that noise was significantly higher at T1 than at T2, $z = 18.02$, $p < .001$. We also found significant two-way interactions between Age and Semantic type, $\chi^2(9) = 24.34$, $p = .004$; Age and Time, $\chi^2(3) = 27.33$, $p < .001$; Semantic type and Time, $\chi^2(3) = 50.18$, $p < .001$; and the three-way interaction of Age, Semantic type, and Time, $\chi^2(9) = 17.98$, $p = .035$. For clarity and interpretability, we reported only pairwise comparisons for the main effects and the two-way interactions.

Pairwise comparison: Age. We found more noise in ladders produced by 10–11-year-olds compared with 9–10-year-olds ($z = -2.80$, $p = .03$) and 12–13-year-olds ($z = 2.62$, $p = .04$). No other comparisons reached significance, all ps $> .27$.

Pairwise comparison: Semantic Type. We found more noise in ladders based on abstract words compared with concrete ($z = 4.67$, $p < .001$), emotional ($z = 2.80$, $p = .03$), and social role words ($z = 4.13$, $p < .001$). No other comparisons reached significance (all ps $> .24$).

Pairwise comparison: Age and Time. The noise level was higher at T1 than at T2 across all age groups (ps $< .001$). At T1, 10–11-year-olds showed more noise compared with 9–10-year-olds ($z = -4.42$, $p < .001$) and 12–13-year-olds ($z = 3.28$, $p = .005$), whereas at T2, no significant differences in noise level were found among age groups (all ps $> .38$).

Pairwise comparison: Age and Semantic Type. Children aged 9–12 years produced more noise in ladders with abstract words compared with concrete (9–10-year-olds: $z = 3.85$, $p = .001$; 10–11-year-olds: $z = 5.01$, $p < .001$; 11–12-year-olds: $z = 5.46$, $p < .001$), emotional (9–10-year-olds: $z = 2.62$, $p = .044$; 10–11-year-olds: $z = 2.71$, $p = .034$; 11–12-year-olds: $z = 3.40$, $p = .004$), and social role words (9–10-year-olds: $z = 3.54$, $p = .002$; 10–11-year-olds: $z = 4.23$, $p = .001$; 11–12-year-olds: $z = 4.62$, $p < .001$). Among 12–13-year-olds, abstract words generated more noise than concrete ($z = 2.95$, $p = .017$) and social role words ($z = 2.87$, $p = .021$), but no difference was found compared with emotional words ($p = .35$). Across all age groups, noise levels did not differ for concrete, emotional, and social role words (ps $> .09$) (see Figure 6). Additionally, we found that 9–10-year-olds and 12–13-year-olds produced less noise than 10–11-year-olds and 11–12-year-olds in ladders with abstract words (9–10 vs. 10–11: $z = -3.08$, $p = .011$; 10–11 vs. 12–13: $z = 3.69$, $p = .001$; 11–12 vs. 12–13: $z = 3.44$, $p = .003$). Additionally, 9–10-year-olds generated less noise than 10–11-year-olds in ladders with emotional words ($z = -2.91$, $p = .02$). There were no significant differences in noise levels across age groups in ladders for concrete (ps $> .55$) and social role words (ps $> .18$).

Pairwise comparison: Semantic Type and Time. Noise levels were higher at T1 than at T2 across all semantic types (ps $< .001$). Noise in ladders for concrete, emotional, and social role words did not differ at either time point (T1: ps $> .18$;

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

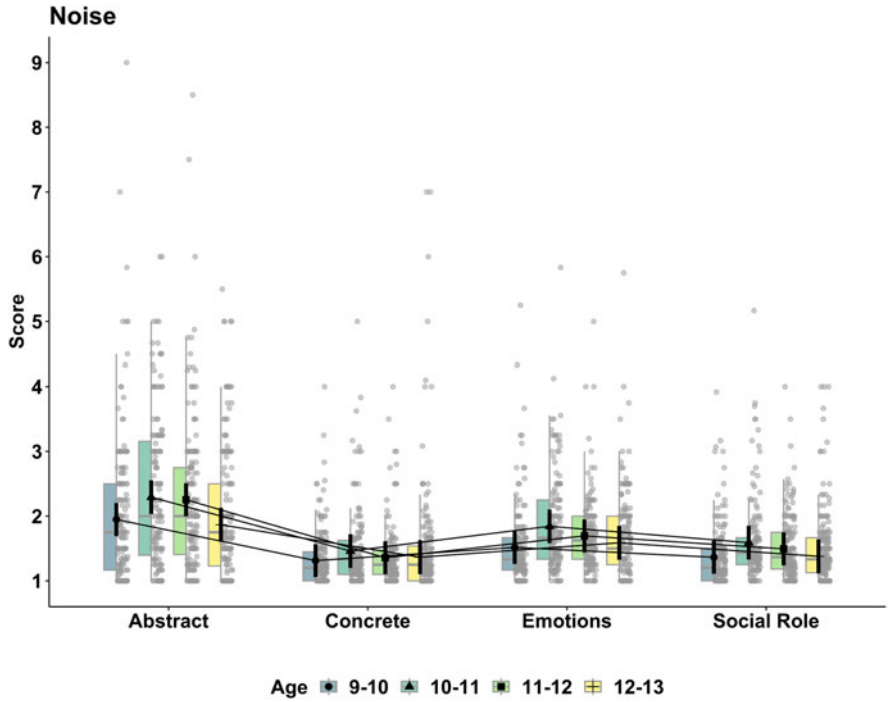


Figure 6. Noise scores (nontaxonomic relation) for each semantic type as a function of children's age. Raw data in the background are aggregated over participants.

T2 *p*s > .37). In contrast, at both time points abstract words yielded more noise than concrete (T1: $z = 6.00$, $p < .001$; T2: $z = 3.10$, $p = .011$) and social role words (T1: $z = 5.34$, $p < .001$; T2: $z = 2.70$, $p = .035$), but only at T1 abstract words yield more noise than emotional words ($z = 3.98$, $p < .001$).

Nontaxonomic relation: noise hypernyms and hyponyms

A Welch two-sample *t*-test showed that noise was significantly higher in hypernyms ($M = 1.58$) than in hyponyms ($M = 1.42$), $t(6269.6) = 7.54$, $p < .001$.

Hypernyms. We found significant main effects of Age, $\chi^2(3) = 10.64$, $p = .014$, Semantic Types, $\chi^2(3) = 11.71$, $p = .008$, and Time $\chi^2(1) = 66.11$, $p < .001$, showing that noise in hypernyms was significantly higher at T1 than at T2 ($z = 7.55$, $p < .001$). We also found significant two-way interactions between Age and Semantic type, $\chi^2(9) = 17.64$, $p = .040$, and between Semantic type and Time, $\chi^2(3) = 12.84$, $p = .005$; and the three-way interaction of Age, Semantic type, and Time, $\chi^2(9) = 20.22$, $p = .017$.

Pairwise comparison: Age. Children aged 9–10 years produced less noise in hypernyms compared with 10–11-year-olds ($z = -3.18$, $p = .008$). No other comparisons reached significance (*p*s > .08).

Pairwise comparison: Semantic Types. Ladders based on concrete words showed less noise in hypernyms than those based on emotional words ($z = -2.94$, $p = .017$). No other comparisons reached significance (all *p*s > .06).

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Pairwise comparison: Age and Semantic Type. Children aged 10–11 years produced more noise in hypernyms than 9–10-year-olds in ladders with abstract ($z = -2.60$, $p = .04$), emotional ($z = -3.46$, $p = .003$), and social role words ($z = -2.88$, $p = .02$). No other comparison reached significance ($ps > .06$). Moreover, 10–11-year-olds ($z = 3.16$, $p = .009$) and 11–12-year-olds ($z = 2.85$, $p = .02$) produced more noise in hyponyms for ladders based on concrete words than those based on abstract words. Additionally, 10–11-year-olds produced less noise in hyponyms for ladders with concrete words compared with emotional ones ($z = -3.95$, $p < .001$). No significant difference was found in the youngest ($ps > .21$) and oldest groups ($ps > .12$).

Pairwise comparison: Semantic Type and Time. Noise in hypernyms decreased significantly from T1 to T2 for ladders with concrete ($z = 2.15$, $p = .031$), emotional ($z = 7.06$, $p < .001$), and social role words ($z = 5.74$, $p < .001$), but not for ladders with abstract words ($z = 1.82$, $p = .069$). At T1, noise levels in hyponyms were higher in ladders for emotional compared with concrete ones ($z = -3.94$, $p < .001$). No other comparisons reached significance at either time point (all $ps > .08$).

Nontaxonomic relation: noise hyponyms. We found a significant main effect of Time $\chi^2(1) = 186.80$, $p < .001$, showing that noise in hyponyms was significantly higher at T1 than at T2, $z = 12.78$, $p < .001$. We also found a significant two-way interaction between Age and Time, $\chi^2(3) = 24.72$, $p < .001$. Noise in hyponyms decreased from T1 to T2 in all age groups, $ps \leq .0004$. However, there were no significant differences between age groups at each time point (T1: all $ps > .42$; T2: all $ps > .10$). No other main effects or interactions reached significance, all $ps > .09$.

Discussion

We observed a significant reduction from T1 to T2 in the proportion of nontaxonomic relations—hereafter referred to as *noise*—across all semantic types and age groups. By *noise*, we mean responses that did not reflect valid hierarchical (*IS-A*) relations within a ladder, including associative, thematic, or otherwise semantically inappropriate links. This overall decrease indicates that, over time, children become increasingly selective and precise in constructing taxonomic relations, producing fewer invalid or loosely connected words. Although noise generally declined, its distribution varied as a function of semantic type and age. For concrete, emotional, and social role concepts, noise levels were relatively stable across ages and time points. In contrast, ladders based on abstract concepts displayed greater variability. At T1, abstract concepts elicited significantly more noise than all other semantic types, suggesting particular difficulty in identifying appropriate taxonomic relations for these concepts. By T2, however, noise levels for abstract concepts no longer differed from those observed for emotional concepts, indicating partial convergence as children refined their semantic representations.

Age-related differences further qualified this pattern. Overall, children aged 10–11 years produced more noise than younger (9–10) and older (11–12, 12–13) groups, although this difference diminished over time. Children aged 9–12 showed elevated noise specifically for abstract concepts relative to other semantic types, whereas the oldest group (12–13) exhibited increased noise for abstract concepts

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

only in comparison with concrete and social role concepts. Notably, both the youngest (9–10) and oldest (12–13) children generated less noise in abstract ladders than the intermediate age groups (10–11 and 11–12), and the youngest group also produced less noise for emotional concepts than 10–11-year-olds. Consistent with our predictions, noise was higher for hypernyms than for hyponyms, reflecting greater difficulty in identifying valid superordinate categories than subordinate ones. Noise in both hypernyms and hyponyms decreased significantly from T1 to T2, indicating improved hierarchical control over time. However, age- and semantics-related differences emerged only for hypernyms. In particular, noise in hypernyms declined over time for most semantic types, with the notable exception of abstract concepts. Moreover, abstract and emotional concepts generated more hypernym-related noise than concrete concepts in the 10–11 and 11–12 age groups, a pattern not observed in the youngest or oldest children.

Taken together, these findings suggest that although children's production of nonvalid hierarchical semantic relations decreases over time, noise is differentially modulated by age and by the semantic properties of the concepts involved. Intermediate age groups (particularly 10–11-year-olds) appear to be in a transitional phase, characterized by increased exploration but reduced precision. Abstract concepts consistently elicit higher levels of noise, especially when children are required to move toward more general levels of abstraction, highlighting the persistent challenge of constructing coherent hierarchical representations for truly abstract meanings.

General discussion and conclusions

In a longitudinal study, we investigated the development of language-mediated abstraction using a generative word-production task. Students aged 9–13 were asked to construct word ladders, linking words through semantic relations of categorical inclusion (IS-A relations), for a set of initial prompts varying in concreteness. The task was administered at two points during the school year. To operationalize and assess language-mediated abstraction skills, we measured ladder productivity (the total number of words produced starting from one prompt) and ladder validity (the number of valid words produced starting from one prompt), both reflecting children's overall ability to hierarchically organize their active vocabulary. Additionally, we quantified the occurrence of non-IS-A semantic relations as ladder *noise*, arguing that this variable captures sources of variation in language-mediated abstraction pathways.

Comparing task performance within and between participants we observed that language-mediated abstraction skills improved over time but are not clearly associated with participants' age. Specifically, across all age groups, ladders produced at the end of the school year were on average longer and contained on average more valid words than those created at the beginning of the school year. This finding partially aligns with previous research showing that children become better at identifying valid taxonomic relations as they progress in their education (Lucariello et al., 1992; Nelson, 1988). The absence of a direct age effect may reflect

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

the influence of task-specific training over the school year, enabling some younger children to reach performance levels comparable to their older peers. Another possible explanation is that the task may be more sensitive to the meaning of words used as prompts than to general cognitive development, with age affecting performance only indirectly through differences in children's knowledge of formal semantic hierarchies associated with different word types.

As shown in the second analysis, word semantics influenced children's ability to perform language-mediated abstraction. As expected, ladders generated in response to concrete concepts were, on average, longer and contained more valid words compared with those produced for other semantic categories. In contrast, ladders for abstract concepts were the shortest and contained the highest proportion of noise. Concepts categorized as emotional and social role yielded intermediate results; ladders for emotional concepts contained more valid words on average but were similar in length to those for abstract concepts, while ladders for social-role concepts were comparable in both length and validity to those for concrete concepts. This suggests that language-mediated abstraction is facilitated when words are grounded in real-world experiences, whereas identifying hierarchical relations for abstract concepts remains more challenging. Our findings support theories that classify abstract concepts as "hard words" to acquire and process (Gleitman et al., 2005), aligning with recent research showing that semantic categories differing in concreteness scores also vary along the specific–general continuum, resulting in either longer or shorter taxonomies (Villani et al., 2024). Importantly, we replicate and extend this line of research by applying an innovative generative task within a developmental framework.

Our findings contribute to language development research, and particularly to the development of abstract language. Indeed, we found that while all age groups showed similar ladder productivity for concrete, emotional, and social concepts, the youngest children (9–10) produced shorter ladders for abstract concepts than for emotional concepts, a pattern absent in older children (10–13). In terms of ladder validity, we found no age-related differences across semantic types, indicating that children of all ages can identify meaningful taxonomic relations, regardless of concept type. However, abstract and emotional concepts consistently yielded less valid ladders than concrete and social role concepts. Interestingly, by the second data collection, the ladder validity for abstract concepts was comparable to emotional one. This suggests that abstract thinking develops later than emotional processing, and as children's emotional vocabulary refines, their ability to form complex associations with abstract concepts improves. These results align with the "Abstract-via-Emotion" hypothesis (Kousta et al., 2011; Vigliocco et al., 2009; Ponari et al., 2018; Lund et al., 2019), which posits that emotional states provide a bootstrap mechanism for the development of abstract words and concepts. However, the observed pattern could also be driven by differences in frequency, arousal, or valence between emotional and abstract words, greater exposure to emotional vocabulary through school and home discourse, and the availability of conventional superordinate categories such as "emotion" and "feeling" that facilitate taxonomic organization for emotion words. More broadly, the intermediate profile of word ladders for emotional and social-role terms may also reflect that these categories draw on grounding mechanisms that partially overlap with those

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

supporting concrete concepts, including embodied experiences of emotional states, perceptual processing of faces, and social interaction (e.g., Catricalà et al., 2020; Dreyer & Pulvermüller, 2018). The fact that different semantic categories yielded systematically different word ladders is consistent with the view that the abstract concepts domain is inherently multidimensional, with distinct sub-categories recruiting different grounding mechanisms (Villani et al., 2019, 2021; for a review see Conca et al., 2021). Our findings extend this view by suggesting that differences in grounding mechanism affect taxonomic organization, with more embodied and perceptually grounded categories supporting more extended and stable taxonomic structure.

Finally, we explored the impact of language on the abstraction process by analyzing the distribution of non-IS-A relations in our data, operationalized as "noise." As predicted, noise was higher in ladders with abstract and emotional concepts than with concrete and social role concepts. A plausible explanation is that abstract and emotional concepts are more complex, and their content is more strongly defined by social constructs (Barrett, 2016). This makes them more reliant on linguistic support to be acquired, and thereby their conceptual representation is more linked to linguistic associations than sensorial experiences (Borghi, 2023). Surprisingly, the noise effect was most pronounced in intermediate-age children (10–11 and 11–12 years), whereas it was not observed in younger (9–10 years old) or older children (12–13 years old). The elevated noise levels in the intermediate group suggest that middle childhood might represent a transitional developmental phase during which children undergo a critical process of constructing, refining, and consolidating their understanding of hierarchical concept organization.

Younger and older children showed lower noise levels. One possible explanation is that younger children (aged 9–10 years in our sample) may rely more heavily on input learned at school and at home through parental scaffolding to construct their vocabulary, resulting in more guided but less flexible generalizations. Older children (aged 12–13), having received more formal education, may have internalized more adult-like taxonomies, consciously applying them during the task. As a result, both younger and older children are less likely to produce alternative nontaxonomic relations, even for abstract and emotional concepts, albeit through different strategies. Conversely, the age range in the middle, which corresponds to the beginning of middle school in the Italian school system, may be a period in which children are facing the beginning of a new school cycle, with new classes and fellow students. We speculate that this transition may encourage children to construct their own identity, relying on their own inner intuitions, also linguistically, rather than conforming to school-imposed knowledge and taxonomies. It is also plausible that 10–12-year-old children, while expanding their vocabulary, may not yet fully understand the different semantic relationships of new words. This partial understanding, combined with their confident use of these words, could lead to increased "noise" in their performance. This pattern differs from that of younger children, who rely on a smaller, well-understood vocabulary, and from older children, who possess a more mature and comprehensive understanding of language. This developmental trajectory is further confirmed by the separate analysis on hypernyms and hyponyms, with noise in hypernyms particularly higher among children aged 10–12. The variability in responses to hypernyms in this intermediate

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

age group likely reflects their ongoing process of consolidating broader category structures. In contrast, hyponyms showed much less variability across age groups, likely because they involve more specific categories, with a larger set of valid terms to choose from. This greater specificity likely makes it easier for children to identify correct words within a taxonomy, thereby minimizing the influence of extra-semantic relations and leading to more stable hyponyms production. Taken together, these findings suggest that the language-mediated abstraction process in children is shaped not only by the inherent structure of semantic categories but also by the dynamic interaction between cognitive development and linguistic experience, particularly during the critical transition from middle to late childhood. However, as social-cognitive factors related to the school transition were not directly measured in the present study, these results should be interpreted with caution and explored in future research.

Methodologically, our study has significant implications. Rather than relying on Likert-scale ratings to collect linguistic data, we employed a generative linguistic approach that prompts students to retrieve and produce both specific and general words from memory, thereby bolstering their linguistic proficiency. Productivity and validity of the word ladders improved across two data collection periods, with notable variations by semantic type and age group. Improvements were especially evident for concrete, emotional, and social-role concepts, indicating advancements in children's taxonomic knowledge and overall language-mediated abstraction. In contrast, abstract concepts showed less improvement, likely due to their inherently shorter taxonomic ladders. Notably, younger children demonstrated the most significant progress, presumably reflecting ongoing cognitive and linguistic development, whereas the oldest group (12–13 years) showed no notable improvement, possibly indicating the consolidation of their cognitive abilities and stabilized abstract language development. Nonetheless, ladder validity increased across all age groups and semantic types, evidencing overall progress in generating valid taxonomic structures. These improvements can be attributed to the "training effect" of the Word Ladders task, which fosters abstraction and language-mediated skills—an outcome often challenging to achieve with traditional methods. Cultivating a rich and fine-grained vocabulary in school children is a key objective, as it underpins a mature conceptual organization that is vital for effective communication, problem-solving, and diverse social interactions. Interventions such as the Word Ladders task can arguably enhance these advanced language skills, fostering lifelong cognitive flexibility and potentially improving educational outcomes.

Theoretically, our findings shed light on the organization of abstract concepts in mind and in language, showing that abstract concepts (and not only concrete ones) vary in specificity. Although taxonomies for abstract concepts tend to be shorter, researchers should account for this factor, particularly in studies comparing the processing of abstract and concrete words and concepts, where specificity is often overlooked (for a discussion, see Bolognesi et al., 2020).

There are some limitations to this study. First, the longitudinal design does not allow us to fully disentangle developmental changes from potential task-familiarity effects. Because the same participants completed the same Word Ladders task at two time points using the same target words, part of the improvement observed at T2 may reflect increased familiarity with the task format or greater awareness of the

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

taxonomic principle involved. Future studies could address this limitation through the inclusion of alternate task versions, transfer measures, or additional comparison tasks assessing related semantic and categorization abilities. Second, our sample does not include children under 9-years-old, a group in which abstract vocabulary undergoes a steep increase (Ponari et al., 2018). This limits our inferences about vocabulary growth and its impact on the abstraction mechanism. For instance, we hypothesize that preschoolers, with underdeveloped reading and writing skills, may demonstrate abstraction abilities limited to taxonomies of concrete concepts, with little to no improvement in constructing taxonomies of more abstract concepts. However, further research is needed to better understand these patterns, including standardized vocabulary assessments to capture children's baseline lexical knowledge. Third, our study focused on categorical inclusion relations and provided a quantitative analysis of non-IS-A relations, which primarily emerge in generalizations, particularly in the production of superordinate terms for abstract concepts. In future, we aim to conduct a detailed qualitative analysis of these semantic relations and test whether computational models can predict these relations, enabling the automatic classification of various semantic relations (e.g., free associations, synonyms, meronyms) within word ladders. Fourth, the present study concerns the use of WordNet and externally defined taxonomic ontologies as reference frameworks for the annotation of children's productions. Although these resources were adopted primarily to ensure methodological consistency in the identification of IS-A relations, they inevitably reflect adult-centered lexical and taxonomic conventions. As a consequence, the present analyses are better interpreted as assessing children's ability to approximate shared taxonomic norms within a school-based lexical task, rather than as a full reconstruction of children's spontaneous semantic organization. Future analyses focusing on the so-called "noise" relations may provide additional insight into alternative forms of conceptual organization that are less constrained by adult taxonomic structures, including figurative or metaphorical interpretation of target words that can emerge early in child language (e.g., Horvat et al., 2024). Finally, another limitation is the small number of words selected for balanced familiarity and time constraints, which prevents broad generalization of our findings across the entire semantic domain, as well as the lack of other socio-demographic factors (e.g., children's reading habits), which we are addressing by scaling up data collection via the Word Ladders mobile application to include a larger sample of Italian and English speakers with different cultural and educational backgrounds.

To conclude, language-mediated abstraction in late childhood is shaped more by experience than by age, revealing that abstract concepts form shorter and noisier semantic hierarchies, a theoretical insight that underscores the graded nature of abstraction and an applied indication that targeted linguistic activities can effectively support the development and consolidation of abstract knowledge.

Replication package. The study was not preregistered. The dataset generated in the current study and analysis scripts are available in the online repository the Open Science Framework (OSF): https://osf.io/a3beq/.

Acknowledgments. The authors would like to thank all members of the ABSTRACTION research group (GRANT AGREEMENT: ERC-2021-STG-101039777) for comments and discussions. They are/have been, in alphabetic order: Claudia Collacciani, Tommaso Lamarra, Laura Occhipinti, Giulia Rambelli, Andrea

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Amelio Ravelli, and Paola Vernillo. The authors wish to express deep gratitude to all the students and teachers of the IC7 public school in Imola (BO) for their participation in this study.

Author contributions.

- **Conceptualization:** Caterina Villani, Adele Loia, and Marianna M. Bolognesi;
- **Data curation:** Caterina Villani, Adele Loia, and Marianna M. Bolognesi;
- **Data collection:** Caterina Villani, Adele Loia, and Marianna M. Bolognesi;
- **Formal analysis:** Caterina Villani;
- **Funding acquisition:** Marianna M. Bolognesi;
- **Methodology:** Caterina Villani, Adele Loia, and Marianna M. Bolognesi;
- **Project administration:** Marianna M. Bolognesi;
- **Supervision:** Marianna M. Bolognesi;
- **Visualization:** Caterina Villani;
- **Writing—original draft:** Caterina Villani and Marianna M. Bolognesi;
- **Writing—review and editing:** Caterina Villani, Adele Loia, and Marianna M. Bolognesi.

Funding statement. Caterina Villani and Marianna M. Bolognesi were funded by the ABSTRACTION European Union (GRANT AGREEMENT: ERC-2021-STG-101039777). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. Adele Loia was founded by PNRR—PRIN 2022 (2022EPTPJ9) "WEMB: Word Embeddings from Cognitive Linguistics to Language Engineering and back."

Competing interests. The authors declare none.

Declaration of generative AI and AI-assisted technologies in the writing process. The authors declare that the generative AI tool (ChatGPT) was used only in the writing process to rephrase and improve the manuscript's readability.

Notes

1 Note that in the present study we refer to NON-IS-A relations with the term "noise," hereby used in a strictly operational and methodological sense. More specifically, it refers to the discrepancy between the raw Word Ladder produced by participants and the cleaned ladder obtained after retaining only IS-A taxonomic relations for the purposes of hierarchical lexical analysis. The label therefore does not imply that the excluded relations are cognitively irrelevant, erroneous, or semantically impoverished. On the contrary, many of these relations (e.g., meronymic, thematic, associative, frame-based, or experiential links) may reflect meaningful forms of conceptual organization and meaning construction, particularly in the case of abstract and emotional concepts. The term "noise" is thus intended only relative to the specific taxonomic criterion adopted in the present analyses and should not be interpreted as a broader evaluation of children's semantic competence or abstraction abilities.

2 At the second time of data collection only one of the two eighth-grade classes took part to the experimental session, due to some school organizational problems. The missing class attempted to provide the missing data under the sole supervision of the teacher, who autonomously ran the data collection session. Due to significant missing data (70%), and methodological issues, these datapoints were excluded from the analyses.

References

Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., & Mazzoleni, M. (2004). Introducing the La Repubblica Corpus: A Large, Annotated, TEI(XML)-compliant Corpus of Newspaper Italian. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04). European Language Resources Association (ELRA). https://www.lrec-conf.org/proceedings/lrec2004/pdf/247.pdf

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, **12**(1), 1–23. https://doi.org/10.1093/scan/nsw154

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48. https://doi.org/10.18637/jss.v067.i01

Bellagamba, F., Borghi, A. M., Mazzuca, C., Pecora, G., Ferrara, F., & Fogel, A. (2022). Abstractness emerges progressively over the second year of life. *Scientific Reports*, **12**(1), 20940. https://doi.org/10.1038/s41598-022-25426-5

Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., & Anna Maria Thornton. (2005). Corpus e Lessico di frequenza dell'Italiano Scritto (CoLFIS). https://www.iris.unisa.it/handle/11386/3818078

Bolognesi, M., Burgers, C., & Caselli, T. (2020). On abstraction: Decoupling conceptual concreteness and categorical specificity. *Cognitive Processing*, **21**(3), 365–381. https://doi.org/10.1007/s10339-020-00965-9

Bolognesi, M. M., & Caselli, T. (2023). Specificity ratings for Italian data. *Behavior Research Methods*, **55**(7), 3531–3548. https://doi.org/10.3758/s13428-022-01974-6

Borghi, A. M. (2023). *The freedom of words: Abstractness and the power of language*. Cambridge University Press. https://doi.org/10.1017/9781108913294

Borghi, A. M., & Binkofski, F. (2014). *Words as social tools: An embodied view on abstract concepts* (pp. xiii, 127). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-9539-0

Brown, R. W. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, **55**(1), 1–5. https://doi.org/10.1037/h0041199

Buccino, G., Colagè, I., Silipo, F., & D'Ambrosio, P. (2019). The concreteness of abstract language: An ancient issue and a new perspective. *Brain Structure and Function*, **224**(4), 1385–1401. https://doi.org/10.1007/s00429-019-01851-7

Catricalà, E., Conca, F., Fertonani, A., Miniussi, C., & Cappa, S. F. (2020). State-dependent TMS reveals the differential contribution of ATL and IPS to the representation of abstract concepts related to social and quantity knowledge. *Cortex*, **123**, 30–41. https://doi.org/10.1016/j.cortex.2019.09.018

Clark, E. V. (1993). *First language acquisition* (1. publ., 6. print). Cambridge University Press.

Cohen, H., & Lefebvre, C. (2005). *Handbook of categorization in cognitive science*. Elsevier.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**(1), 37–46. https://doi.org/10.1177/001316446002000104

Conca, F., Borsa, V. M., Cappa, S. F., & Catricalà, E. (2021). The multidimensionality of abstract concepts: A systematic review. *Neuroscience & Biobehavioral Reviews*, **127**, 474–491. https://doi.org/10.1016/j.neubiorev.2021.05.004

Crutch, S. J. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, **128**(3), 615–627. https://doi.org/10.1093/brain/awh349

Davis, C. P., Altmann, G. T. M., & Yee, E. (2020). Situational systematicity: A role for schema in understanding the differences between abstract and concrete concepts. *Cognitive Neuropsychology*, **37**(1–2), 142–153. https://doi.org/10.1080/02643294.2019.1710124

Della Rosa, P. A., Catricalà, E., Vigliocco, G., & Cappa, S. F. (2010). Beyond the abstract—concrete dichotomy: Mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behavior Research Methods*, **42**(4), 1042–1048. https://doi.org/10.3758/BRM.42.4.1042

Desai, R. H., Reilly, M., & van Dam, W. (2018). The multifaceted abstract brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **373**(1752), 20170122. https://doi.org/10.1098/rstb.2017.0122

Dreyer, F. R., & Pulvermüller, F. (2018). Abstract semantics in the motor system? – An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex*, **100**, 52–70. https://doi.org/10.1016/j.cortex.2017.10.021

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. MIT Press.

Ferrucci, F. (2022). Neology pathways: An analysis of holonymy/meronymy and hypernymy/hyponymy. *Cognitive Linguistic Studies*, **9**(2), 323–340. https://doi.org/10.1075/cogls.20014.fer

Foster-Hanson, E., & Rhodes, M. (2019). Normative social role concepts in early childhood. *Cognitive Science*, **43**(8), e12782. https://doi.org/10.1111/cogs.12782

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. SAGE.

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Gangemi, A., Guarino, N., Masolo, C., & Oltramari, A. (2010). Interfacing WordNet with DOLCE: Towards OntoWordNet (pp. 36–52). https://doi.org/10.1017/CBO9780511676536.004

Gangemi, A., Guarino, N., & Oltramari, A. (2001). Conceptual analysis of lexical taxonomies: The case of WordNet top-level. Proceedings of the International Conference on Formal Ontology in Information Systems - Volume 2001, 285–296. https://doi.org/10.1145/505168.505195

Gentner, D., & Asmuth, J. (2019). Metaphoric extension, relational categories, and abstraction. *Language, Cognition and Neuroscience*, **34**(10), 1298–1307. https://doi.org/10.1080/23273798.2017.1410560

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, **12**(4), 395–427. https://doi.org/10.3758/BF03201693

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, **1**(1), 23–64. https://doi.org/10.1207/s15473341lld0101_4

Guarino, N. (1998). Some Ontological Principles for Designing Upper Level Lexical Resources (arXiv:cmp-Lg/9809002). arXiv. https://doi.org/10.48550/arXiv.cmp-lg/9809002

Guarino, N., & Welty, C. (2000). A Formal Ontology of Properties. In R. Dieng & O. Corby (Eds.), *Knowledge Engineering and Knowledge Management Methods, Models, and Tools* (pp. 97–112). Springer. https://doi.org/10.1007/3-540-39967-4_8

Hajibayova, L. (2013). Basic-level categories: A review. *Journal of Information Science*, **39**(5), 676–687. https://doi.org/10.1177/0165551513481443

Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development*, **51**(3), 708–719. https://doi.org/10.2307/1129456

Horvat, A. W., Despot, K. Š., & Hržica, G. (2024). Early acquisition of figurative meanings in polysemous nouns and verbs. *Language and Cognition*, **16**(4), 2055–2080. https://doi.org/10.1017/langcog.2024.46

Katja Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science*, **29**(5), 719–736. https://doi.org/10.1207/s15516709cog0000_33

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, **140**(1), 14–34. https://doi.org/10.1037/a0021446

Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **12**(1), 92–107. https://doi.org/10.1037/0278-7393.12.1.92

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, **44**(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Lamarra, T., Villani, C., & Bolognesi, M. M. (2026). Specificity effect in concrete/abstract semantic categorization task. *Cognitive Processing*, **27**(1), 25–43. https://doi.org/10.1007/s10339-025-01286-5

Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Piaskowski, J., Riebl, H., & Singmann, H. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.10.4) [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html

Lewis, M., Colunga, E., & Lupyan, G. (2021). Superordinate Word Knowledge Predicts Longitudinal Vocabulary Growth. Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43). https://escholarship.org/uc/item/15w9t3vc

Lucariello, J., Kyratzis, A., & Nelson, K. (1992). Taxonomic knowledge: What kind and when? *Child Development*, **63**(4), 978–998. https://doi.org/10.2307/1131248

Lucariello, J., & Nelson, K. (1985). Slot-filler categories as memory organizers for young children. *Developmental Psychology*, **21**(2), 272–282. https://doi.org/10.1037/0012-1649.21.2.272

Lund, T. C., Sidhu, D. M., & Pexman, P. M. (2019). Sensitivity to emotion information in children's lexical processing. *Cognition*, **190**, 61–71. https://doi.org/10.1016/j.cognition.2019.04.017

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, **3**, 54. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G., & Lewis, M. (2019). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience*, **34**(10), 1319–1337. https://doi.org/10.1080/23273798.2017.1404114

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **373**(1752), 20170137. https://doi.org/10.1098/rstb.2017.0137

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

Markman, E. M. (1989). *Categorization and Naming in Children: Problems of Induction*. MIT Press.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, **14**(1), 57–77. https://doi.org/10.1207/s15516709cog1401_4

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, **16**(1), 1–27. https://doi.org/10.1016/0010-0285(84)90002-1

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development*, **53**(1), 258–266. https://doi.org/10.2307/1129660

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database*. *International Journal of Lexicography*, **3**(4), 235–244. https://doi.org/10.1093/ijl/3.4.235

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the affective norms for english words (ANEW) for Italian. *Behavior Research Methods*, **46**(3), 887–903. https://doi.org/10.3758/s13428-013-0405-3

Muraki, E. J., Reggin, L. D., Feddema, C. Y., & Pexman, P. M. (2025). The development of abstract word meanings. *Journal of Child Language*, **52**(1), 195–207. https://doi.org/10.1017/S0305000923000569

Murphy, G. L. (2002). *The big book of concepts*. MIT Press.

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin*, **84**(1), 93–116.

Nelson, K. (1988). Where do taxonomic categories come from? *Human Development*, **31**(1), 3–10. https://doi.org/10.1159/000273198

Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract Nouns. *Frontiers in Human Neuroscience*, 8, 6:275. https://doi.org/10.3389/fnhum.2012.00275

Nook, E. C., Stavish, C. M., Sasse, S. F., Lambert, H. K., Mair, P., McLaughlin, K. A., & Somerville, L. H. (2020). Charting the development of emotion comprehension and abstraction from childhood to adulthood using observer-rated and linguistic measures. *Emotion (Washington, D.C.)*, **20**(5), 773–792. https://doi.org/10.1037/emo0000609

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, **45**(3), 255–287. https://doi.org/10.1037/h0084295

Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, **21**(2), e12549. https://doi.org/10.1111/desc.12549

Ravelli, A. A., Bolognesi, M. M., & Caselli, T. (2024). Specificity ratings for English data. *Cognitive Processing*, **26**(2), 283–302. https://doi.org/10.1007/s10339-024-01239-4

RCoreTeam. (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Reilly, J., Shain, C., Borghesani, V., Kuhnke, P., Vigliocco, G., Peelle, J. E., Mahon, B. Z., Buxbaum, L. J., Majid, A., Brysbaert, M., Borghi, A. M., De Deyne, S., Dove, G., Papeo, L., Pexman, P. M., Poeppel, D., Lupyan, G., Boggio, P., Hickok, G., … **Vinson, D.** (2025). What we mean when we say semantic: Toward a multidisciplinary semantic glossary. *Psychonomic Bulletin & Review*, 32(1), 243–280. https://doi.org/10.3758/s13423-024-02556-7

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, **4**(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Schwanenflugel, P. (1991). Why are abstract concepts hard to understand? In P. Schwanenflugel (Ed.) *The Psychology of Word Meanings* (pp. 223–250). Hillsdale, NJ: Erlbaum.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, **27**(5), 499–520. https://doi.org/10.1016/0749-596X(88)90022-8

Suffill, E., Schonberg, C., Vlach, H. A., & Lupyan, G. (2022). Children's knowledge of superordinate words predicts subsequent inductive reasoning. *Journal of Experimental Child Psychology*, **221**, 105449. https://doi.org/10.1016/j.jecp.2022.105449

Troyer, M., & McRae, K. (2022). Thematic and other semantic relations central to abstract (and concrete) concepts. *Psychological Research*, **86**(8), 2399–2416. https://doi.org/10.1007/s00426-021-01484-8

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press

VandenBos, G. R. (Ed.). (2015). *APA dictionary of psychology* (2nd ed.). American Psychological Association. https://doi.org/10.1037/14646-000

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, **1**(2), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Vigliocco, G., Ponari, M., & Norbury, C. (2018). Learning and processing abstract words and concepts: Insights from typical and atypical development. *Topics in Cognitive Science*, **10**(3), 533–549. https://doi.org/10.1111/tops.12347

Villani, C., & Bolognesi, M. M. (2026). Specificity: Metrics and norms. In H. Nesi & P. Milin (Eds.), *International Encyclopedia of Language and Linguistics* (Third Edition) (pp. 678–683). Elsevier. https://doi.org/10.1016/B978-0-323-95504-1.00894-2

Villani, C., Loia, A., & Bolognesi, M. M. (2024). The semantic content of concrete, abstract, specific, and generic concepts. *Language and Cognition*, **16**(4), 867–894. https://doi.org/10.1017/langcog.2023.64

Villani, C., Lugli, L., Liuzza, M. T., & Borghi, A. M. (2019). Varieties of abstract concepts and their multiple dimensions. *Language and Cognition*, **11**(3), 403–430. https://doi.org/10.1017/langcog.2019.23

Villani, C., Lugli, L., Liuzza, M. T., Nicoletti, R., & Borghi, A. M. (2021). Sensorimotor and interoceptive dimensions in concrete and abstract concepts. *Journal of Memory and Language*, **116**, 104173. https://doi.org/10.1016/j.jml.2020.104173

Vossen, P. (Ed.). (1998). *EuroWordNet: A multilingual database with lexical semantic networks*. Springer Netherlands. https://doi.org/10.1007/978-94-017-1491-4

Waxman, S., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development*, **1**(2), 139–156. https://doi.org/10.1016/S0885-2014(86)80016-8

Winter, B. (2022). Abstract concepts and emotion: Cross-linguistic evidence and arguments against affective embodiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **378**(1870), 20210368. https://doi.org/10.1098/rstb.2021.0368

Cite this article: Villani, C., Loia, A., & Bolognesi, M. M. (2026). Development of language-mediated abstraction: Insights from word ladders task. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716426100897

https://doi.org/10.1017/S0142716426100897 Published online by Cambridge University Press