Learning Abstract Words and Concepts:

Insights from Developmental Language Disorder

Marta Ponari[a], Courtenay Frazier Norbury[b], Armand Rotaru[c], Alessandro Lenci[d], Gabriella Vigliocco[c*]

a. School of Psychology, University of Kent, UK.

b. Language and Cognition, Psychology and Language Sciences, University College London

c. Institute for Multimodal Communication, Psychology and Language Sciences, University College London, 26 Bedford Way, London WC1H 0AP, UK.

d. Computational Linguistics Laboratory, Department of Philology, Literature and Linguistics, University of Pisa, Italy.

*Corresponding author

Abstract

Some explanations of abstract word learning suggest that these words are learnt primarily from linguistic input, using statistical co-occurrences of words in language whereas concrete words can also rely on non-linguistic, experiential information. According to this hypothesis, we expect that, if the learner is not able to fully exploit the information in the linguistic input, abstract words should be affected more than concrete ones. Embodied approaches, instead, argue that both abstract and concrete words can rely on experiential information and, therefore, there might not be any linguistic primacy. Here, we test the role of linguistic input in the development of abstract knowledge with children with Developmental Language Disorder (DLD) and Typically Developing (TD) children aged 8-13. We show that DLD children, who by definition have impoverished language, do not show a disproportionate impairment for abstract words in lexical decision and definition tasks. These results indicate that linguistic information does not have a primary role in the learning of abstract concepts and words, rather, it would play a significant role in semantic development across all domains of knowledge.

Keywords: developmental language disorder; language acquisition; abstract concepts; distributional semantics; semantic representation; vocabulary development

Introduction

Learning the meaning of words is one of the most complex and remarkable of human achievements. Learning words is hard because even when the referent is present in the physical environment, rarely is it isolated in the visual scene (1). To make the situation worse, referents are not always present in the physical environment, either because they are spatially and/or temporally displaced (e.g., talk about past or future events), or because they are abstract and have no tangible referent.

A number of theories argue that abstract concepts are grounded (solely or primarily) in our linguistic experience (2–5) whereas concrete words could benefit also from non-linguistic information. For example, it has been shown that the richness of featural representations (used as a proxy of sensory-motor and affective content) predicts behavioural effects (e.g., lexical decision, semantic priming) better for concrete than abstract words; whereas, the richness of the linguistic contexts in which a word appears (semantic neighbourhood density, used as a proxy for language-based information) predicts behavioural effects better for abstract than concrete words (6).

Embodied theories of semantic representation, instead, argue that learning and representing both concrete and abstract concepts are grounded in our experience with the world. There is now plenty of evidence that processing concrete concepts in adults engages to some extent the same cognitive and neural systems involved in perceiving and acting upon the physical world (7). There is also growing evidence that processing abstract concepts in adults involves motor representations (8,9), simulation of specific situations (10) and the emotion system (11,12). In development, Ponari et al. (13) showed that abstract words with emotional connotations are learnt earlier than neutral abstract words suggesting that emotion could serve as a bootstrapping mechanism for the learning of abstract words and concepts. Scholars who argue for a role of embodied information in the learning and representation of abstract concepts also assume that linguistic information matters, but do not claim “language primacy” (8,14,15).

Here, we present a test of the role of linguistic information in learning semantic representations for abstract words comparing the knowledge of abstract and concrete words by children with Developmental Language Disorder (DLD) and their typically developing peers (TD).

DLD is a neurodevelopmental disorder affecting approximately 7.5% of children at school entry (16). Children with DLD typically present with severe deficits in morphosyntax and other aspects of grammar (17) as well as vocabulary that is reduced in both breadth and depth relative to typically-developing peers (18). Vocabulary reduction in children with DLD has been linked to a number of different causes among which are working memory deficits (19), statistical learning (20) and attention (21). However, no previous study to our knowledge has focused on abstract words, despite the anecdotal report by Speech and Language Professionals that these children are especially impaired with these words. Here, we investigate knowledge of abstract and concrete word meanings in children with DLD and typically developing (TD) peers matched for chronological age (TD_{age}) or receptive vocabulary scores (TD_{voc}). As DLD is assumed to affect vocabulary development (18), it follows that, if learning abstract words is based primarily on linguistic information, then abstract words should be disproportionately impaired relative to concrete words in children with DLD when compared to their TD peers. The inclusion of both age- and vocabulary-matched control groups allows us to assess both quantitative as well as qualitative differences in knowledge of words: the comparison with age-matched TD children can tell us whether DLD children show any quantitative difference with their peers. Thus, if DLD children show larger impairment for abstract than concrete words, this could be either because these words are learnt later by DLD children, or because there are qualitative differences in the manner in which DLD and TD children learn vocabulary. The comparison with younger vocabulary-matched TD children will then allow us to make inferences about whether any difference we find in the DLD-TD_{age} comparison depends on qualitative differences in the way DLD children use and organise their word knowledge, or whether DLD children are simply behind in their linguistic development.

We chose to use both definitions and lexical decision tasks: defining words, provides a direct window into what children know about concepts; it is, however, a challenging task as it further requires expressive language, which is often compromised in children with DLD. Thus, the definition task may underestimate word knowledge in this group. Lexical decision does not require language production although it provides a more indirect window into children's knowledge of word.

Methods

Participants

Eighteen children with DLD (14 males; mean age = 10.03, SD = 1.76) were recruited from schools in Southeast England. All children had a clinical diagnosis from a speech-language therapist external to the research team. Children in the control groups were selected from a pool of 73 TD children who completed both tasks: 18 children (14 males; mean age = 10.34, SD = 1.44) were matched to the DLD children on gender and age (TD_{age}), and 18 (14 males; mean age = 8.16, SD = 2.12) were matched to the DLD children on gender and raw scores on the *British Picture Vocabulary Scale* (BPVS, (22); TD_{voc}). TD children were recruited from local schools and did not have any reported special educational needs, or history of language delay. Non-verbal cognitive abilities were assessed using the Matrix Reasoning test of the Wechsler Abbreviated Scale of Intelligence (WASI; (23)). DLD children were also administered the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals: Core Language Scales (CELF; (24)). Children characteristics are summarised in Table 1. The protocol was approved by the Research Ethics Committee at University College London; informed, written consent was obtained from all parents and verbal assent was obtained from all children prior to assessment. The same children participated in both tasks.

Table 1

Materials

Thirty-six abstract and 36 concrete words were selected from a pool of 3,505 words for which normative data on a range of lexical variables could be obtained. These variables included: age of acquisition (AoA; (25)), concreteness (26), valence (27), and log-frequency (28). AoA ratings were used to ensure the items selected were appropriate for our participants' age: words were divided into age of acquisition bands (1: words acquired at 4-5 years; 2: 6-7 years; 3: 8-9 years; 4: 10-11 years). Within each AoA band, concrete and abstract words were matched for valence, length (number of letters) and log-frequency. Concrete and abstract words also did not differ on familiarity, and on a measure of frequency taken from subtitles from a UK TV channel targeted at children aged 6-12 (CBBC; (29)). Lexical and sublexical characteristics of the words are listed in Table 2, see Supplementary Materials for a list of all words and the non-words used in the lexical decision task.

Among these 72 words, 24 (12 abstract and 12 concrete) were shared between the two tasks; 24 (12 abstract and 12 concrete) were used for the definitions task only, and the remaining 24 were used for the auditory lexical decision task only. Additionally, for the lexical decision task, forty-eight pronounceable non-words were created by changing one phoneme from 48 words matched to the experimental words on length, AoA, valence and concreteness. All words and non-words were recorded by a native English speaker using Audacity v. 1.2.2 (30).

Table 2

Procedure

All children were assessed in their school and received stickers for participation. Stimuli were presented verbally using E-Prime v. 2.0 (31) running on a laptop with a touchscreen display. Participants were presented with short computer games in which they were asked to help a cartoon alien learn English. The Lexical Decision task was always presented before the Definition task, in a single session. Children received verbal instructions from the experimenter, and were asked to wear headphones prior to the beginning of each task.

Lexical decision. In each trial, a cartoon alien was presented in the middle of the screen for 1000 ms, followed by the auditory presentation of either a real English word or a non-word. Immediately after the offset of the word (average stimulus duration = 830 ms), two touch screen buttons appeared at the bottom left (a red thumbs-down icon) or the bottom right (a green thumbs-up icon) of the screen and children were asked to indicate whether what they heard was a word they knew (green button), or a “funny, made-up” word (red button). Six practice trials (three non-words and three words not used in the experiment) included visual feedback of either a smiling (correct trial) or frowning (incorrect) cartoon alien after each response. No feedback was provided for the remaining 96 trials (24 abstract and 24 concrete words, plus 48 non-words), which were presented in a randomised order. Presentation of each subsequent word was prompted by the experimenter to ensure the child was on-task. To minimise fatigue, children were given the choice to take a break every 24 trials. Accuracy and reaction times were recorded; however, to ensure child attention and compliance to task instructions, the experimenter controlled stimulus presentation and did not ask the children to respond quickly, but rather as accurately as possible. Therefore, only accuracy data is analysed below. Note that this does not limit our ability to observe semantic effects, as we have shown in a previous study using the same materials and procedure (13).

Definition. Children were encouraged to provide an accurate and comprehensive definition, including as much information as they could on the meaning of each word. Each trial included the presentation of the alien in the center of the computer screen, along with the acoustic presentation of a word. Children's responses were audio-recorded and then scored off-line but "don't know" or definitely inaccurate responses were recorded online by the experimenter. The presentation of subsequent words was prompted by the experimenter. The 48 words were presented in four blocks of 12 items arranged in blocks corresponding to the AoA bands described previously. Words within each block were presented in random order. The task ended when the child was unable to define three words within a single block or responded to all 48 words.

Definitions were transcribed off-line and scored according to the following criteria:

a) *Definitions accuracy*. Definitions were scored according to the Wechsler Intelligence Scale for Children (WISC) vocabulary sub-test scoring criteria (32).Scoring was performed by two independent researchers who were blind to the study hypotheses and diagnosis of the children. A third independent researcher moderated instances in which only one scorer awarded a score of 0; all other scores were averaged.

b) *Definitions' quality ratings*. All definitions that were scored > 0 following the above criteria (N = 959) were arranged in lists of about 200 and presented to a minimum of N = 10 (range = 10-13) adult native English speakers, which were recruited on the crowdsourcing website Prolific (https://www.prolific.ac/). Participants were asked to rate how accurate each definition was in defining the concept. The procedure and the instructions given to raters are detailed in Supplementary Materials. These ratings allow us to assess at a more fine-grained level the extent to which definitions of abstract words by DLD children may be of lesser quality than those by TD children.

c) *Definitions' conceptual features*. Definitions were scored based on the 11 conceptual categories used by Barca, Mazzuca and Borghi (33). This classification allows us to have some initial insight on the conceptual features of concepts known by DLD and TD children differ. The procedure and results of this analysis are reported in Supplementary Materials.

Data analysis

DLD children were contrasted to: 1) a group of TD children matched on age (TD_{age}), to see whether DLD children had lower scores than they TD peers, especially for abstract words; 2) a group of (younger) TD children matched on vocabulary (TD_{voc}) to further assess qualitative differences in their knowledge of concrete and abstract words. Quantitative data was analysed using mixed-effect models running in R version 3.2.1, running in R version 3.2.1 (34). Lexical decision accuracy was analysed using mixed effects logistic regression models (GLME; package 'lmerTest' (35)); definition scores were treated as ordinal and analysed using cumulative link mixed models (CLMM; package ‘ordinal’ (36)); and average definition quality ratings were analysed using linear mixed models (LME; package ‘lmerTest’ (35)). In all analyses, the baseline models included as continuous predictor the children’s non-verbal reasoning scores, which significantly differ between our DLD and TD groups, and our categorical variables of interest: concreteness (abstract, concrete) and group (DLD vs TD_{age}; DLD vs TD_{voc}), as well as the two-way interaction between the two. The categorical variables were contrast coded and the continuous predictor was centered on the mean. Log-likelihood ratio tests were used to compare fitted models. Supplementing these analyses, we performed Bayesian mixed-effects model analysis using the ‘brms’ package (37) for R, which fits Bayesian multilevel models using the Stan programming language. Model fit was performed using default priors, running 4 chains of 10000 iterations each. We compared models pairwise by computing the Bayes Factor (BF10). A BF10 of 3 is considered sufficient evidence to favour a model over another (38,39), while a BF10 between 1/3 and 3 indicates that there is not enough evidence in the data to provide support for either model, and a BF10 < 1/3 indicates definite evidence against the model and in favour of the null hypotheses. Bayes Factors are reported in Table 3.

Qualitative data analysis of the Definitions’ conceptual features was carried out using Correspondence Analysis (CA; (40,41)) running in R version 3.2.1 (package ‘CAinterprTools’; (42)), and it is reported in Supplementary Materials.

Finally, for both lexical decision and definition tasks, case-series analyses was performed using the Revised Standardized Difference Test (RSDT; (43)), in which the difference in performance between concrete and abstract words per each DLD child is compared to the difference in performance exhibited by the TD groups (either TD_{age} or TD_{voc}); this is reported in Supplementary Materials.

Results

Below, we report the results of model selection in mixed-effects models. p-values from the model comparisons and corresponding Bayes Factors are summarised in Table 3.

Lexical decision

One DLD child did not complete the task. Data from the remaining 17 children and matched controls were inspected to check whether any children showed a bias toward either answering “word” or “non-word”. We computed the response bias (or *criterion, c*), by multiplying the sum of the normalised *hit* rate (correctly identifying a word) and the normalised *false alarm* rate (incorrectly claiming that a non-word was a word) by -0.5 (44–46). A criterion with a negative value would indicate that responses are biased toward answering “word” (both words and non-words are more likely to be indicated as words); a criterion of positive value would, conversely, indicate a response bias toward answering “non-word” (both words and non-words are more likely to be indicated as non-words). The average criterion bias was -0.002 (SD = 0.33) for TD children, and -0.02 (SD = 0.50) for DLD children. Children who showed a criterion bias higher than 1.5 standard deviations above their group mean (indicating a strong bias toward “non-word” responses) or lower than 1.5 standard deviations below their group mean (indicating a strong bias toward “word” responses) were excluded from further analyses. 3 children were therefore excluded from the DLD group (DLD9: *c* = -0.97; DLD12: *c* = -0.74; DLD17: *c* = -0.97); to maintain the matching between the DLD and TD groups, we also excluded the corresponding TD children.

DLD vs TD_{age}

Proportion of correct responses of the two groups for concrete and abstract words is shown in Figure 1 (left). We started by comparing the baseline model including the interaction between concreteness and group (see details above) against a model that included the main effects only. Including the two-way interaction did not significantly improve the fit of the model (log-likelihood ratio for interaction model = -542.3; log-likelihood ratio for main effects model = -542.4; $\chi^2(1)$ = 0.071, p = .790).

In the main effects model, non-verbal abilities (coefficient estimate = 0.008, SE = 0.002, p = .007) was a significant predictor of children's performance and it was therefore kept in subsequent models. We then tested whether the main effects were significant by removing them from the model, one by one. Removing the main effect of group significantly reduced the fit (log-likelihood ratio for the model including the main effect of group = -542.4; log-likelihood ratio for the model not including it = -544.5; $\chi^2(1)$ = 4.335, p = .037), with TD_{age} children recognising more words overall compared to DLD children (coefficient estimate = -0.69, SE = 0.32). Removing the main effect of concreteness did not affect the fit (log-likelihood ratio for the model including the main effect of concreteness = -542.4; log-likelihood ratio for the model not including it = -542.5; $\chi^2(1)$ = 0.243, p = .622).

DLD vs TD_{voc}

Two TD children did not complete the task due to time constraints; therefore, they were excluded along with their matched DLD peer; this left 12 children per group. The proportion of correct responses is shown on Figure 1 (right). The interaction between concreteness and group was not warranted (log-likelihood ratio for interaction model = -330.43; log-likelihood ratio for model not including it = -330.44; χ2(1) = 0.028, p = .866). There was no significant main effect of concreteness (log-likelihood ratio for model including the main effect of concreteness = -330.44; log-likelihood ratio for model not including it = -330.69; $\chi2(1) = 0.497$, $p = .481$), and no main effect of group (log-likelihood ratio for model including the main effect of group = -330.69; log-likelihood ratio for model not including it = -330.85; $\chi2(1) = 0.323$, $p = .570$).

Figure 1

Definition

Only 13.4% of our TD children could provide any definition for words of AoA block 4 (words acquired at 10-11); therefore, we excluded block 4 from further analysis, reducing the total number of items to 36 words (18 abstract and 18 concrete). Overall, definitions provided by DLD children were significantly shorter (M = 7.21 words, SD = 4.03) than both definitions provided by TD_{age} (M = 9.04 words, SD = 7.19; $p < .001$) and TD_{voc} children (M = 10.02 words, SD = 8.76; $p < .001$), plausibly reflecting the expressive difficulties of DLD children.

a) Definition score.

DLD vs TD_{age}

Definition accuracy (raw total score) for concrete and abstract words is depicted in Figure 2 (left).

Including the two-way interaction did significantly improve the fit of the model (log-likelihood ratio for interaction model = -1191.4; log-likelihood ratio for main effects model = -1194.6; LRtest = 6.455, $p = .011$). In this model, non-verbal abilities (coefficient estimate = 0.15, SE = 0.02, $p < .001$) was a significant predictor of children's performance, so it was kept in subsequent models.

To interpret the significant interaction, we first looked at the main effect of concreteness separately in the two groups. We found no difference between definition scores for abstract and concrete words in both TD_{age} children (coefficient estimate = -0.45, SE = 0.64, p = .485) and DLD children (coefficient estimate = -1.02, SE = 0.67, p = .132). Looking separately at concrete and abstract words, TD_{age} children's performance was significantly better than DLD children for both concrete words (coefficient estimate = -1.04, SE = 0.33, p = .002) and abstract words (coefficient estimate = -1.57, SE = 0.36, p < .001).

DLD vs TD_{voc}

One TD child did not complete the task and his definitions were excluded along with data from the matched DLD child. Definition accuracy (raw total score) is illustrated in Figure 2 (right). The interaction between concreteness and group was not warranted (log-likelihood ratio for interaction model = -1064.4; log-likelihood ratio for model not including it = -1064.5; LRtest = 0.346, p = .556). The main effect of group was significant (log-likelihood ratio for model including the main effect = -1065.5; log-likelihood ratio for model not including it = -1068.9; LRtest = 6705, p = .010). There was no main effect of concreteness (log-likelihood ratio for model including the main effect = -1064.5; log-likelihood ratio for model not including it = -1065.5; LRtest = 1.993, p = .158).

Figure 2

b. Definitions' quality ratings.

DLD vs TD_{age}.

In the online study, we obtained ratings for 247 definitions provided by DLD children, and 439 definitions provided by their TD_{age} peers.

The concreteness × group interaction was not warranted (log-likelihood ratio for interaction model = -1002.3; log-likelihood ratio for main effects model = -1003.1; $\chi^2(1) = 1.583$, p = .208). The main effect of group was significant (log-likelihood ratio for the model including the main effect of group = -1003.1; log-likelihood ratio for the model not including it = -1006.8; $\chi^2(1) = 7.399$, p = .007), with definitions of TD_{age} children rated as more accurate overall than those provided by their DLD peers. The main effect of concreteness was also significant (log-likelihood ratio for the model including the main effect of concreteness = -1003.1; log-likelihood ratio for the model not including it = -1001.1; $\chi^2(1) = 3.963$, p = .047). Crucially, definitions of abstract words were rated as more accurate than definitions of concrete words, for both DLD and TD_{age} children (see Figure 3, left).

DLD vs TD_{voc}.

We analysed ratings for 244 definitions provided by DLD children, and 314 definitions provided by their TD_{voc} peers.

The concreteness × group interaction was not warranted (log-likelihood ratio for interaction model = -810.4; log-likelihood ratio for main effects model = -811.0; $\chi^2(1) = 1.123$, p = .268). The main effect of group was significant (log-likelihood ratio for the model including the main effect of group = -810.4; log-likelihood ratio for the model not including it = -814.3; $\chi^2(1) = 6.667$, p = .010), with definitions of TD_{voc} children rated as more accurate than those provided by their DLD peers. The main effect of concreteness was marginally significant (log-likelihood ratio for the model including the main effect of concreteness = -810.4; log-likelihood ratio for the model not including it = -812.9; $\chi^2(1) = 3.817$, p = .051), with definitions of abstract words rated as more accurate than definitions of concrete words (see Figure 3, right).

Figure 3

Table 3

Discussion

This study aimed to assess whether linguistic development has a greater role – a primacy – in the learning of abstract compared to concrete concepts as predicted by theories such as Dual Coding (3) and Context Availability (4). We tested knowledge of abstract and concrete words in children with DLD and age-matched as well as (younger) vocabulary-matched peers, using both a lexical decision and a definition task.

In the lexical decision task, we found that children with DLD recognised significantly less words overall compared to their age-matched TD peers, however this was a small effect and not confirmed by the Bayes Factor analysis. What is of most interest here, however, is that while DLD children's performance was impaired with all words, they did not show a disproportionate impairment with abstract words compared to concrete, as confirmed by the lack of a concreteness by group interaction, supported by a Bayes Factor in favour of the null hypothesis. Interestingly, when looking at the comparison between DLD children and their vocabulary-matched peers, we found no significant differences at all. The lexical decision task however only gives us an indication of how many words children could recognise, and it cannot tell us anything about children's appreciation of word meaning.

In the definition task, when looking at definition accuracy, we do find a significant interaction between concreteness and group. What the results of the definition task suggest is that TD_{age} children give more accurate definitions compared to DLD children for both abstract and concrete words, and they define abstract words with similar accuracy compared to concrete words (although Bayes factor analysis suggests there is no enough evidence to accept the null hypothesis), while children with DLD show a larger difference between accuracy for abstract and concrete definitions. When compared to younger TD_{voc} children, children with DLD are worse at defining all words, not only abstract words. The additional analyses reported in Supplementary Materials, which contrast the difference in performance between abstract and concrete words for each individual DLD children against the average difference shown by TD_{age} and TD_{voc} children, confirm that any difference in definition scores for abstract and concrete words is equivalent to that exhibited by the TD groups.

When we look at the quality ratings that adults provided of how accurate each definition is in defining the concept, we find again strong support for a difference between the quality of definitions provided by TD children (both TD_{age} and TD_{voc}) and children with DLD, but the lack of an interaction between group and concreteness suggests that the difference, if any, between DLD children's definitions of abstract and concrete concepts is not significantly larger than the difference, if any, exhibited by TD children. Interestingly, although the Bayes Factor suggests the evidence to argue for or against a main effect of concreteness is inconclusive, the marginally significant p-values suggest that adults rate definitions of abstract words (overall) as slightly more accurate than definitions of concrete words. This is an unexpected but interesting result which might be linked to task expectations. Adult raters were recruited over the internet and they knew that the definitions were provided by children. It may be that in general they were less strict for the abstract words as these are typically considered to be harder for children.

To summarise, DLD children show impaired performance at recognising both abstract and concrete words compared to their age-matched peers, but they can correctly recognise an equivalent number of words (both abstract and concrete) compared to their vocabulary-matched peers (as showed by the lexical decision task). However, they cannot provide the same level of quality of definitions. It is worth noting here that TD_{voc} children were matched to our DLD children on receptive vocabulary scores, but the definitions task requires expressive language skills, which are impaired in children with DLD (as also supported by the fact that their definitions were shorter than those of TD children).

Taken together, our lexical decision results as well as the results from analysis of definition accuracy, definition quality ratings and the comparisons of individual DLD with the TD groups, do not provide clear support for linguistic primacy in the learning of abstract words and concepts. When language development is impaired, as is the case for DLD children, both knowledge of abstract and concrete words is impaired. When expressive vocabulary is not required, children with DLD perform like younger TD children with equivalent receptive vocabulary. This suggests that the same factors might support learning of new words in young children and children with DLD.

A number of theoretical accounts assume that embodied information contributes to the semantic representation of words. For example, Kousta et al. (11) suggested that while words referring to concrete objects and actions would be learnt by associating sensory-motor experience with the word, abstract words would be learnt by associating emotional states with the word. Ponari et al. (13) showed that TD children up to the age of 8-9 (about the age range of our TD_{voc} children) have better knowledge of emotionally valenced abstract words. They suggest that emotion might be particularly important for the acquisition of abstract words early in childhood, when vocabulary is mainly acquired through social interactions, providing a bootstrapping mechanism. Emotional valence could support children in discovering that some words – those that trigger emotional reactions – refer to internal states, rather than to objects and actions in the environment, thus, providing the building blocks for establishing the general category of abstract concepts. Later on, after the age of 9, the effect of valence declines. They suggests that as vocabulary and linguistic competence increases, children make greater use of linguistic information (e.g., from text), and are more able to make use of correlational patterns in discourse in order to extrapolate abstract meaning from the linguistic context (13). Children with DLD have reduced vocabulary and deficits in syntactic competence, and it has been shown that they are not as attuned as TD peers to statistical co-occurrences in language input (48). However, they do not have sensory-motor, or emotional/social impairments. Thus, they can benefit of the same embodied mechanisms for learning both concrete and abstract words as their TD peers.

The qualitative analysis of the content of the definitions provided by DLD as well as TD children (both TD_{age} and TD_{voc}) reported in Supplementary Materials supports the idea that sensory-motor associations are crucial for concrete words while affective associations are crucial for abstract words (11). Here, we found that definitions of concrete concepts include more perceptual features of the referent, their spatial location or function, as well as superordinate levels of the taxonomy, while abstract concepts' definitions include more situational and emotional features. These different features provide a clear distinction between abstract and concrete categories, at least in TD children. According to the same analysis however, definitions of abstract and concrete words in DLD children are less clearly distinct, and it seems less clear whether children with DLD make use of embodied emotional and situational features when defining abstract concepts. In summary, while children with DLD do not seem to be more impaired with abstract vs concrete words compared to their TD peers in terms of how accurate their definitions are, DLD children might not use embodied (sensory-motor, emotional and situational) information to the same extent as their TD peers. However, we can only speculate on the basis of the current data as the differences might just reflect expressive deficits of the DLD children.

In conclusion, the study presented the first investigation of abstract word knowledge by children with DLD. Our results confirm the role of linguistic information on the representation of concepts across domains of knowledge. Children with DLD show poorer vocabulary when compared to age-matched TD children. We do not support, however, a special role for linguistic information in the learning of abstract concepts. It is for future studies to further investigate to what extent children with DLD can take advantage of sensory-motor and emotional information in learning the semantics of both concrete and abstract words.

Acknowledgment

This research was supported by a grant from Nuffield Foundation (EDU/40477) to GV and CFN, but the views expressed are those of the authors and not necessarily those of the Foundation.

References

1. Medina TN, Snedeker J, Trueswell JC, Gleitman LR. How words can and cannot be learned by observation. Proc Natl Acad Sci [Internet]. 2011 May 16 [cited 2016 Mar 30];108(22):9014–9. Available from: http://www.pnas.org/content/108/22/9014.full
2. Binder JR, Desai RH. The neurobiology of semantic memory. Trends Cogn Sci [Internet]. Elsevier Ltd; 2011 Nov [cited 2014 Jul 21];15(11):527–36. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3350748&tool=pmcentrez&rendertype=abstract
3. Paivio A. Mind and its evolution: A dual coding theoretical approach. [Internet]. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2007 [cited 2016 Jan 7]. 517 p. Available from: http://psycnet.apa.org/psycinfo/2006-22587-000
4. Schwanenflugel PJ. Why are abstract concepts hard to understand? In: Schwanenflugel PJ, editor. The psychology of word meanings. Hillsdale, NJ, England: Lawrence Erlbaum Associates; 1991. p. 223–50.
5. Hoffman P, Binney RJ, Lambon Ralph MA. Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge. Cortex [Internet]. Elsevier; 2015 Feb [cited 2018 Jan 22];63:250–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25303272
6. Recchia G, Jones MN. The semantic richness of abstract concepts. Front Hum Neurosci [Internet]. Frontiers; 2012 Nov 27 [cited 2018 Apr 26];6:315. Available from: http://journal.frontiersin.org/article/10.3389/fnhum.2012.00315/abstract
7. Meteyard L, Cuadrado SR, Bahrami B, Vigliocco G. Coming of age: a review of embodiment and the neuroscience of semantics. Lotte Meteyard, Sara Rodriguez Cuadrado, Bahador Bahrami, Gabriella Vigliocco. :1–57.

8. Borghi AM, Binkofski F, Castelfranchi C, Cimatti F, Scorolli C, Tummolini L. The challenge of abstract concepts. Psychol Bull [Internet]. 2017 Mar [cited 2018 Apr 26];143(3):263–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28095000

9. Moseley R, Carota F, Hauk O, Mohr B, Pulvermüller F. A role for the motor system in binding abstract emotional meaning. Cereb Cortex [Internet]. 2012 Jul [cited 2013 Mar 2];22(7):1634–47. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3377965&tool=pmcentrez&rendertype=abstract

10. Barsalou LW, Wiemer-Hastings K. Situating abstract concepts. Grounding cognition: The orle of perception and action in memory, language and thought. 2005. p. 129–63.

11. Kousta S-T, Vigliocco G, Vinson DP, Andrews M, Del Campo E. The representation of abstract words: why emotion matters. J Exp Psychol Gen [Internet]. 2011 Feb [cited 2012 Jul 23];140(1):14–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21171803

12. Vigliocco G, Kousta S-T, Della Rosa PA, Vinson DP, Tettamanti M, Devlin JT, et al. The neural representation of abstract words: the role of emotion. Cereb Cortex [Internet]. 2014 Jul 13 [cited 2016 Jan 7];24(7):1767–77. Available from: http://cercor.oxfordjournals.org/content/early/2013/02/13/cercor.bht025

13. Ponari M, Norbury C, Vigliocco G. Acquisition of abstract concepts is influenced by emotional valence. Dev Sci [Internet]. 2017; Available from: http://onlinelibrary.wiley.com/doi/10.1111/desc.12549/full

14. Dove G. Beyond perceptual symbols: A call for representational pluralism. Cognition [Internet]. 2009 Mar [cited 2018 Apr 26];110(3):412–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19135654

15. Andrews M, Vigliocco G, Vinson D. Integrating experiential and distributional

data to learn semantic representations. Psychol Rev [Internet]. 2009 Jul [cited 2013 Mar 11];116(3):463–98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19618982

16. Norbury CF, Gooch D, Wray C, Baird G, Charman T, Simonoff E, et al. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. J Child Psychol Psychiatry [Internet]. 2016 Nov [cited 2017 Feb 28];57(11):1247–57. Available from: http://doi.wiley.com/10.1111/jcpp.12573

17. Rice ML. Language growth and genetics of specific language impairment. Int J Speech Lang Pathol [Internet]. 2013 Jun [cited 2016 Mar 30];15(3):223–33. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3684183&tool=pmcentrez&rendertype=abstract

18. McGregor KK, Oleson J, Bahnsen A, Duff D. Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. Int J Lang Commun Disord [Internet]. 2013 May [cited 2018 Apr 26];48(3):307–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23650887

19. Henry LA, Botting N. Working memory and developmental language impairments. Child Lang Teach Ther [Internet]. SAGE PublicationsSage UK: London, England; 2017 Feb 26 [cited 2018 Apr 26];33(1):19–32. Available from: http://journals.sagepub.com/doi/10.1177/0265659016655378

20. Romberg AR, Saffran JR. Statistical learning and language acquisition. Wiley Interdiscip Rev Cogn Sci [Internet]. Wiley-Blackwell; 2010 Nov 1 [cited 2018 Apr 26];1(6):906–14. Available from: http://doi.wiley.com/10.1002/wcs.78

21. Ebert KD, Kohnert K. Sustained attention in children with primary language impairment: a meta-analysis. J Speech Lang Hear Res [Internet]. NIH Public

Access; 2011 Oct [cited 2018 Apr 26];54(5):1372–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21646419

22. Dunn LM, Dunn LM, Whetton C, Burley J. British Picture Vocabulary Scale, 2nd edition [Internet]. Windsor, UK: NFER-Nelson; 1997 [cited 2016 Jan 7]. Available from: http://www.researchconnections.org/childcare/resources/7656

23. Wechsler D. Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II) [Internet]. San Antonio, TX: NCS Pearson; 2011 [cited 2016 Jan 7]. Available from: http://www.pathwaysstudy.pitt.edu/codebook/wasi-sb.html

24. Semel E, Wiig EH, Secord WA. Clinical Evaluation of Language Fundamentals (4th ed.). San Antonio, TX: PsychCorp; 2006.

25. Kuperman V, Stadthagen-Gonzalez H, Brysbaert M. Age-of-acquisition ratings for 30,000 English words. Behav Res Methods [Internet]. 2012 Dec [cited 2016 Jan 7];44(4):978–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22581493

26. Brysbaert M, Warriner AB, Kuperman V. Concreteness ratings for 40 thousand generally known English word lemmas. Behav Res Methods [Internet]. 2014 Sep [cited 2016 Jan 7];46(3):904–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24142837

27. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behav Res Methods [Internet]. 2013;45:1191–207. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23404613

28. Balota D a, Yap MJ, Cortese MJ, Hutchison K a, Kessler B, Loftis B, et al. The English Lexicon Project. Behav Res Methods [Internet]. 2007 Aug;39(3):445–59. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3193910&tool=pmcentrez&rendertype=abstract

29. van Heuven WJB, Mandera P, Keuleers E, Brysbaert M. SUBTLEX-UK: a new and improved word frequency database for British English. Q J Exp Psychol (Hove) [Internet]. 2014 Jan [cited 2016 May 10];67(6):1176–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24417251

30. Audacity Team X. Audacity ® [Internet]. Pittsburgh: Carnegie Mellon University; Available from: http://audacityteam.org/

31. Psychology Software Tools. E-Prime 2.0 [Internet]. 2012. Available from: https://www.pstnet.com

32. Wechsler D. Wechsler intelligence scale for children-fifth edition. Bloomington, MN: Pearson; 2014.

33. Barca L, Mazzuca C, Borghi AM. Pacifier Overuse and Conceptual Relations of Abstract and Emotional Concepts. Front Psychol [Internet]. Frontiers; 2017 Dec 1 [cited 2018 Apr 26];8:2014. Available from: http://journal.frontiersin.org/article/10.3389/fpsyg.2017.02014/full

34. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria; 2017. Available from: https://www.r-project.org/

35. Kuznetsova A, Brockhoff P, Christensen R. lmerTest Package: Tests in Linear Mixed Effects Models. J Stat Softw. 2017;82(13):1–26.

36. Christensen R. ordinal - Regression Models for Ordinal Data. R package version 2018.4-19. 2018.

37. Bürkner P-C. brms: An R Package for Bayesian Multilevel Models Using Stan. J Stat Softw [Internet]. 2017 [cited 2018 Apr 26];80(1). Available from: http://www.jstatsoft.org/v80/i01/

38. Jeffreys H. Theory of probability [Internet]. Clarendon Press; 1998 [cited 2018 Apr 26]. 459 p. Available from: https://global.oup.com/academic/product/theory-of-probability-9780198503682?cc=us&lang=en&

39. Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc [Internet]. 1995 Jun [cited 2018 Apr 26];90(430):773–95. Available from: http://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572

40. Greenacre MJ, Blasius J. Correspondence analysis in the social sciences: recent developments and applications [Internet]. 1994 [cited 2018 Apr 26]. Available from: http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=SUV.xis&method=post&formato=2&cantidad=1&expresion=mfn=010330

41. Greenacre MJ. Correspondence analysis in practice. CRC Press; 2016.

42. Alberti G. CAinterprTools: An R package to help interpreting Correspondence Analysis' results. SoftwareX [Internet]. Elsevier; 2015 Sep 1 [cited 2018 Apr 26];1–2:26–31. Available from: https://www.sciencedirect.com/science/article/pii/S2352711015000060

43. Crawford JR, Garthwaite PH. Evaluation of criteria for classical dissociations in single-case studies by Monte Carlo simulation. Neuropsychology [Internet]. 2005 [cited 2016 Mar 30];19(5):664–78. Available from: http://homepages.abdn.ac.uk/j.crawford/pages/dept/pdfs/Neuropsychology_2005_classical_dissociations.pdf

44. Fox JR. A Signal Detection Analysis of Audio/Video Redundancy Effects in Television News Video. Communic Res [Internet]. 2004 Oct 1 [cited 2016 Jan 8];31(5):524–36. Available from: http://crx.sagepub.com/content/31/5/524.short

45. Macmillan NA, Creelman CD. Detection Theory: A User's Guide [Internet]. Cambridge, UK: Cambridge University Press; 2004 [cited 2016 Jan 8]. 512 p. Available from: https://books.google.com/books?hl=en&lr=&id=P094AgAAQBAJ&pgis=1

46. Shapiro MA. Signal Detection Measures of Recognition Memory. In: Lang A, editor. Measuring Psychological Responses To Media Messages [Internet].

Hillsdale, NJ, England: Lawrence Erlbaum Associates; 1994 [cited 2016 Jan 8]. p. 133–48. Available from: https://books.google.com/books?hl=en&lr=&id=D-ZQAwAAQBAJ&pgis=1

47. Kousta S-T, Vigliocco G, Vinson DP, Andrews M, Del Campo E. The representation of abstract words: why emotion matters. J Exp Psychol Gen [Internet]. 2011 Feb [cited 2013 Oct 21];140(1):14–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21171803

48. Evans JL, Saffran JR, Robe-Torres K. Statistical learning in children with specific language impairment. J Speech Lang Hear Res [Internet]. 2009 Apr [cited 2016 Apr 12];52(2):321–35. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3864761&tool=pmcentrez&rendertype=abstract

Table 1. *Demographic characteristics of DLD and TD children and performance at the background tests, means (SD).*

	Age-matched				Vocabulary-matched		
	DLD	TD_{age}	*t*-test	*p*	TD_{voc}	*t*-test	p
Age	10.40 (1.83)	10.33 (1.44)	0.127	.899	8.16 (2.12)	3.383	.002
Matrix reasoning	40.33 (10.67)	49.22 (9.37)	2.656	.012	51.41 (8.31)	3.413	.002
BPVS	108.72 (25.03)	129.66 (14.74)	3.059	.004	109 (24.25)	0.034	.973
CELF recall sentence	4.83 (4.23)	NA	-	-	NA	-	-

Table 2. *Lexical and sublexical characteristics of the words used, means (SD).*

	Variable	Concreteness category: Abstract	Concreteness category: Concrete	*t*-test	*p*
AoA band 1	Concreteness[a]	337.22 (45.7)	576.44 (35.8)	12.338	< 0.001
	Length (no. of letters)	5.44 (1.0)	5.44 (1.0)	0.000	1.000
	Valence[b]	5.10 (1.9)	5.10 (1.9)	0.006	0.995
	Age of acquisition[c]	5.22 (0.94)	4.80 (0.80)	1.401	0.180
	CBBC frequency[d]	4.7 (0.43)	4.64 (0.31)	0.505	0.621
	Familiarity[a]	566.33 (30.74)	565.33 (27.47)	0.049	0.962
AoA band 2	Concreteness[a]	319.56 (50.55)	509.22 (70.78)	6.542	< 0.001
	Length (no. of letters)	4.67 (0.5)	4.67 (0.5)	0.000	1.000
	Valence[b]	4.93 (1.69)	4.90 (1.66)	0.025	0.980
	Age of acquisition[c]	7.16 (1.23)	6.67 (1.53)	0.756	0.460
	CBBC frequency[d]	4.45 (0.45)	4.39 (0.40)	0.280	0.783
	Familiarity[a]	543.0 (16.18)	533.67 (23.02)	0.995	0.335
AoA band 3	Concreteness[a]	334.11 (26.7)	517.78 (71.0)	7.265	< 0.001
	Length (no. of letters)	5.67 (1.32)	5.78 (1.20)	0.000	1.000
	Valence[b]	4.94 (1.27)	4.75 (1.81)	0.28	0.807

	Age of acquisition[c]	9.04 (1.44)	9.14 (1.53)	0.157	0.877
	CBBC frequency[d]	3.56 (1.32)	3.56 (120)	0.005	0.996
	Familiarity[a]	464.67 (61.01)	463.56 (74.72)	0.035	0.973
AoA band 4	Concreteness[a]	322.78 (41.37)	495.13 (67.38)	6.442	< 0.001
	Length (no. of letters)	6.33 (1.32)	6.22 (1.09)	0.194	0.848
	Valence[b]	4.90 (1.47)	5.08 (1.42)	0.272	0.789
	Age of acquisition[c]	10.71 (0.78)	10.74 (0.47)	0.079	0.938
	CBBC frequency[d]	3.26 (0.74)	3.05 (0.52)	0.680	0.506
	Familiarity[a]	430.56 (60.0)	448.38 (56.19)	0.630	0.538

Note: a. (31); b. (56, 57); c. (30); d. (34).

Table 3. *Summary of p-values from the mixed-effects model comparisons, Bayes Factors (BF10), and their interpretation. Asterisks indicate significant p-values (at p < .05) or BF10 indicating either support for H0 (BF10 < 1/3) or support for H1 (BF10 > 3).*

		Effect	p-value	BF10	BF notes
Lexical Decision	DLD vs TD_{age}	group:concreteness	.790	0.099*	H_0 favoured
		group	.037*	0.85	inconclusive
		concreteness	.622	0.172*	H_0 favoured
	DLD vs TD_{voc}	group:concreteness	.866	0.144*	H_0 favoured
		group	.570	0.12*	H_0 favoured
		concreteness	.481	0.25*	H_0 favoured
Definition score	DLD vs TD_{age}	group:concreteness	.011*	17.2*	H_1 favoured
		(TD_{age})concreteness	.485	2.22	inconclusive
		(DLD)concreteness	.132	5.06*	H_1 favoured
		(abstract)group	.002*	54.6*	H_1 favoured
		(concrete)group	< .001	1524.9*	H_1 favoured
	DLD vs TD_{voc}	group:concreteness	.556	0.89	inconclusive
		group	.010*	31.2*	H_1 favoured
		concreteness	.158	4.86*	*H_1 favoured*
Definition quality ratings	DLD vs TD_{age}	group:concreteness	.208	0.892	inconclusive
		group	.007*	11.68*	H_1 favoured
		concreteness	.049*	2.13	inconclusive
	DLD vs TD_{voc}	group:concreteness	.268	0.79	inconclusive
		group	.010*	7.78*	H_1 favoured
		concreteness	.051	0.78	inconclusive

List of figure legends

Figure 1 - Proportion of correct responses to abstract and concrete words, comparing performance of DLD with TD_{age} (N = 14; left), and and DLD with TD_{voc} (N = 12; right) children. Error bars indicate standard error of the mean.

Figure 2 – Average total score of definitions to abstract and concrete words, comparing performance of DLD with TD_{age} (N = 18; left), and with TD_{voc} (N = 17; right) children. Error bars indicate standard error of the mean.

Figure 3. Average ratings (from adult native English speakers) for abstract and concrete words' definitions provided by DLD children and their matched TD_{age} peers (left), and by DLD children and their matched TD_{voc} peers (right). Error bars indicate standard deviations.

Figure 4 – Representational Similarity Analysis for Concrete and Abstract words. Error bars indicate 95% confidence intervals. The dashed line corresponds to the RM value for each parameter. Left: correlations for the set of simulations involving all the valid combinations of values for the three parameters. For example, in the case of window size (and similarly for the other parameters, in turn), we first consider the correlations obtained from all the models for which *winSize* has the reference value of 5 (regardless of the values for *learnRate* and *novel Bias*), and compare them to correlations from all the models that have "lesioned" values for *winSize*, namely 3 and 1 (regardless of the values for *learnRate* and *novelBias*). Right: correlations for the set of simulations where only one parameter was allowed to vary, while the other two were kept to their reference values.