languages

MDPI

Article

How Children With and Without Developmental Language Disorder Use Prosody and Gestures to Process Phrasal Ambiguities

Albert Giberga [1,*], Ernesto Guerra [2], Nadia Ahufinger [1], Alfonso Igualada [1], Mari Aguilera [3] and Núria Esteve-Gibert [1]

[1] Faculty of Psychology and Education Sciences, Universitat Oberta de Catalunya, 08018 Barcelona, Spain; nadiahufinger@uoc.edu (N.A.); nesteveg@uoc.edu (N.E.-G.)
[2] Center for Advanced Research in Education, Institute of Education (IE), Universidad de Chile, Santiago 8330015, Chile; ernesto.guerra@ciae.uchile.cl
[3] Department of Cognition, Development and Educational Psychology, Universitat de Barcelona, 08007 Barcelona, Spain; mari.aguilera@ub.edu
* Correspondence: agiberga@uoc.edu

Abstract: Prosody is crucial for resolving phrasal ambiguities. Recent research suggests that gestures can enhance this process, which may be especially useful for children with Developmental Language Disorder (DLD), who have impaired structural language. This study investigates how children with DLD use prosodic and gestural cues to interpret phrasal ambiguities. Catalan-speaking children with and without DLD heard sentences with two possible interpretations, a high (less common) and low (more common) attachment interpretation of the verb clause. Sentences were presented in three conditions: baseline (no cues to high-attachment interpretation), prosody-only (prosodic cues to high-attachment interpretation), and multimodal (prosodic and gestural cues to high-attachment interpretation). Offline target selection and online gaze patterns were analysed across linguistic (DLD vs. TD) and age groups (5–7 vs. 8–10 years old) to see if multimodal cues facilitate the processing of the less frequent high-attachment interpretation. The offline results revealed that prosodic cues influenced all children's comprehension of phrasal structures and that gestures provided no benefit beyond prosody. Online data showed that children with DLD struggled to integrate visual information. Our findings underscore that children with DLD can rely on prosodic cues to support sentence comprehension and highlight the importance of integrating multimodal cues in linguistic interactions.

Keywords: developmental language disorder; phrasing; prosody; gestures

check for updates

Academic Editor: Michael Robb

Received: 20 December 2024
Revised: 12 March 2025
Accepted: 18 March 2025
Published: 26 March 2025

Citation: Giberga, A., Guerra, E., Ahufinger, N., Igualada, A., Aguilera, M., & Esteve-Gibert, N. (2025). How Children With and Without Developmental Language Disorder Use Prosody and Gestures to Process Phrasal Ambiguities. *Languages*, *10*(4), 61. https://doi.org/10.3390/languages10040061

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Multimodal signals accompanying speech convey information about the meanings intended by the speaker, to an extent that many authors consider that the combination of visual and spoken signals is part of a communication system that is multimodal by nature (Hagoort & Özyürek, 2024; Holler & Levinson, 2019). In this context, prosodic modulations of the voice and communicative movements of the body (hand gestures, facial expressions, head movements and body posture) significantly contribute to sentence meaning, also for language development. Prosodic information is crucial for infants' perception of word and phrase boundaries at early stages of language acquisition, as well as for children's later production and comprehension of sentence structure and sociopragmatic meanings (see reviews in Chen et al., 2020; Prieto & Esteve-Gibert, 2018). Similarly, infants' and children's

communicative body movements predict and scaffold the development of other linguistic milestones (Rowe et al., 2022) and are tightly connected with the prosodic structure of speech (Esteve-Gibert & Prieto, 2014) from very early on.

The present study focuses on one domain in which prosodic cues and body movements are found to contribute to linguistic meaning: the processing of sentences that are ambiguous at the phrasal level. When two different interpretations of the same sentence are possible, such as in Someone shot the maid of the actress who was on the balcony (Fernández & Sekerina, 2015), where the relative clause 'who was in the balcony' could be attached to 'actress' to indicate that the actress was on the balcony or instead be attached to the 'maid' to indicate that the maid was on the balcony, the position of a prosodic boundary is found to help English adult listeners to decide between one of the two possible interpretations (Jun & Bishop, 2014; Nagel et al., 1996; Pynte & Prieur, 1996; Snedeker & Trueswell, 2004; Weber et al., 2006). Thus, a prosodic boundary after 'maid' would favour the first interpretation, whereas a prosodic boundary break after 'actress' would favour the second one. The presence of gestures accompanying speech also aids in interpreting complex syntactic structures (e.g., object–cleft–construction; Theakston et al., 2014); they contribute to the interpretation of phrasal ambiguous sentences (Guellaï et al., 2014), and they reduce the cognitive effort when processing complex structures (Holle et al., 2012). While this previous evidence shows that both prosody and gesture signal phrasal structures are used for their comprehension (Guellaï et al., 2014), the relative weight of prosodic cues and visual information in processing phrasal boundaries is yet to be compared, especially in children with diverse linguistic abilities. Mouth movements seem to help in speech segmentation when auditory signals are limited or noisy (Mitchel & Weiss, 2014), and head nods, added to prosodic breaks, help in segmenting an artificial language into phrase-like units (De La Cruz-Pavía et al., 2020, 2022). Whether and how typically developing children use prosodic and gestural information to process phrasal ambiguities, and whether the presence of these highlighting multimodal cues can be especially beneficial for children with difficulties in sentence comprehension, remains to be understood.

Cross-linguistic differences have been observed in the way adult listeners process phrasal ambiguities. English speakers generally show a preference for low attachment (i.e., in the example above, to interpret that the actress was on the balcony, instead of the maid) when processing ambiguous sentences (Felser et al., 2003; Grillo et al., 2015). In Romance languages like Spanish or Catalan, the listeners' preference seems to be different. In Spanish, adult speakers exhibit a preference for high attachment (i.e., in the example above, to interpret that the maid was on the balcony, instead of the actress), although the preferences may vary depending on the nature of the syntactic ambiguity (see, for instance, Cuetos & Mitchell, 1988; Gilboy et al., 1995; Hemforth et al., 2015). In Catalan, the language under investigation in the present study, attachment preferences are less clear, and research is relatively scarce. Only one study (Prieto, 1997) has investigated whether the position of phrasal boundaries affected the interpretation of ambiguous sentences such as in (1) below. In their study, stimuli were presented only in a written modality, and the results showed that Catalan adult speakers tend to prefer the low-attachment interpretation as in (a) if no written cues are presented. Instead, when a break was signalled after the last referent llança ('lance'), a high-attachment interpretation was preferred in 40% of the cases. Prieto (1997) concluded that the low-attachment interpretation is the 'default' interpretation in Catalan and that prosodic cues may play a significant role in guiding less-frequent high-attachment preferences when the input is presented orally. In this context, it could well be that speech-accompanying gestures can amplify this effect, especially when language is under development or impaired.

(1) a. [La vella]NP [llança l'amenaça]VP
[The old woman] [throws the threat]
b. [La vella llança]NP [l'amenaça]NP
['The old lance'] ['threats (somebody)']

Developmental studies on how children use prosodic signals to resolve phrasal ambiguities suggest that this ability develops around 6 years of age. Some studies conducted in Korean and English report that 3 to 6 year olds do not detect prosodic phrasing for sentence comprehension yet (Choi & Mazuka, 2003; Snedeker & Yuan, 2008; Vogel & Raimy, 2002; Wiedmann & Winkler, 2015). For example, Snedeker and Trueswell (2001) investigated whether mothers use prosodic cues in their productions to structure phrasally ambiguous sentences, such as 'Tap the frog with the flower', and whether children use these prosodic cues for comprehension. The study found that while mothers varied their production to mark prosodically the referent in the sentence that should be attached to the preposition 'with' (and so, they indicated prosodically whether the flower was used to tap the frog or whether the frog that was tapped had a flower), children did not show different responses to the different versions of the sentences. On the contrary, Wiedmann and Winkler (2015) reported that 6 year olds were indeed more accurate at resolving ambiguities when a prosodic boundary indicated the scope of the intonational phrase (e.g., [Mandy plays the boy's drum.] versus [Mandy plays.] [The boys drum.]), and Snedeker and Yuan (2008) also found that children as young as 4 years of age could make use of prosody to resolve ambiguous prepositional-phrase attachments. These previous findings suggest that the children's use of prosodic cues for sentence interpretation develops around age 6 and may be mastered some years later.

Next to prosody, gestures also scaffold the children's interpretation of sentence meaning (Armstrong et al., 2014; Hübscher et al., 2017). Whether gestures are also used by children to interpret sentence structure remains to be investigated, but it seems reasonable to think that this may be the case in light of previous evidence in adults (Biau et al., 2018; Guellaï et al., 2014) as well as in young infants (De La Cruz-Pavía et al., 2019; Hollich et al., 2005). The importance of prosodic and gesture cues in driving children's interpretation of sentence structure might be even stronger in the case of children with language difficulties, such as those who have Developmental Language Disorder (DLD), because they experience difficulties with sentence comprehension (Bishop & Adams, 1992), especially when the sentences are syntactically complex (Leonard, 1998), and additional enhancing cues have been shown to facilitate comprehension in other linguistic domains (see, for instance, Giberga et al., 2024 for pragmatic comprehension).

Children with DLD experience ongoing language difficulties that cannot be explained by medical conditions (Bishop et al., 2017). A key characteristic of children with DLD is their struggle with grammar, particularly in the area of morpho-syntax. These difficulties are mostly found when children with DLD are compared to age-matched TD controls but reduced when linguistic levels are controlled for (Marshall et al., 2009), showing that children with DLD may overcome such difficulties as they move forward in the language development process (Christou et al., 2020; Coloma et al., 2024). Some studies also point at general processing deficits in this population (Joanisse & Seidenberg, 1998; Miller et al., 2001; Montgomery et al., 2018; Plym et al., 2021). Interestingly, an eye-tracking study with Catalan-speaking children with DLD found that subtle processing difficulties in this population are not always observed in more traditional offline behavioural measures (Andreu et al., 2011). This underscores the importance of using methodologies such as eye tracking in our study, as it allows for the detection of nuanced deficits in real-time language processing. Children with DLD also exhibit poorer phonological abilities at the segmental level (Aguilar-Mediavilla et al., 2002; Leonard, 2014; Maillart & Parisse, 2006;

Vukovic et al., 2022), while at the suprasegmental (prosodic) level, results are more mixed and depend on the children's age and the linguistic function in which prosody is evaluated (Calet et al., 2021; Marshall et al., 2009; Sabisch et al., 2009; Van Der Meulen et al., 1997). To our knowledge, only one study investigated how children with DLD use prosodic information for syntactic disambiguation (Caccia & Lorusso, 2019), which was conducted in an Italian population. In this study, the authors designed a picture-matching task with sentences such as in (2), with temporary syntactic closure ambiguity. Their results showed that both children with DLD and TD children (aged 10 to 13) used prosody to disambiguate the syntactic structures.

(2) a. [Quando Marta guida] [la macchina fuma]
['When Mary drives'] ['the car smokes']
b. [Quando Marta guida la macchina] [fuma]
['When Mary drives the car'] ['(she) smokes']

Instead, children with DLD of the same age may experience difficulties using prosody for other linguistic functions such as in focus interpretation (Marshall et al., 2009) or in processing passive structures (Sabisch et al., 2009), and younger children with DLD struggle using prosody to distinguish a statement from a question (Calet et al., 2021; Giberga et al., 2024). Altogether, these findings suggest that younger children with DLD may use prosody for sentence comprehension differently than older children with DLD and that these age-related effects may be reduced when children with DLD are compared to TD peers with language-matched abilities (Marshall et al., 2009).

As for gestures, children with DLD might exploit multimodal cues for sentence comprehension to a larger extent than TD children. There is robust evidence that gestures improve vocabulary acquisition (Lüke et al., 2020; Vogt & Kauschke, 2017) and sentence comprehension (Modyanova et al., 2024) of children with DLD and that gestures seem to help children with DLD achieve the same level of comprehension of complex pragmatic inferences (Giberga et al., 2024; Kirk et al., 2010) that, typically, developing children reach using only prosodic cues in speech. Children with DLD, thus, might benefit from the presence of gestures also to disambiguate phrasal structure, but the way gestures interact with prosodic cues for that matter and whether this relationship evolves with children's age remain to be explored.

The aims of the present study were (1) to investigate the extent to which Catalan-speaking children with DLD use prosodic cues to interpret phrasal structures, compared to children with TD, (2) to determine whether body gestures enhance this process in children with DLD compared to children with TD, and (3) to explore whether multimodal cues impact differently at distinct developmental stages. To do so, we designed a visual-world eye-tracking experiment in which we presented ambiguous sentences at the phrasal level to four subgroups of children that differed in their linguistic profile (DLD vs. TD) and age (younger vs. older). The sentences were presented in three different conditions (baseline, prosody-only or multimodal) in a within-subjects design. We recorded the children's looking patterns toward a target referent (and its competitors and distractors) during the unfolding of the sentence, as well as the children's offline behavioural response when asked to select the target referent.

We formulated four main general predictions, namely that (1) children with DLD would rely on prosodic cues to a lesser extent than their typically developing peers due to their deficits at the phonological and prosodic level; that (2) multimodal cues (i.e., the presence of gesture accompanying prosodic cues in speech) would impact the interpretation of phrasal structures, compared to when only prosodic information is presented, especially in children with DLD; and (3) that gestures would facilitate comprehension more signifi-

cantly for younger children, compared to older children. Regarding the online processing of the phrasally ambiguous sentences, we predicted (4) that prosodic information would impact the preferred interpretation (i.e., more and faster looks to images representing the sentence to be processed), especially in the TD and older subgroups, and that multimodal information would compensate for the lack of impact of prosodic cues on gaze preferences in the gaze preferences in the DLD and younger children.

2. Materials and Methods

2.1. Participants

A total of 34 children with DLD (16 female) and 45 TD children (22 female) participated in this study. The TD children were recruited from a primary school in Barcelona (Escola Poblenou). Children with DLD were mainly recruited through educational services providing support to schools (i.e., the CREDA centres) in the Barcelona in helping children with speech, language and communication needs), as well as through the research group's social media. All children were bilingual in Catalan and Spanish, with high proficiency in Catalan, which is the main language of instruction at schools in Catalonia, the language used during the testing and widely spoken in the area (90% according to the Idescat institute). Families also reported that children spoke other languages, including English (n = 6), Romanian (n = 2), French (n = 1), and Arabic (n = 1). Despite the children's multilingual background, it was confirmed with the school and/or the families that all the children included in the sample were able to complete the task in Catalan. Participants were divided into two age subgroups: a younger group (age range 5:0–7:11; average 6:2) and an older group (age range 8:0–10:11; average 8:9). None of the children had self-reported hearing disorders (see Table 1 for more detail).

Table 1. Average measures per group for age, nonverbal IQ (KBIT-MAT) and core language (CELF-5).

	DLD		**TD**	
Age Group	**Younger**	**Older**	**Younger**	**Older**
Age (mean)	6.18	8.88	6.05	8.67
Age (std)	0.97	0.76	0.92	0.82
NONV IQ [1] (mean)	97.08	89.29	106.65	99.14
NONV IQ (std)	15.71	12.35	13.12	12.54
NONV IQ (min)	74	70	74	71
NONV IQ (max)	120	120	130	119
Core Language [2] (mean)	71.75	75.71	93.29	92.14
Core Language (std)	9.91	14.53	14.33	14.64
Core Language (min)	50	57	50	73
Core Language (max)	85	103	125	118

[1] Kaufman Brief Intelligence Test. [2] Clinical Evaluation of Language Fundamentals (CELF-5).

This study was approved by the Ethics Committee of the Name University. All families were asked to sign a consent form to participate.

In order to divide children in the two linguistic groups (DLD or TD), all children were assessed for their oral language and nonverbal intelligence. Oral language was evaluated using the Core Language Score of the Clinical Evaluation of Language Fundamentals—Fifth Edition, Spanish (Wiig et al., 2013), and the nonverbal intelligence was assessed through the Kaufman Brief Intelligence Test (Kaufman, 1990). All children in this study scored $\geq$70 in the nonverbal IQ test (KBIT-MAT). Criteria for the inclusion in the DLD group were scoring -1 SD on the CELF Core Language and, at the same time, having parents or educators reporting language difficulties (Bishop et al., 2017; Castilla-Earls et al., 2020). Criteria for the

inclusion in the TD group were no history nor diagnosis of language learning disability and parents and or educators reporting no concern about their language and learning skills, as well as scoring above −1 SD on the CELF core language. None of the children included in this study had a reported history of hearing impairments, speech sound disorder, or autism.

2.2. Experimental Materials

2.2.1. Preliminary Task for Sentence Creation

We designed a preliminary study to confirm previous results by Prieto (1997), which suggested that, unlike Spanish and English speakers, Catalan speakers interpret low attachment as the 'default' interpretation in ambiguous phrasal structures, especially because, in that previous study, only written stimuli were presented. The results of this preliminary study would then allow us to associate our 'baseline' condition with the most frequent interpretation and to evaluate whether the presence of prosodic and gestural cues in the experimental (prosody-only and multimodal) conditions can activate the less frequent interpretation. Following Gilboy et al. (1995), ten adult participants read a list of two sentence types, a noun phrase ('L'home saluda la dona amb el barret'/'The man greeted the woman with the hat') and an adjective phrase ('*L'home parla amb el nen malalt*'/literally, 'The man talks to the boy sick'), and were asked to indicate their initial fast interpretation upon reading them. The results showed that adjective phrases like '*L'home parla amb el nen malalt*' were less ambiguous (with a 78% preference for low attachment) than noun phrases like '*L'home saludava la dona amb el barret*' (which had a 48% preference for low attachment). The less ambiguous adjective phrases were then chosen to be used in our study, because we wanted listeners to have a strong preference for a particular interpretation in the baseline condition that could then be revised in case prosodic and multimodal cues were presented.

We then created twelve target sentences with two potential phrasal interpretations, a more frequent 'default' low-attachment interpretation such as in (3a) and a less frequent high-attachment interpretation such as in (3b). Target sentences included high-frequency verbs to maximize lexical accessibility, especially for children with DLD.

(3) a. [L'home parla] [amb el nen malalt] - default interpretation in Catalan
[The man talks]H- [to the sick boy]
b. [L'home parla amb el nen] [malalt]
[The man talks to the boy]H- [sick]

2.2.2. Oral Presentation of Target Sentences

All the sentences were presented in three conditions. In the baseline condition, sentences were produced with a continuation rise right after the verb 'talks' and preceding the predicate, favouring a low-attachment interpretation (the boy is sick) due to the lack of prosodic break between the last referent and the adjective, as in 3a. In the prosody-only and multimodal conditions, instead, sentences were produced with a marked continuation rise after the last referent ('boy'), as in (3b), provoking a prosodic break in that position and, thus, favouring the high attachment of the verb clause (the man is sick). See Figure 1 for pitch track differences between conditions. Phonologically, both phrases were characterised by a L+H* H- at the intermediate phrase, which coincided with the verb, followed by a downstepped H* L% nuclear configuration at the end of the intonational phrase. However, in (3b), the internal break was stronger (level 4 in ToBI terms) and occurred later in the utterance, coinciding with the second referent (see Figure 1). In addition, the pitch range was, on average, wider in the prosody-only and multimodal conditions (103.6 Hz), compared to the baseline (34.9 Hz).

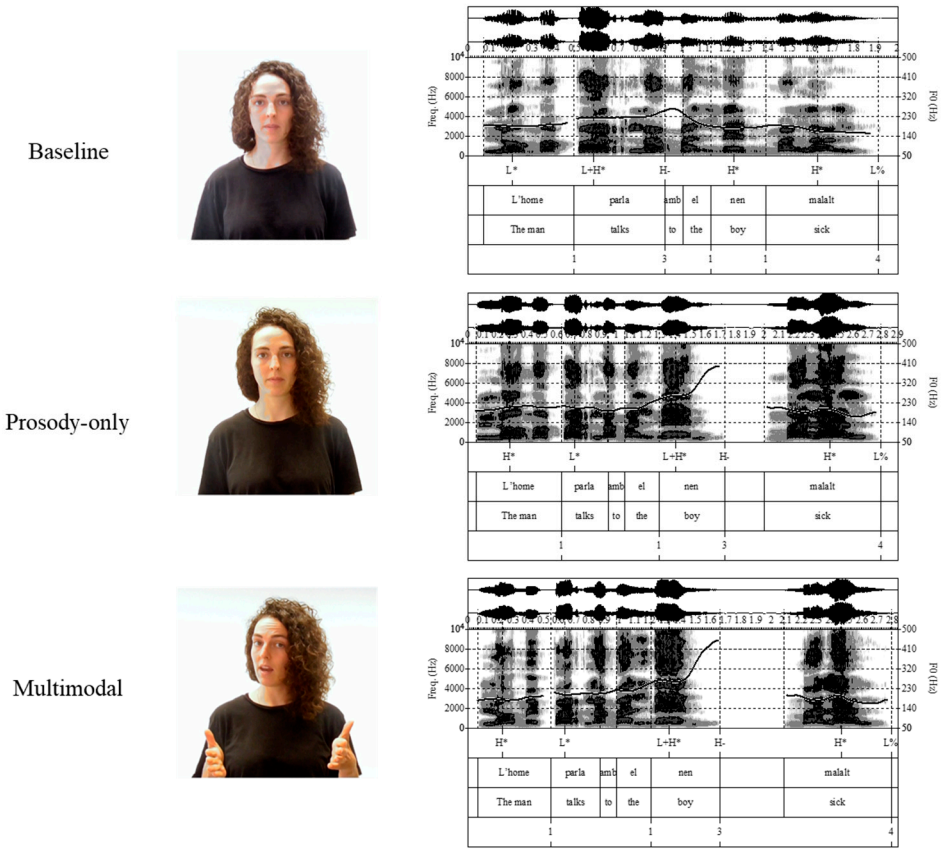

Figure 1. Screen caption of the gestures as they were produced during the internal phrasal break (indicated with number 3 in the lower row of the Praat annotations) (**left** panels), and pitch track, spectrogram and waveform (**right** panels) for the three conditions.

2.2.3. Visual Display

During the experimental items, a speaker appeared in the centre of the screen in video format and produced the target sentence to be processed by the children (see Figure 2B). In the baseline and prosody-only conditions, the speaker made no gestural cues during the production of the sentence and, thus, remained as static as possible. Instead, in the multimodal condition, the speaker produced a manual beat gesture and a head nod whose strokes were aligned to the nuclear pitch-accented syllable. To ensure precise alignment between prosodic and gestural cues, the speaker was trained by a researcher specialized in gesture–speech coordination. The production of manual beat gestures and head nods was rehearsed and recorded multiple times. After an initial pilot phase, adjustments were made to enhance the naturalness of the multimodal cues. The final stimuli were carefully examined to confirm that the apex of each manual beat gesture coincided in time with the nuclear pitch-accented syllable, following numerous previous findings (Loehr, 2012; Rohrer et al., 2023; Shattuck-Hufnagel & Ren, 2018), ensuring consistency across trials.

2.2.4. Response Images

Simultaneously to the speaker, four images appeared in the corners of the screen (see Figure 2B): the target image (bottom right in Figure 2B, corresponding to a less frequent high-attachment interpretation), the competitor image (bottom left in Figure 2B, corresponding to a low-attachment 'default' interpretation), a relevant distractor (top left in Figure 2B, with two relevant characters in the scene but performing completely different actions) and an irrelevant distractor (top right in Figure 2B, with only one relevant character making this interpretation completely impossible). To ensure that no specific image was inherently more salient than others, all response images were designed in black and white, and we

conducted pilot tests and several rounds of discussions among co-authors to ensure that they had comparable levels of detail and visual complexity. This standardization applied to all image types: target, competitor, relevant distractor, and irrelevant distractor.

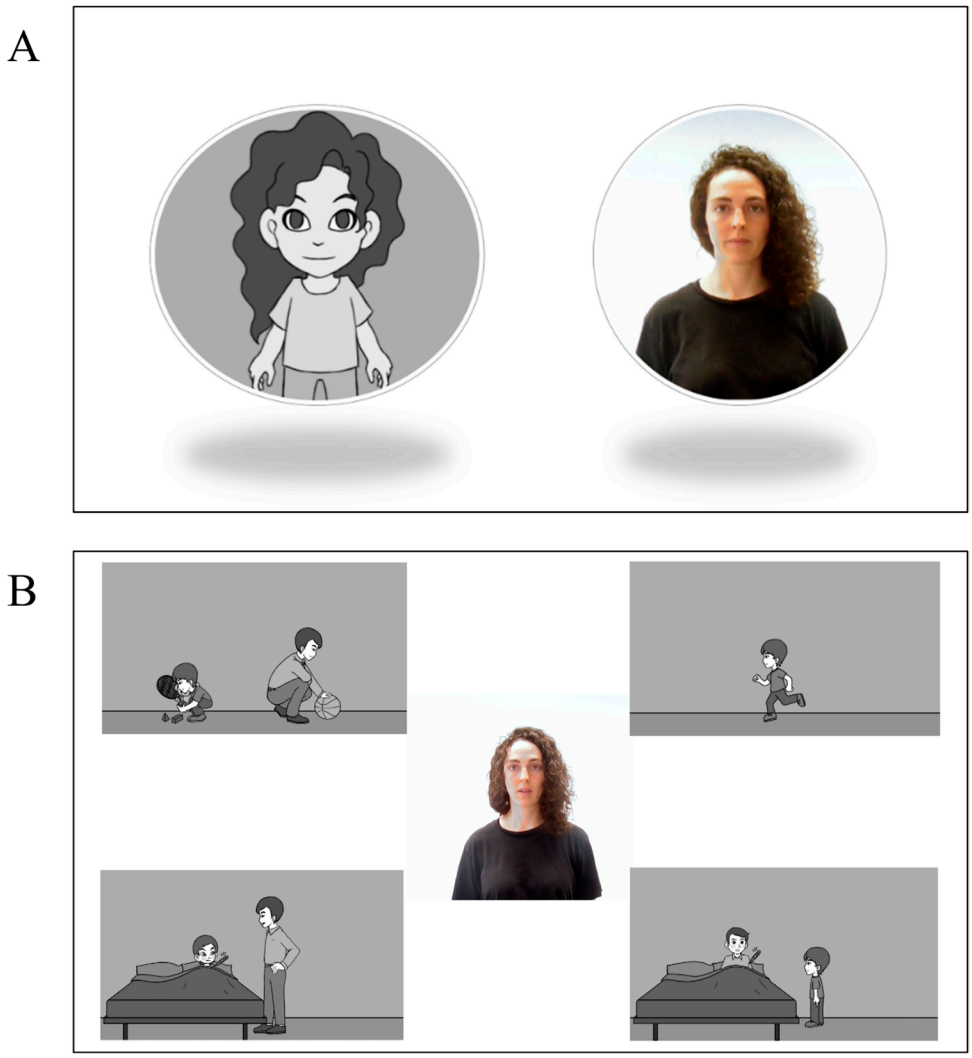

Figure 2. Visual depiction of the procedure during the eye-tracking experiment. The image (**A**) was used to introduce the main character of the story (Martina). The image (**B**) was the visual display during the unfolding of the target sentence produced by the main speaker, with the four response images around the video.

2.3. Procedure

In line with Silverman et al.'s (2010) work, we modified the original visual-world paradigm (Tanenhaus et al., 1995) to investigate the real-time processing of audio-visual stimuli. Children were positioned in front of a screen (approximately 60 cm away), and the experimenter gave the following instructions to each participant: 'Now you will play a game with a girl named Martina. Martina will appear in the centre of the screen and she will say a sentence. Simultaneously, four images will appear in the four corners of the screen. After Martina speaks, you will have to point to the image that best represents what Martina just said'.

The task began with a slide designed to connect the image of Martina speaking at the centre of the target scenes in the experimental trials to the character shown in the images to be selected (Figure 2A). Before the presentation of the target sentence, a context sentence introduced the two characters depicted in the corresponding trial, ensuring participants were familiarized with relevant vocabulary before the task. The four response images

were displayed 2000 milliseconds before Martina began speaking, and children had up to 10 s after the end of the sentence to select an image. This extended duration was chosen to ensure that children with DLD had sufficient time to process the sentence and make a decision before the scene concluded. Target sentences were presented under three conditions following a within-subject design: prosody-only, multimodal, and baseline (see Section 2.2 for details on the conditions). Eye gaze patterns were recorded using a Tobii Pro Nano 60 Hz eye tracker through the iMotions version 9.2 software (iMotions, 2023) on a 21.5-inch screen (1920 × 1080 resolution), and children's offline responses (pointing towards the target object) were manually annotated in a scoresheet.

The experiment consisted of 12 trials in total: 3 in the prosody condition, 3 in the multimodal condition, and 6 in the baseline condition. The larger number of baseline trials was intended to ensure that the competitor image, which was expected to be chosen in this condition, was just as plausible as the target image expected in the prosody-only and multimodal conditions. Three of the baseline trials were designated as fillers (excluded from the analysis), with the first serving as a familiarisation trial at the beginning of the task. This first trial was used to provide task instructions and its correspondent feedback to the children on how to relate the sentences Martina said with the possible interpretations represented by the images.

2.4. Data Analysis

2.4.1. Offline Responses

We computed the proportion of trials in which participants selected the target image across conditions. We ran Generalised Linear Mixed Models (GLMMs) in R (using the lme4 package (Bates et al., 2014)) with the proportion of target image selection as dependent variable. The fixed factors were experimental condition (three levels: prosody-only, multimodal, and baseline), linguistic group (two levels: DLD and TD), and age group (two levels: older and younger). We used the maximal random effect structure that converged with the model, which included random effects for participants and items and a by-participant random slope for experimental condition. We then used R's anova() function to compare pairs of models for assessing main effects and interactions between predictors.

2.4.2. Eye Movement Data

We analysed the first 8000 ms of the stimuli presentation. The last seconds of the stimuli presentation were not analysed due to the low number of eye movements collected (M looks per second until 8000 ms = 15,935; M looks per second after 8000 ms = 8121). Thus, we considered the 8000 ms from the start of the stimulus to be the end of the trial. We divided the screen into 4 main areas of interest (AOI) using the iMotions version 9.2 software (iMotions, 2023) corresponding to the target image, the competitor image, the relevant distractor, and the irrelevant distractor. Unlike the conventional design of the visual-world paradigm (Tanenhaus et al., 1995), the adapted design used in our study (inspired by Silverman et al., 2010) required children to visually process not only the 4 response images but also the speaker positioned at the centre of the screen. Because children with DLD may be affected differently by this additional non-conventional processing demand in a visual-world paradigm, we created a fifth area of interest to explore the impact of the speaker's presence in the centre of the screen in this design.

Using R version 4.3.3 software (https://www.R-project.org/, accessed on 17 March 2025), we calculated the proportion of looks to each AOI across items and participants for every 100 ms time window. Next, we ran nonparametric random permutation cluster analysis (Barr et al., 2014), which consisted of mixed-effect linear regressions, with the proportion of fixations as the dependent variable and the type of AoI (target; competitor;

speaker) as fixed effects (including random intercepts and random slopes for object for both participants and items). Fixations on the distractors were excluded from the analysis because they received a low proportion of fixations (M = 0.077 for relevant distractors and M = 0.054 for irrelevant distractors). Thus, two main comparisons were considered: the proportion of looks to target versus competitor (aimed at assessing the interpretation processing of children) and the proportion of looks to target versus speaker (aimed at evaluating the impact of the speaker's presence during the sentence processing). These two comparisons were run per each linguistic group, experimental condition and age group. We then applied a cluster-based approach to identify the 100 ms time bins with significant differences between objects. Clusters were formed when at least three 100 ms time bins showed significant differences. For each cluster, we generated a t-value distribution through the random permutation of object labels within each time point. This process involved 2000 iterations for each 100 ms time bin. We aggregated the t-values obtained from the data for each cluster and calculated the sum of absolute values of the largest t simulated for each time bin. This resulted in one sum t value for each cluster. To determine the statistical significance of our observed clusters, we computed the proportion of greater summed simulated t-values relative to the t-values obtained from our data. Following Helo et al. (2021), we considered proportions smaller than 0.025 as significant. This methodology allowed us to control Type 1 errors while identifying robust significant differences in fixation proportions between objects. We analysed the presence of clusters as appearing at three distinct time windows: Time window 1 extended from the beginning of the target sentence to the appearance of the phrasal break marked with prosodic and gestural cues (or absence thereof in the baseline condition, on average, 3623 ms after the onset of the trial). Time window 2 spanned from the appearance of these phrasal cues to the end of the sentence (on average, 4911 ms after the onset of the trial). Time window 3 extended from the end of the sentence until the end of the trial. All scripts used for the analysis can be found in the OSF project.

2.5. Predictions

Regarding the offline selection of the response images, we hypothesised a main effect of condition by which the highest percentage of target image selections would occur in the multimodal condition, followed by the prosody-only condition, and then finally the baseline condition. We also predicted an interaction between condition and linguistic group, expecting that the effect of multimodality would be more pronounced for children with DLD, and an interaction between condition and age by which younger children would rely on bodily signals more than older children.

As for the eye movement analysis, from the onset of the target sentence to the presentation of prosodic and multimodal signals (time window 1), we predicted no gaze preferences to target or competitor across linguistic groups, age groups and conditions. From the unfolding of prosodic and multimodal signals to the end of the target sentence (time window 2), we expected a weak preference of looks to the competitor image in the baseline condition across subgroups. We expected an increase in looks to the target in the prosody-only condition in the groups with TD and in all children in the case of the multimodal condition. From the end of the target sentence to 8 s after the start of the trial (time window 3), we expected a strong preference for the competitor in the baseline condition. We also expected more looks to the target image in the prosody-only condition, but we expected this preference to be stronger in the older (compared to younger) and TD (compared to DLD) groups. In the multimodal condition, we expected all children to prefer the target image, with gestures reducing the potential differences between groups in the prosody-only condition.

3. Results

3.1. Offline Selection of the Target Image

Figure 3 shows the proportion of selections to the target image across conditions, age and linguistic groups. Overall, we found a main effect of condition ($X^2 = 31.68$, $p < 0.01$), showing that children selected the target image more frequently in the multimodal and prosody-only conditions, compared to the baseline (baseline vs. prosody: $\beta = 5.180$, SE = 1.773, z = 2.921, $p < 0.01$; baseline vs. multimodal: $\beta = 5.015$, SE = 1.765, z = 2.841, $p < 0.01$). There was no significant difference between the prosody-only and multimodal conditions ($\beta = -0.1649$, SE = 0.4641, z = −0.355, $p = 0.72$). We also found a two-way interaction between condition and age group ($X^2 = 9.401$, $p < 0.01$), by which younger children selected the target image more than older children in the baseline condition ($\beta = -2.0836$, SE = 0.8440, z = −2.469, $p = 0.01$). Although this result was unexpected because children were supposed to select the competitor image in the baseline condition, it should be noted that, as Figure 3 shows, younger children still selected the competitor more than the target in the baseline condition, which was preferred in only 25% of the cases. There was a (non-significant) tendency of older children to select the target image more than younger children in the multimodal condition ($\beta = 0.9720$, SE = 0.5433, z = 1.789, $p = 0.07$). There was no significant interaction between condition and linguistic group ($X^2 = 1.56$, $p = 0.46$), meaning that the impact of the presence of prosodic and multimodal cues was similar in children with and without DLD.

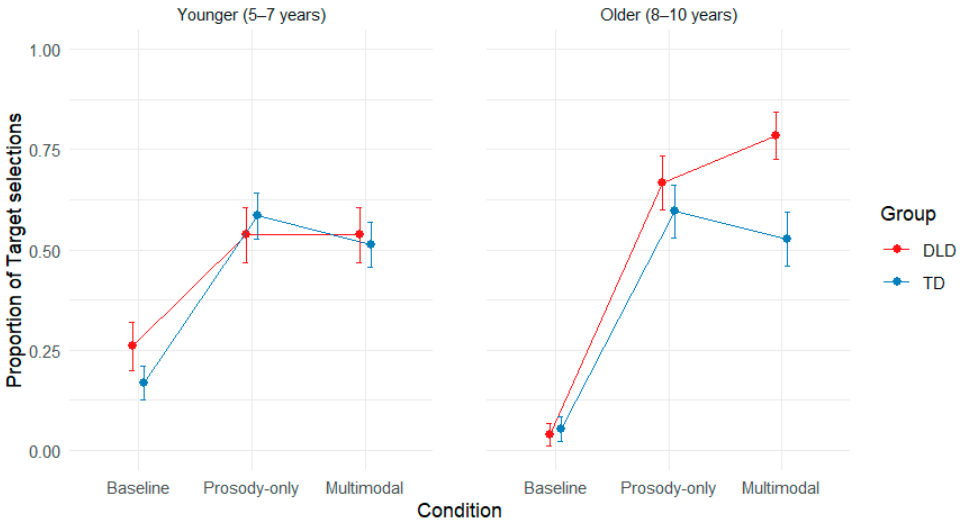

Figure 3. Accuracy in target selection for younger (**left** panel) and older children (**right** panel), across linguistic groups (DLD: red; TD: blue). The X axis shows the three experimental conditions: baseline (left columns), prosody (centre columns) and multimodal (right columns).

There was also a non-significant tendency in the three-way interaction between condition, linguistic group and age group, where older children with DLD select the target image more, compared to older TD children in the multimodal condition ($\beta = 1.4352$, SE = 0.8546, z = 1.679, $p = 0.09$). The full results of the models and all comparisons between conditions and subgroups can be found in the OSF project.

3.2. Gaze Preferences During the Eye-Tracking Task

The results reported below are structured according to distinct time windows (see Section 2.4) to provide a clear analysis of the online processing data. We focus on the progression of gaze preferences across these time windows, highlighting the differences between object preferences that are significant under different experimental conditions. For

detailed information about the specific statistical effects within each time point, see the Supplementary Material available on OSF.

3.2.1. Gaze Preference to Target vs. Competitor

Figure 4 shows the children's gaze preferences to target, competitor and speaker over time across conditions and linguistic groups. Prior to the unfolding of the prosodic and multimodal cues (or lack thereof, time window 1), in general, all children looked equally at the competitor and at the target in all three experimental conditions. However, in the baseline condition, older children with DLD looked at the competitor more, compared to the target (sum t = 16.2). The model also revealed a significant albeit unexpected difference between target and competitor at the end of this time window in the prosody condition, by which younger TD children shifted their gaze towards the target image in the prosody condition, even if no relevant prosodic cues were presented yet (sum t = 45.7). It is very unlikely that young TD children anticipated the target interpretation when no prosodic signals were presented yet, and, thus, we hypothesise that this effect could be attributable to an exploration of the alternatives to the baseline interpretation prior to the unfolding of other disambiguating signals.

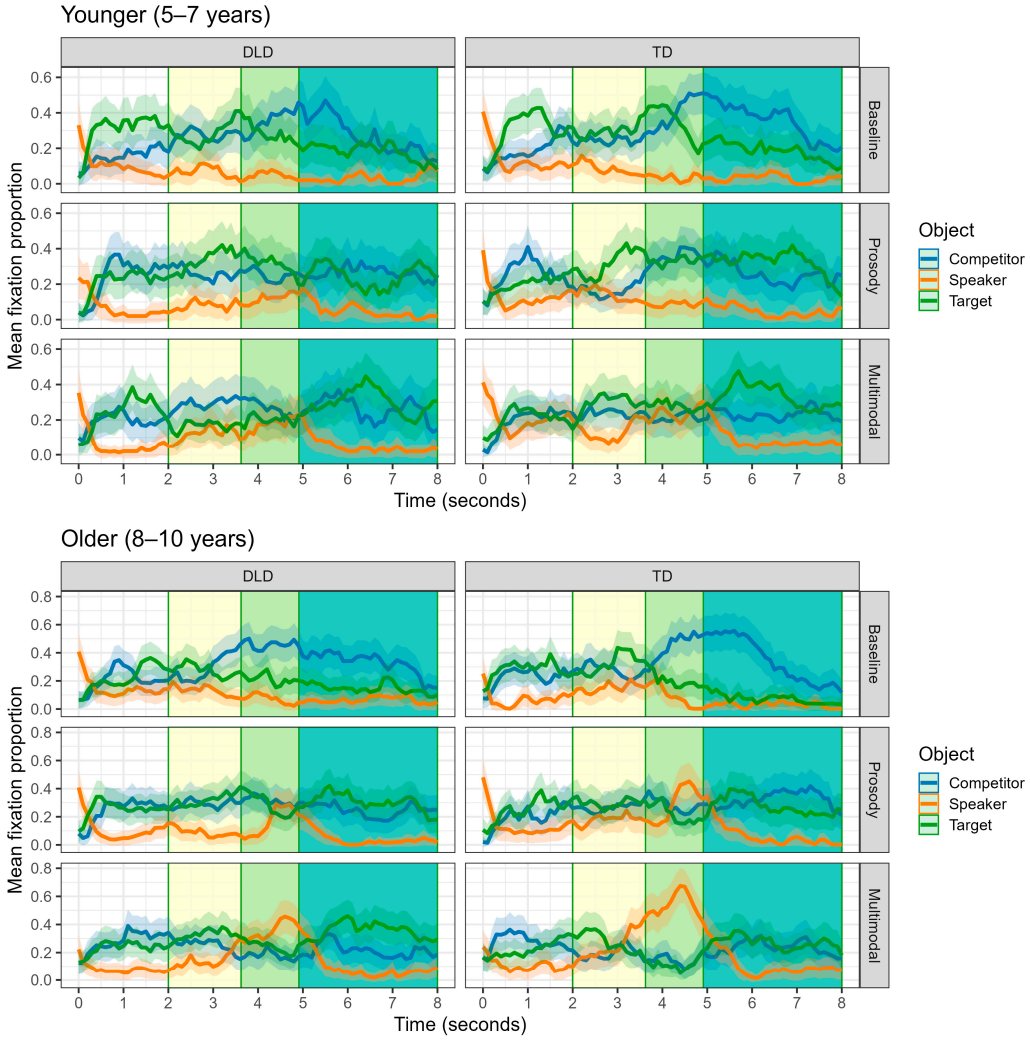

Figure 4. Mean fixation proportion of fixations to target, competitor and speaker across conditions and linguistic groups by the younger (**top** panel) and older (**bottom** panel) groups. The 0 in the X axis represents the start of the trial. The three coloured areas represent the time window 1 (yellow area), time window 2 (green area) and time window 3 (blue area).

From the presentation of prosodic and multimodal signal cues to the end of the target sentence (time window 2), older children with and without DLD looked at the competitor more in the baseline condition, as expected, showing a preference for the default low-attachment interpretation when no cues to the alternative interpretation unfolded (sum t TD = 73.2; sum t DLD = 83.7). Instead, in the prosody-only and multimodal conditions, this preference for the competitor was reduced, and children's gaze preferences were divided (or alternated) between the target and competitor, as no significant difference between these two objects was found (sum t = 0 for all subgroups).

From the end of the target sentence to 8 s from the start of the stimuli (time window 3), we found that the baseline condition elicited a preference for the competitor for both the older (sum t TD = 244.0; sum t DLD = 125.1) and younger (sum t TD = 86.6; sum t DLD = 34.8) groups. Additionally, in the multimodal condition, the younger children (both TD and DLD) and the older DLD group showed a preference for the target image (sum t TD = 36.5; sum t DLD = 26.5 for younger, and sum t = 57.3 for older DLD). No such effect of multimodality was observed in older TD children (sum t = 0).

Altogether, this comparison showed that children, especially in the case of the older groups, preferred the competitor image (a default low-attachment interpretation) in the baseline condition when no prosodic or multimodal signals of an alternative interpretation were presented yet. As soon as children perceived relevant prosodic and multimodal signals in the corresponding conditions, their preference for the default interpretation observed in the baseline condition decreased in the prosody and multimodal conditions, with all children dividing their attention between the target and competitor. After the target sentence (time window 3), all children kept looking at the competitor in the baseline condition, kept dividing their attention between target and competitor in the prosody-only condition, but instead shifted their gaze towards the target image when multimodal signals unfolded. Crucially, the processing effect of the presence of multimodal signals was especially significant in younger children and older children with DLD.

3.2.2. Gaze Preferences to Target vs. Speaker

This comparison does not directly derive from the research questions, but it was important due to the innovative nature of the methodological paradigm, which is inspired by Silverman et al. (2010) and represented a modification of the classic visual-world paradigm (Tanenhaus et al., 1995). Contrary to the classical visual-world paradigm, which requires the processing of audio stimuli together with a visual display of images representing (or not) the audio stimuli, in our case, children had to process both the visual display with response images and also a speaker located at the centre of the screen that utters the linguistic input. By comparing the proportion of looks to the target image and the speaker across groups and conditions, we are able to investigate how the speaker's presence influenced the children's processing of the stimuli.

Before the presentation of prosodic and multimodal signals (time window 1), children barely looked at the speaker, even though the sentence was already being produced. Instead, in this region, there was a preference for the target image, compared to the speaker, that was significant in the case of the younger subgroups (sum t TD = 90.4; sum t DLD = 170.5).

When prosodic and gestural signals unfolded (time window 2), older TD children then looked significantly more at the speaker than at the target in the multimodal condition (sum t = 149.8). That was not the case of the baseline and prosody-only conditions, where most children divided their looks between these two objects (except for younger TD children, who still fixated more on the target; sum t = 69.2, sum t = 96.0, respectively).

Finally, after the end of the sentence (time window 3), the looking patterns of older TD children shifted again from the speaker to the target image in the multimodal condition

(sum t = 38.2). The target image was also the preferred one in this time window by the other children (younger and DLD) in the prosody-only condition, maintaining the gaze preference observed in the previous region. The full results from the cluster analysis can be found in the OSF link.

In sum, this comparison showed that the presence of the speaker at the centre of the screen attracted the children's attention as soon as multimodal cues were presented but not before or after. This effect was especially strong in the older TD children. This suggests that the presence of a speaker producing hand gestures might temporarily attract the (older TD) children's looks towards that object. However, the fact that after the end of the sentence, children's looks were again distributed between the target and competitor images suggests that it did not fundamentally impact the processing of the utterance and, thus, that the adaptation of the visual-world paradigm was valid to assess the children's processing and comprehension of the audio-visually presented linguistic input.

4. Discussion

Sentences can be ambiguous at the phrasal level. In a sentence in Catalan, such as *L'home parla amb el nen malalt* (literally, 'The man talks to the boy sick'), listeners may derive a low-attachment interpretation (i.e., 'the boy was sick'), which is the preferred default interpretation in English and in Catalan (Felser et al., 2003; Grillo et al., 2015), or instead a high-attachment interpretation ('the man was sick'). Our study investigated whether prosodic and multimodal cues help children with DLD in processing phrasally ambiguous sentences, compared to TD children. In a visual-world eye-tracking experiment, younger (5 to 7 years old) and older (8 to 10 years old) children had to select the image representing the two possible interpretations of ambiguous sentences that could have a low- or high-attachment interpretation, while their gaze preferences were recorded and offline responses were scored. Sentences were presented in three different conditions: baseline, prosody-only cues, and multimodal cues. The analyses of the offline selection responses showed that, when sentences are presented orally and prosody or multimodality cues signal the phrasal structure, children with and without DLD are able to modify their initial default preference for a low-attachment interpretation and then choose a less frequent high-attachment meaning.

More specifically, we saw that (1) younger children were as capable as the older group of resolving the ambiguities through prosodic and multimodal cues; that (2) children with DLD, similar to their TD peers, can use prosody and multimodal cues to interpret the phrasal structure of phrasally ambiguous sentences; and that (3) prosodic cues are actually enough to indicate the phrasal structure, since gestures in the multimodal condition do not provide a significant advantage over stimuli that present only prosodic cues. Online results on the children's gaze preferences nuanced this last finding by showing that, while multimodal cues did not increase the children's accuracy in offline target selection, they did induce younger children and children with DLD to look at the target more once speech-accompanying gestures unfolded. Additional results showed that the presence of gestures accompanying speech may temporarily increase the attention to the speaker during the processing of sentences, as we observed in older TD children, but that this effect soon disappears and does not impact on children's sentence processing. These results are an important contribution in three main debates: (1) the discussion about the age in which children develop the ability to process phrasal prosody, (2) the discussion on whether children with DLD can reliably use prosodic cues (which are phonological in nature) in language comprehension, and (3) the discussion on when (and in which linguistic contexts) children with and without DLD use co-speech gestures for language comprehension.

First, our study revealed that younger children (5–7 years of age) performed comparably to the older (8- to 10-year-old) children in interpreting phrasally ambiguous sentences through prosody. This indicates that the ability to use prosodic cues for the resolution of sentence ambiguities emerges at least as early as 5–6 years of age. Previous studies reported mixed findings regarding the age at which children develop the ability to use prosodic cues for the interpretation of phrasally ambiguous sentences. Our results align with studies by Snedeker and Yuan (2008) and Wiedmann and Winkler (2015), which report that children are capable of using prosody to parse sentences at the age of 4 and 6, respectively, and with neurophysiological evidence (Männel et al., 2013) showing that the neural correlates for intonational phrase boundary detection are already operational by the age of 6. However, our results stand in contrast to previous evidence proposing that children aged 3 to 6 years have not yet developed the ability to detect certain prosodic cues for sentence comprehension (Choi & Mazuka, 2003; Vogel & Raimy, 2002). These differences might be explained as methodological reasons, as, for instance, in Vogel and Raimy's (2002) study, the authors examined compound (e.g., *hotdog*) versus phrasal stress (e.g., *hot dog*), reflecting lexical rather than syntactic processing. Overall, while the proficiency in using prosody for phrasal structures is still developing during the early school-age years, our findings suggest that children possess a great capacity for using prosody in understanding phrasal ambiguities already at five years of age.

Second, our results indicate that children with DLD are capable of using prosodic cues to comprehend phrasal structures. We initially predicted that children with DLD would rely on prosodic cues to a lesser extent than their TD peers due to these documented difficulties, and, instead, our results revealed that children with DLD benefited from enhanced prosodic cues as much as their TD peers. While we believe that these results are important and show that children with DLD can benefit from prosodic information, we do not want to imply that the phonological or prosodic dimensions of language are completely unaffected in this population. Extensive research has shown that children with DLD can exhibit deficits in these domains (Marshall et al., 2009; Calet et al., 2021), which can influence their language development. Our findings suggest that, despite these challenges, children with DLD can successfully use prosody to aid sentence interpretation in the context of phrasal ambiguities. Our findings also highlight the role of age in the ability of children with DLD to use prosodic cues for phrasal disambiguation, underscoring that the ability to effectively integrate these cues emerges as early as five years of age. This finding extends previous research by providing evidence that, despite their phonological deficits, children with DLD at a younger age already show sensitivity to prosody for syntactic parsing. It is important to note that, in our study, the prosodic (and multimodal) cues provided were strongly marked, providing a clear highlight of the potential alternative phrasal meaning. It could be that children with DLD can use prosodic cues to sentence comprehension when these prosodic cues are highly marked, like the ones we used in our task. Future research should clarify whether children with DLD require such highly explicit prosodic cues to consistently benefit from prosodic information.

Our results are, thus, more similar to those by Caccia and Lorusso (2019), who found that children with DLD (aged 10–13) were able to use prosody effectively for syntactic parsing when dealing with phrasal ambiguities. Relatedly, Sabisch et al. (2009) found that children with DLD struggled with sentence comprehension but that these difficulties were not due to an incapability of processing prosody but rather due to the syntactic complexity of sentences, such as passive voice. Our findings suggest that prosodic cues that guide phrase structure can indeed help children with DLD in sentence comprehension. In typical language development, prosodic cues are found to help segment linguistic input into phrasal groups, making sentence parsing more efficient (Fromont et al., 2017;

Stack & Watson, 2023). In the context of DLD, where syntactic processing is often compromised (Leonard, 2014), exaggerated prosodic cues might ease children's processing of syntactic information and, therefore, serve as additional cues to scaffold comprehension, even if phonological deficits are present. Future research should further explore how varying levels of sentence complexity affect the comprehension abilities of children with DLD, to better understand the conditions under which children with DLD can effectively use prosodic information.

Third, we found a limited effect of gestures on children's phrasal disambiguation. While our initial predictions were that children with DLD would have difficulties interpreting sentences based solely on prosodic information, and that multimodality would compensate for these difficulties, prosodic information seems to be highly relevant for the children's comprehension of phrasal structures, leaving little room for the impact of gestures for that matter. While previous studies have shown that children with DLD use more gestures in language production to compensate for their communicative deficits with oral language (Lavelli & Majorano, 2016; Wray et al., 2017), in comprehension, the impact of gestures accompanying speech is less clear. In fact, previous results on how children with DLD use body gestures for the offline comprehension and online processing of pragmatic meanings suggest that the contribution of gestures may depend on the complexity of the meaning to be understood and on a trade-off between the informativeness of gestures and of prosodic cues (Giberga et al., 2024). The trade-off relation between the informativity of prosody and gestural cues might also be what explains why our results differ from the ones in De La Cruz-Pavía et al. (2020). While they concluded that adults benefited from the presence of gestures to parse linguistic input into phrase-like units in artificial languages, these benefits were observed mainly in the condition in which only visual cues were available, and small differences were found when prosody and gestures were both available. In addition, they used an artificial language, which may have maximised the listener's use of acoustic and gestural input, compared to in our study, where children also had semantic and contextual information to aid in sentence interpretation, thus potentially reducing the reliance on gestural information.

While gestural cues did not make children more accurately choose the high-attachment interpretation, the analysis of the online gaze patterns showed that some children (especially at a younger age and when they have DLD) indeed processed these multimodal cues to process phrasal ambiguities. In our experiment, the inputs without prosodic or gestural cues elicited (as we expected) looks towards the low-attachment interpretation once these cues unfolded, while the prosody-only condition made the looking patterns distributed between both interpretations. Notably, the multimodal condition led to a significant increase in looks toward the high-attachment interpretation in all the subgroups, except for the older TD children. Similar to Holle et al.'s (2012) findings in adults, this pattern could be framed within a *multimodal bootstrapping* perspective, whereby co-speech gestures may lead to an increase in the children's attention towards a referent and, therefore, appear to boost the children's interpretation of phrasal ambiguities. The differences in gaze patterns during stimulus processing did not affect the children's offline comprehension, a finding consistent with previous studies using eye-tracking methodology in children with DLD (Giberga et al., 2024; Andreu et al., 2011) and also in adults (Esteve-Gibert et al., 2020). Another interesting observation in our results is that younger TD children looked at the competitor more already in time window 1 in the baseline condition. Although our stimuli were not designed to test the role of prenuclear prosodic cues for sentence processing, this suggests that children might be able to anticipate sentence interpretations based on the prosodic structure of the prenuclear material. Previous research has shown that early prosodic phrasing and pre-boundary lengthening can serve as cues for upcoming syntactic

structures (Christophe et al., 2008; Petrone & Niebuhr, 2014), and future studies could control for these prosodic cues to better understand their role in guiding children's early sentence processing and whether children with DLD can also benefit from these early cues.

In our adapted version of the visual-world paradigm (that was inspired by Silverman et al., 2010), children effectively processed prosodic and multimodal cues while distributing attention between the speaker and the response images, indicating a successful integration of auditory and visual information. TD children temporarily fixated more on the speaker in the centre of the screen when gestural cues indicated the phrasal structure (but not when prosodic cues were presented). We believe that this temporal effect did not interfere with the auditory sentence processing because listeners redirected their gaze back to the response images soon after multimodal cues were presented. The fact that only TD children but not DLD children show this significant temporary effect might have different explanations. First, it is plausible that children with DLD concentrate their attention to prosodic information, which seems to be sufficient for them to shift interpretations, and, therefore, the visual processing of the stimulus in the multimodal condition is diminished. Alternatively, it is possible that children with DLD had difficulties with the processing of complex audio-visual information, as seen in previous research on visual processing (Smolak et al., 2020; Wright et al., 2000) and impaired oculomotor functioning (Bilkova et al., 2021; Kelly et al., 2013) and, therefore, reduced their looks to the speaker producing the gestural input. Further studies are needed to disentangle which of these two hypotheses can explain this effect.

Our study has limitations that should be acknowledged. First, the highly controlled experimental design used to manipulate and present prosodic and gestural cues precisely may have resulted in stimuli that were less naturalistic and ecologically valid. The nature and timing of gestures used in the stimuli were carefully controlled so that they aligned with the exact same prosodic cues. This controlled approach might have reduced their naturalness, limiting their generalizability to everyday communication scenarios. Second, the focus on a specific type of sentence ambiguity—attachment ambiguity—allowed us to investigate a well-defined linguistic phenomenon, but it also narrows the scope of our findings. Attachment ambiguities may not represent the full range of challenges encountered by children with DLD when processing ambiguous sentences. Additionally, the task's cognitive demands could have influenced performance, particularly given the known processing deficits associated with DLD (Miller et al., 2001; Joanisse & Seidenberg, 1998). The need to process auditory, visual, and linguistic information simultaneously may have placed a strain on participants' cognitive resources. Although our materials and procedure were designed to minimize lexical and syntactic complexity, individual differences in cognitive capacities—such as working memory, attention, or visuospatial processing—may have contributed to variability in task performance. Future studies should investigate the influence of these individual factors on children's ability to integrate prosodic and multimodal cues in sentence processing. Another potential concern relates to the multilingual background of the child participants, specifically, whether the high-attachment preference typically observed in Spanish could influence the low-attachment preference expected in Catalan. Indeed, previous studies using relative clauses found a high-attachment preference in Spanish (Cuetos & Mitchell, 1988; Gilboy et al., 1995; Hemforth et al., 2015), in contrast to Catalan's typical low-attachment interpretation. However, those studies used different sentence types (relative clauses) rather than the phrasally ambiguous sentences we tested. Furthermore, all participating children were schooled in Catalan, and their proficiency was carefully verified with educators and caregivers before inclusion in this study. Importantly, preliminary results from our reading task with bilingual adults living in the same region indicated that, despite their strong proficiency in both languages, they exhibited an 80%

preference for low attachment in Catalan, consistent with the expected default preference. This suggests that any potential cross-linguistic influence from Spanish was minimised by the sociolinguistic context of the region, where Catalan is understood by over 90% of the population according to institutional data. Nevertheless, future studies could systematically assess language dominance and directly compare attachment preferences across both languages, clarifying how bilingual experience might shape default parsing strategies.

Our study sheds light on how children with and without DLD use prosodic and multimodal cues to resolve phrasal ambiguities in sentence comprehension. Contrary to previous assumptions, we found that children with DLD can use prosodic cues similarly to their TD peers, suggesting that prosodic cues could help them overcome difficulties in structural language. This is especially relevant in light of previous findings that children with DLD might also have impairment at the phonological level. It seems that phonological cues at the suprasegmental level, i.e., prosody, might be less problematic and rather be an important additional element that facilitates sentence comprehension. While gestures did not provide a clear advantage over prosody alone, multimodal cues did influence children's attention, particularly in guiding them toward the speaker cues and the images resolving the ambiguities. These findings provide evidence that language comprehension in natural settings often involves integrating oral and visual cues, even for children with language disorders. Our results are significant for the field of language acquisition, as they suggest that prosody is not only crucial in early language acquisition but continues to play a significant role in later stages.

Supplementary Materials: The following supporting information can be downloaded at: https://osf.io/ks2h5/?view_only=d3f5fb999734407aa00b82d2ced7d9ed, accessed on 17 March 2025. All supplementary information has been linked to the OSF project.

Author Contributions: Conceptualization, A.G. and N.E.-G.; methodology, A.G., E.G., N.A., A.I., M.A. and N.E.-G.; formal analysis, A.G.; investigation, A.G.; resources, N.E.-G.; data curation, A.G.; writing—original draft preparation, A.G.; writing—review and editing, E.G., N.A., A.I., M.A. and N.E.-G.; supervision, N.E.-G.; project administration, N.E.-G.; funding acquisition, N.E.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by a grant PID2020-115385GA-I00 from the Ministerio de Ciencia e Innovación, the grant 2021SGR01102 from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) and the grant 2024AFB240004 from the Agencia Nacional de Investigacion y Desarrollo (ANID).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Universitat Oberta de Catalunya (protocol code 20201210_nesteveg_Prosody on 20 January 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original data presented in the study are openly available in the OSF project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Aguilar-Mediavilla, E. M., Sanz-Torrent, M., & Serra-Raventós, M. (2002). A comparative study of the phonology of pre-school children with specific language impairment (SLI), language delay (LD) and normal acquisition. *Clinical Linguistics & Phonetics*, *16*(8), 573–596. [CrossRef]

Andreu, L., Sanz-Torrent, M., Guàrdia Olmos, J., & Macwhinney, B. (2011). Narrative comprehension and production in children with SLI: An eye movement study. *Clinical Linguistics & Phonetics*, *25*(9), 767–783. [CrossRef]

Armstrong, M. E., Esteve Gibert, N., & Prieto Vives, P. (2014). The acquisition of multimodal cues to disbelief. In N. Campbell, D. Gibon, & D. Hirst (Eds.), *Speech prosody 2014, Dublin, Ireland, 20–23 May 2014* (pp. 1139–1143). International Speech Communication Association.

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General*, *143*(1), 404–413. [CrossRef] [PubMed]

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*, 1.

Biau, E., Fromont, L. A., & Soto-Faraco, S. (2018). Beat gestures and syntactic parsing: An ERP study. *Language Learning*, *68*(S1), 102–126. [CrossRef]

Bilkova, Z., Dobias, M., Dolezal, J., Fabian, V., Havlisova, H., Jost, J., & Malinovska, O. (2021). Eye tracking using nonverbal tasks could contribute to diagnostics of developmental dyslexia and developmental language disorder. In *Dyslexia* (p. 226). IntechOpen.

Bishop, D. V. M., & Adams, C. (1992). Comprehension problems in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *35*(1), 119–129. [CrossRef]

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of catalise: A multinational and multidisciplinary delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*, 1068–1080. [CrossRef]

Caccia, M., & Lorusso, M. L. (2019). When prosody meets syntax: The processing of the syntax-prosody interface in children with developmental dyslexia and developmental language disorder. *Lingua*, *224*, 16–33. [CrossRef]

Calet, N., Martín-Peregrina, M. Á., Jiménez-Fernández, G., & Martínez-Castilla, P. (2021). Prosodic skills of Spanish-speaking children with developmental language disorder. *International Journal of Language and Communication Disorders*, *56*(4), 784–796. [CrossRef]

Castilla-Earls, A., Bedore, L., Rojas, R., Fabiano-Smith, L., Pruitt-Lord, S., Restrepo, M. A., & Peña, E. (2020). Beyond scores: Using converging evidence to determine speech and language services eligibility for dual language learners. *American Journal of Speech-Language Pathology*, *29*(3), 1116–1132. [CrossRef]

Chen, A., Esteve-Gibert, N., Prieto, P., & Redford, M. A. (2020). Development of phrase-level prosody from infancy to late childhood. In C. Gussenhoven, & A. Chen (Eds.), *The oxford handbook of language prosody* (pp. 552–562). Oxford University Press.

Choi, Y., & Mazuka, R. (2003). Young children's use of prosody in sentence parsing. *Journal of Pyscholinguistic Research*, *32*(2), 197–217.

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech*, *51*(1&2), 61–75. [PubMed]

Christou, S., Guerra, E., Coloma, C. J., Barrachina, L. A., Araya, C., Rodriguez-Ferreiro, J., Pereda, M. J. B., & Sanz-Torrent, M. (2020). Real time comprehension of spanish articles in children with developmental language disorder: Empirical evidence from eye movements. *Journal of Communication Disorders*, *87*, 106027. [CrossRef]

Coloma, C. J., Guerra, E., De Barbieri, Z., & Helo, A. (2024). Article comprehension in monolingual spanish-speaking children with developmental language disorder: A longitudinal eye tracking study. *International Journal of Speech-Language Pathology*, *26*(1), 105–117. [CrossRef] [PubMed]

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, *30*(1), 73–105. [CrossRef]

De La Cruz-Pavía, I., Elordieta, G., Villegas, J., Gervain, J., & Laka, I. (2022). Segmental information drives adult bilingual phrase segmentation preference. *International Journal of Bilingual Education and Bilingualism*, *25*(2), 676–695. [CrossRef]

De La Cruz-Pavía, I., Gervain, J., Vatikiotis-Bateson, E., & Werker, J. F. (2019). Finding phrases: On the role of co-verbal facial information in learning word order in Infancy editado por I. Laka. *PLoS ONE*, *14*(11), e0224786. [CrossRef]

De La Cruz-Pavía, I., Werker, J. F., Vatikiotis-Bateson, E., & Gervain, J. (2020). Finding phrases: The interplay of word frequency, phrasal prosody and co-speech visual information in chunking speech by monolingual and bilingual adults. *Language and Speech*, *63*(2), 264–291. [CrossRef]

Esteve-Gibert, N., & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication*, *57*, 301–316. [CrossRef]

Esteve-Gibert, N., Schafer, A. J., Hemforth, B., Portes, C., Pozniak, C., & D'Imperio, M. (2020). Empathy influences how listeners interpret intonation and meaning when words are ambiguous. *Memory and Cognition*, *48*(4), 566–580. [CrossRef]

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*(3), 453–489. [CrossRef]

Fernández, E. M., & Sekerina, I. A. (2015). The interplay of visual and prosodic information in the attachment preferences of semantically shallow relative clauses. In L. Frazier, & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in theoretical psycholinguistics* (Vol. 46, pp. 241–261). Springer International Publishing.

Fromont, L. A., Soto-Faraco, S., & Biau, E. (2017). Searching High and Low: Prosodic Breaks Disambiguate Relative Clauses. *Frontiers in Psychology*, *8*, 96. [CrossRef]

Giberga, A., Igualada, A., Ahufinger, N., Aguilera, M., Guerra, E., & Esteve-Gibert, N. (2024, July 2–5). *Prosody and gesture in the comprehension of pragmatic meanings: The case of children with developmental language disorder*. Speech Prosody 2024 (pp. 697–701), Leiden, The Netherlands. [CrossRef]

Gilboy, E., Sopena, J. M. M., Cliftrn, C., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, *54*(2), 131–167. [CrossRef]

Grillo, N., Costa, J., Fernandes, B., & Santi, A. (2015). Highs and Lows in English Attachment. *Cognition*, *144*, 116–122. [CrossRef]

Guellaï, B., Langus, A., & Nespor, M. (2014). Prosody in the hands of the speaker. *Frontiers in Psychology*, *5*, 700. [CrossRef]

Hagoort, P., & Özyürek, A. (2024). Extending the architecture of language from a multimodal perspective. *Topics in Cognitive Science*, 1–11. [CrossRef]

Helo, A., Guerra, E., Coloma, C. J., Reyes, M. A., & Räma, P. (2021). Objects shape activation during spoken word recognition in preschoolers with typical and atypical language development: An eye-tracking study. *Language Learning and Development*, *18*(3), 324–351. [CrossRef]

Hemforth, B., Fernandez, S., Clifton, C., Frazier, L., Konieczny, L., & Walter, M. (2015). Relative clause attachment in German, English, Spanish and French: Effects of position and length. *Lingua*, *166*, 43–64. [CrossRef]

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., & Gunter, T. C. (2012). Gesture Facilitates the Syntactic Analysis of Speech. *Frontiers in Psychology*, *3*, 74. [CrossRef]

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. [CrossRef] [PubMed]

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, *76*(3), 598–613. [CrossRef] [PubMed]

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language*, *37*(1), 24–41. [CrossRef]

iMotions. (2023). *iMotions Biometric Research Platform* (Version 9.1) [Computer software]. iMotions A/S.

Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: A deficit in grammar or processing? *Trends in Cognitive Sciences*, *2*(7), 240–247. [CrossRef]

Jun, S. A., & Bishop, J. (2014). Implicit prosodic priming and autistic traits in relative clause attachment. *Proceedings of the International Conference on Speech Prosody*, *7*, 854–858. [CrossRef]

Kaufman, A. S. (1990). *Kaufman brief intelligence test: KBIT*. American Guidance Service Circle Pines (AGS).

Kelly, D. J., Walker, R., & Norbury, C. F. (2013). Deficits in volitional oculomotor control align with language status in autism spectrum disorders. *Developmental Science*, *16*(1), 56–66. [CrossRef]

Kirk, E., Pine, K. J., & Ryder, N. (2010). I hear what you say but I see what you mean: The role of gestures in children's pragmatic comprehension. *Language and Cognitive Processes*, *26*(2), 149–170. [CrossRef]

Lavelli, M., & Majorano, M. (2016). Spontaneous gesture production and lexical abilities in children with specific language impairment in a naming task. *Journal of Speech, Language, and Hearing Research*, *59*(4), 784–796. [CrossRef]

Leonard, L. B. (1998). *Children with specific language impairment* (2nd ed.). MIT Press.

Leonard, L. B. (2014). Specific language impairment across languages. *Child Development Perspectives*, *8*(1), 1–5. [CrossRef]

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, *3*(1), 71–89. [CrossRef]

Lüke, C., Ritterfeld, U., Grimminger, A., Rohlfing, K. J., & Liszkowski, U. (2020). Integrated communication system: Gesture and language acquisition in typically developing children and children with LD and DLD. *Frontiers in Psychology*, *11*, 118. [CrossRef]

Maillart, C., & Parisse, C. (2006). Phonological deficits in French speaking children with SLI. *International Journal of Language & Communication Disorders*, *41*(3), 253–274. [CrossRef]

Marshall, C. R., Harcourt-Brown, S., Ramus, F., & Van Der Lely, H. K. J. (2009). The link between prosody and language skills in children with specific language impairment (sli) and/or dyslexia. *International Journal of Language & Communication Disorders*, *44*(4), 466–488. [CrossRef]

Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience*, *5*, 86–94. [CrossRef] [PubMed]

Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of Processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *44*(2), 416–433. [CrossRef]

Mitchel, A. D., & Weiss, D. J. (2014). Visual speech segmentation: Using facial cues to locate word boundaries in continuous speech. *Language, Cognition and Neuroscience*, *29*(7), 771–780. [CrossRef]

Modyanova, N. N., Bolton, A. P., Storrusten, C., & McCrory, B. (2024). Improving language comprehension via hand gestures in children with autism spectrum disorders and/or language impairment in rural montana. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, *13*(1), 191–196. [CrossRef]

Montgomery, J. W., Evans, J., Fargo, J., Schwartz, S., & Gillam, R. B. (2018). Structural relationship between cognitive processing and syntactic sentence comprehension in children with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *61*(12), 2950–2976. [CrossRef]

Nagel, H. N., Shapiro, L. P., Tuller, B., & Nawy, R. (1996). Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 319–344. [CrossRef]

Petrone, C., & Niebuhr, O. (2014). On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech*, *57*(1), 108–146. [CrossRef]

Plym, J., Lahti-Nuuttila, P., Smolander, S., Arkkila, E., & Laasonen, M. (2021). Structure of cognitive functions in monolingual preschool children with typical development and children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *64*, 3140–3158. [CrossRef] [PubMed]

Prieto, P. (1997). Prosodic manifestation of syntactic structure in Catalan. In F. Martínez-Gil, & A. Morales-Front (Eds.), *Issues in the phonology of the Iberian languages* (pp. 179–199). Georgetown University Press.

Prieto, P., & Esteve-Gibert, N. (Eds.). (2018). *The development of prosody in first language acquisition* (Vol. 23). John Benjamins Publishing Company.

Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, *11*(1–2), 165–192. [CrossRef]

Rohrer, P. L., Delais-Roussarie, E., & Prieto, P. (2023). Visualizing Prosodic Structure: Manual Gestures as Highlighters of Prosodic Heads and Edges in English Academic Discourses. *Lingua*, *293*, 103583. [CrossRef]

Rowe, M. L., Wei, R., & Salo, V. C. (2022). Early gesture predicts later language development. In *Gesture in language: Development across the lifespan, language and the human lifespan* (pp. 93–111). American Psychological Association.

Sabisch, B., Hahne, C. A., Glass, E., Von Suchodoletz, W., & Friederici, A. D. (2009). Children with specific language impairment: The role of prosodic processes in explaining difficulties in processing syntactic information. *Brain Research*, *1261*, 37–44. [CrossRef]

Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*, *9*, 1514. [CrossRef]

Silverman, L. B., Bennetto, L., Campana, E., & Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition*, *115*(3), 380–393. [CrossRef]

Smolak, E., McGregor, K. K., Arbisi-Kelm, T., & Eden, N. (2020). Sustained attention in developmental language disorder and its relation to working memory and language. *Journal of Speech, Language, and Hearing Research*, *63*(12), 4096–4108. [CrossRef]

Snedeker, J., & Trueswell, J. (2001). Prosodic guidance: Evidence for the early use of capricious parsing constraint. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *23*(23), 2.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299. [CrossRef]

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, *58*(2), 574–608. [CrossRef]

Stack, C. M. H., & Watson, D. G. (2023). Pauses and Parsing: Testing the Role of Prosodic Chunking in Sentence Processing. *Languages*, *8*(3), 157. [CrossRef]

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. [CrossRef] [PubMed]

Theakston, A. L., Coates, A., & Holler, J. (2014). Handling agents and patients: Representational Cospeech gestures help children comprehend complex syntactic constructions. *Developmental Psychology*, *50*(7), 1973–1984. [CrossRef] [PubMed]

Van Der Meulen, S., Janssen, P., & Den Os, E. (1997). Prosodic abilities in children with specific language impairment. *Journal of Communication Disorders*, *30*(3), 155–170. [CrossRef]

Vogel, I., & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: The role of prosodic constituents. *Journal of Child Language*, *29*(2), 225–250. [CrossRef]

Vogt, S. S., & Kauschke, C. (2017). With some help from others' hands: Iconic gesture helps semantic learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3213–3225. [CrossRef]

Vukovic, M., Jovanovska, M., & Jerkic Rajic, L. (2022). Phonological awareness in children with developmental language disorder. *Archives of Public Health*, *14*, 1–10. [CrossRef]

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*(2), B63–B72. [CrossRef]

Wiedmann, N., & Winkler, S. (2015). The influence of prosody on children's processing of ambiguous sentences. In S. Winkler (Ed.), *Ambiguity: Language and Communication* (pp. 185–197). Walter de Gruyter. [CrossRef]

Wiig, E. H., Semel, E. M., & Secord, W. (2013). *CELF-5: Screening test*. Pearson/PsychCorp.

Wray, C., Saunders, N., McGuire, R., Cousins, G., & Norbury, C. F. (2017). Gesture production in language impairment: It's quality, not quantity, that matters. *Journal of Speech, Language, and Hearing Research*, *60*(4), 969–982. [CrossRef]

Wright, B. A., Bowen, R. W., & Zecker, S. G. (2000). Nonlinguistic perceptual deficits associated with reading and language disorders. *Current Opinion in Neurobiology*, *10*(4), 482–486. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.