# Probing deaf oral linguistic competence with minimal morphosyntactic pairs
## The COnVERSA test

Cristiano Chesi°, Giorgia Ghersi*, Valentina Musella•,
Debora Musola•
°IUSS Pavia5, *University of Pisa, •Cooperativa Logogenia

The COnVERSA test (Test di Comprensione delle Opposizioni morfo-sintattiche VERbali attraverso la ScritturA) is a tool for grammatical competence assessment in Italian. It uses informal grammaticality (binary) judgments on written linguistic minimal pairs and it has been developed for testing deaf children, for which no suitable comprehension test was available. Here we discuss the results obtained both from a normally developing hearing children group (NC, 6-10 y.o.) and a deaf children population (DF, 7-16 y.o.). We will focus on children's minimal morphosyntactic pairs sensitivity in three domains: (i) argument structure, (ii) question formation, and (iii) pronominal usage. We will then compare these results with those obtained on agreement-based dependencies (Chesi et al., 2023). Our findings support a complexity metric, based on two structural dimensions (functional height and intervention), which is used to guide a dynamic administration modality, in the end significantly reducing the assessment time without decreasing the discriminative power of the test. Our data confirm the poor discriminative ability of DF (especially those without a cochlear implant) for configurations targeting the highest functional domains and involving intervention.

**Keywords**: deaf children, language acquisition, competence assessment, minimal pairs, argument structure, wh-dependencies, pronouns

## 1.    Introduction

Deaf children's (DF) grammatical sensitivity can hardly be assessed using standard comprehension tests usually tuned on the hearing children (NC) population: even in the case the acoustic gain after implantation is sufficient to capture linguistic signals of decent quality, deaf children often rely on lip-reading and compensatory strategies to understand their interlocutor during oral interac-

tion. In this context, standard comprehension-based tests, based on picture-matching tasks, require a huge effort for deaf children with an impact on their performance (Cardinaletti, 2018). Using standard tests developed for age-matching, sometimes pathological, populations to assess deaf children's linguistic competence is then ineffective: first, most of these tests strongly rely on the auditory modality (WISC IV, Wechsler, 2014), second, they target aspecific constructions (TCGB, Chilosi et al., 2006; PVCL, Rustioni & Lancaster, 2007) of little interest for the DF population, for which we have a specific description of the linguistic deficit (Chesi 2006; Franchi 2006; Volpato 2010 a.o.). Informal grammaticality judgments, on the other hand, can be effectively used to test children's minimal linguistic discriminative abilities in a simpler and more accessible way (Chesi et al., 2023).

The goal of this paper is twofold: a complete assessment tool (COnVERSA) based on grammaticality judgments is presented, which is suitable for deaf children from 6/7 y.o.; then, the norming data will be discussed, both including a deaf group of 54 (7-16 y.o.) and a group of hearing children (90 subjects, 6-10 y.o.). Three major characteristics of the tool are discussed here: (i) its simplicity and accessibility (a necessary condition for the target population), (ii) its repeatability (necessary for tracking precisely longitudinal progresses in children with ongoing logopedic support), and (iii) its specificity in targeting subtle morpho-syntactic phenomena, useful for the therapist/teacher to precisely map the linguistic competence of the child tested.

In the following sections, we will focus on specific deficiencies in standard tests (§0) and introduce the original ideas for a simpler and richer competence assessment (§0, §0). The linguistic background that inspired the relevant contrasts used in this test (§0) will be motivated by precise complexity considerations (§0) that will also guide a "dynamic" administration modality (§0). We will then present the results of the norming study, focusing on featural sensitivity in argumental structure, pronominal usage, and question formation (§3). The results will be discussed (§4) presenting the relevant contrasts obtained in both groups. A general discussion (§5) of these results as well as their coherence concerning the relevant acquisition literature will precede the conclusion about the impact of this assessment approach (§6).

## 2.    Linguistic background: testing implicit grammatical competence

### 2.1    On the inadequacy of standard linguistic tests

Describing the phenomena underlying language processing is not trivial, given the complexity of the variables involved and the difficulty in the identification of strategies adopted by children during the comprehension process. Currently, the main standard linguistic tests for the assessment of morphosyntactic comprehension in Italian are TCGB, (Chilosi et al., 2006), PVCL (Rustioni & Lancaster, 2007), TROG – 2 (Bishop, 2009), WISC-IV - VCI (Wechsler, 2014) and COMPRENDO (Cecchetto et al., 2012). These tests aim at assessing subjects' comprehension of different morphosyntactic structures mainly through picture-matching tasks. TCGB and PVCL are specifically designed for children, TROG-2 is suitable for children aged 4 and over, WISC-IV is for children between 6 and 16 y.o. while COMPRENDO is aimed at adults aged between 20 and 80.

Despite their efficacy and robustness, these standard linguistic tests show some criticalities: First, these tests, except for COMPRENDO, can be administered only by healthcare professionals (speech therapists, psychologists), preventing other professional categories (e.g., teachers and linguists) to collect data using these tools. This is a necessity since only fully controlled administration environments and trained personnel guarantee reliable results.

Secondly, these tests are not specifically designed for tracking grammatical deficiencies and discourage compensatory strategies typical of deaf children (e.g. tendency to consider only lexical information neglecting the functional elements, linear proximity, and world knowledge).

Finally, important morphosyntactic areas (such as the person morphology, the use of determiners, clitics, agreement, and dislocation phenomena) are not sufficiently addressed in these tests.

In the absence of specific tools, TCGB is often used, although it has been reported that the test contains inadequate pictures: occasionally, pictures are not really distractors for the subject and the lack of fictional images depicting unlikely situations (such as *The girl is eaten by the apple*) suggests that the subjects could be guided in their choice by their world knowledge (Bertone et al., 2011). This introduces a critical bias in the assessment.

### 2.2    Grammaticality judgments as a precursor of comprehension

All these linguistic tests target the comprehension ability to infer the level of linguistic competence. Although each experimental item is usually controlled for pragmatic biases, comprehension is a complex and time-demanding task involv-

ing not only lexical and grammatical knowledge but also focalized attention (crucial for interpreting the relevant details in any picture) and a sufficient working memory span.

Grammaticality, according to Chomsky (1957), is a more primitive and fundamental property associated with our implicit linguistic competence: any native speaker is readily able to reject clearly ill-formed expressions (e.g. *runs boy the*). Even though this notion is a purely abstract one, we can probe such a "perception of structural incongruency" by informally asking "can you say this or not?" (Schütze, 2016). To avoid bias in testing children (McDaniel et al., 1996), we developed a solid experimental set-up to elicit natural answers also in metalinguistic terms (Gordon 1996) by concentrating on solid contrasts that do not leave much room for interpretation in adult grammar (e.g. *The child *play/plays*). Among various approaches to assessing acceptability/grammaticality, opting for a binary judgment task ("ok/grammatical" vs. "wrong/ungrammatical") has a series of advantages: (i) it is simple and quick, (ii) we can consider the higher acceptability observed in pathological populations for ungrammatical sentences as a signature of non-standard performance.

## 2.3    Using minimal pairs contrasts

One prominent idea proposed to explain the impairment in oral language acquisition for deaf children relies on the fact that the primary linguistic input that these children receive is impoverished both in quality and quantity, hence it is insufficient to trigger the natural language acquisition device any normally developing child uses for acquiring the first language (Chomsky, 1981): The linguistic input is impoverished in quality since the perceptual ability in phonemes discrimination is low or absent, and in quantity, since visual access to the linguistic input is not as immersive as the oral channel. Moreover, reading ability is a prerequisite for the visual gathering of linguistic information through writing, hence this modality is hardly available in deaf children younger than 6 years. Nevertheless, Bruna Radelli (Radelli, 1999) discussed a simple method that seems to support the natural tendency to acquire a language in deaf children during the critical period (Curtiss, 1978) and possibly later: In her seminal work, she presented preliminary evidence in favor of the improvement of deaf children in their mastering of subtle morphosyntactic oppositions if they get prompted with the relevant contrasts in a minimal pair format. For instance, the role of a prepositional marker such as "to" used to mark the indirect object in dative constructions will be favored by prompting the child with minimal oppositions such as *John gives a book Mary* vs *John gives a book to Mary*, explicitly marking

the ungrammaticality of one sentence in the minimal pair. Neither grammatical explanation nor metalinguistic reasoning is necessary: simple minimal pairs in form of written declarative sentences, questions to be answered, or commands to be executed, seem to constitute a sufficient acquisition trigger. This approach, dubbed "Logogenia", is now widely used as a linguistic support practice for deaf children (Franchi & Musola, 2011, 2012). The absence of a systematic, complete, reliable, and quick assessment before and after a specific linguistic logopedic or logogenic treatment justifies the test presented here.

## 2.4    On structural complexity

Some sentences are simpler than others and this difference can be measured in various ways by relying on explicit complexity metrics. Our aim is to use one such metric to guide our testing procedure under the reasonable assumption that if a child performs well on an item of complexity $c$, he/she should perform equally well on any item of complexity lower than $c$. In this chapter, we explore this option by defining a simple complexity metric.

### 2.4.1 *Functional projections*

The linguistic framework adopted in this study is the generative one (Chomsky, 1995). More precisely, we assume a rich representation of the functional information (namely the structural scaffolding of the lexical content in each sentence) following the so-called "cartographic approach" (Belletti, 2004; Cinque, 1999, 2002; Rizzi, 2004): A universal hierarchy of functional projections is postulated and it is assumed to be mastered incrementally starting from the first functional specification attached to the lexical kernel up to the higher functional layers (Friedmann et al., 2021).

Two simple examples are provided to explain this idea:

(1)    Peter$_j$ [$_{IP}$ eatsi [vP _$_j$ _$_i$ [VP _$_i$ [a candy]]]]

(2)    What$_k$ [$_{CP}$ does$_j$ [Peter$_k$ [$_{IP}$ _$_i$ [vP _$_j$ eat$_i$ [vP _$_i$ [ _$_k$]]]]]]?

Examples (1) and (1) illustrate a clear structural contrast: Right above the Verbal Phrasal (VP) thematic shells (VP hosts the internal argument, the direct object; vP hosts the external argument, the subject, cf. Hale & Keyser 1993) the first functional specification we find is an Inflectional Phrase, IP (possibly split into different agreement projections AgrS(ubject), T(ense) AgrO(bject), Belletti 2017, see Figure 1 in §2.5). According to the hierarchical distribution of these

functional projections, we expect the child to go through a phase in which inflection drop is a real option and *wh-* argumental movement targeting the complementizer phase (CP) as sketched in (1) is not yet available. This is known as the "root infinitives" phase (Hoekstra & Hyams, 1998). Also the unavailability of the left peripheral CP, hence the impossibility of productively asking *wh-* questions is attested until age 3-4 (Haegeman, 1996). A recent cross-linguistic systematic research supports this idea that hierarchical functional structures essentially develop from bottom to top (Friedmann et al., 2021). We adopt this intuition here and we assign an increasing complexity cost to each relevant field, starting from the basic argumental shell to the highest functional expression, that is, other things being equal, an operation involving a higher functional layer is predicted to be more complex than one targeting a lower position.

### 2.4.2 *Locality*

Functional height, per se, is not the sole source of complexity in phrase structure: to cope both with thematic role assignment and with discourse properties, an argument is necessarily displaced from one position into another. This is what happens in English (and Italian), for instance, in argumental *wh-* questions formation presented in (1) above. It is experimentally well attested that the longer a (filler-gap) dependency, the harder it is to process it (Lewis & Vasishth, 2005). Notice that the length is not just a matter of bare distance: a relevant factor is represented by the number and quality of the elements intervening between the displaced element and its base position. In example (1), *what* is moved to a left peripheral CP position to check interrogative features and is moved from the position next to "eat" where it received the appropriate thematic role (direct object). On the way to the left peripheral position, *what* crosses another argument (a Determiner Phrase, DP) *Peter*. The similarity, expressed in terms of relevant features, is what matter as a predictor of complexity (Gordon, Hendrick & Johnson 2004): the more you share, the more difficult the dependency. Friedmann et al. (2009) suggest that the relevant features counting as "interveners" are those triggering displacement. If a *+wh* feature (allegedly responsible for the displacement of *what* to the left periphery of the sentence) would have been present in the subject argument, this would have led to ungrammaticality (\**what_i did who ask \_i a favor?*); since the subject is "different" in the relevant sense (no *+wh* features on it), the dependency is possible, but if subject and object would have shared a similar "lexical restriction" (in Friedmann et al.'s terms), the sentence would have been more difficult (e.g. *This is [the lawyer]_i that [the banker] knows \_i?*). Here we adopt this intuition (cf. Chesi & Canal 2019).

### 2.4.3 *Factor interaction*

Functional height and distance interact in a direct way. One simple phenomenon that makes this hypothesis fully explicit is agreement. Agreement in Italian comes in (at least) three flavors (Moscati & Rizzi, 2014): the simplest case is the local relation determiner-noun (D-N) "the.F.SG child.F.SG" in (3).a (classically defined as specifier-head agreement), then a less local dependency involving agreement and displacement is subject-verb (S-V) agreement, again in (3).a (the subject appears in an IP position hence it results displaced with respect to the thematic VP position). The hardest case is the non-local dependency involving an extra movement operation, plus a morphological reduction operation, namely cliticization, triggering past-participle agreement (Obj.CL-PstPrt), (3).b.

(3)     a. [IP [la bambina]i       [TP ha [VP _i mangiato [i gelati]]]]
            the.F.SG child.F.SG       has.3.SG eaten.M.SG [the ice-creams].M.PL
        b. [IP [la bambina]i      lij                 [TP ha [VP _i mangiati [_j]]]]
            the.F.SG child.F.SG them.CL.M.PL       has.3.SG  eaten.M.PL
            *the child has eaten them*

These cases constitute a natural scale of complexity, starting from the allegedly simpler one (D-N) to the most complex case (Obj.CL-PstPrt). We expected deaf children to perform better on D-N than on S-V and better on S-V than on Obj.CL-PstPrt. Both preliminary evidence (Chesi, 2006) and more advanced testing of more minimal contrasts (Chesi et al., 2023) go in this direction: the combination of hierarchical height and displacement assumptions predicts this scale: D-N (requires only one Merge operation) < S-V (requiring Merge + Move) < Obj.CL-PstPrt (requiring, at least, two Merge and two Move operations), where "<" means "being less complex than".

### 2.5     Phenomena tested with COnVERSA

Given the preliminary discussion on complexity and linguistic framework adopted, we will now introduce the morphosyntactic areas under scrutiny and the targeted dependencies. The functional "fields" (cluster of functional projections) considered here are CP, IP, and the verbal lexical domain (vP, VP). These fields are the morphosyntactic locus of specific phenomena and can be subdivided into various functional projections as indicated in the schematic diagrams in Figure 1. Notice that the first two stages identified in (Friedmann et al., 2021) correspond, roughly, to IP (stage 1), Q(uestion)/Foc(us)P (stage 2) which are respectively involved in cliticization (T head) and questions formation (Q head).

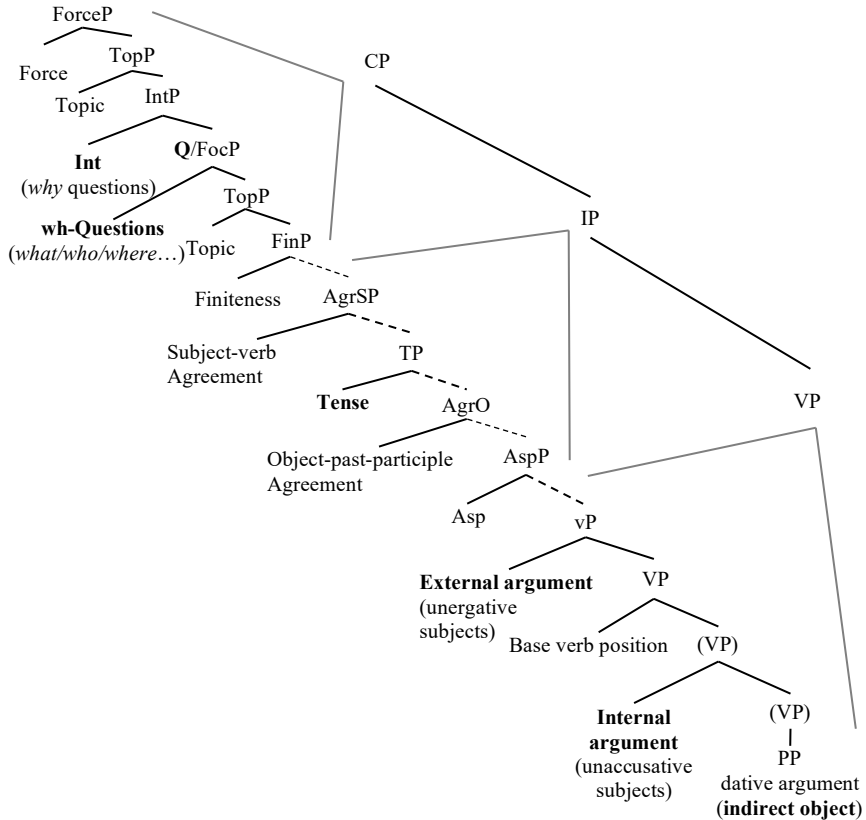*Why* questions are allegedly related to Int and predicted to be mastered only at stage 3.



Figure 1. Syntactic fields above VP and relevant exemplificative phenomena associated with these. The items in bold are targets of the current version of COnVERSA (the C-command relation between Internal and indirect argument is here simplified, see Belletti & Rizzi 1988 vs. Larson 1988).

The dependencies that involve the highlighted positions are thematic role assignment (VP argumental shells), auxiliary realization and cliticization (Tense), and questions formation (IntP/Q/FocP). Each dependency requires the activation of one or more fields, and it will be predicted to be less or more complex in accordance with the relative height of the functional levels activated and the locality of the dependency established. Notice that predicting a difference between constructions A and B does not imply that such a difference is detectable in per-

formance. However, if A is expected to be minimally simpler than B and B is expected to be minimally simpler than C, if we do not detect a difference between A and C, we expect to find no difference between A and B either. Moreover, if the difference between A and B is detectable in a mature population, we might expect this difference to remain undetectable in a younger or pathological population (Grillo, 2008).

### 2.5.1 *Agreement*

Chomsky defines Agreement as a dependency between a Probe and a Goal often triggering movement (Chomsky, 2001). Under the lens of the complexity metrics previously discussed, we refined Moscati & Rizzi's (2014) scale (D-N < S-V < Obj.CL-PstPrt) by adding minimal agreement contrasts, in the end, considering the following oppositions: Subject–adjectival predicate (S-AP), post-verbal subject–unaccusative verb (V-S), Subject–past participle, with unaccusative predicates (S-PstPrt), subject-verb with transitives (S-V) and cumulative subjects–verb (CumS-V) agreement, again with transitive predicates:

(4)   a.   I *giorno/giorni.                                      (D-N)
           The.PL *day/days
      b.   Il bambino                è malato/*malata.       (S-AP)
           [The child].M.SG          is sick.M.SG/*sick.F.SG
      c.   Arriva   la maestra/*le maestre.                    (V-S)
           Arrives [the teacher].SG /*[the teachers].PL
           The teacher arrives
      d.   [Il signore]        è entrato/*entrata in casa.  (S-PstPrt)
           [The man].M.SG   is entered.M.SG/*.F.SG in (the) house
           The man entered home
      e.   Il maestro corregge/*correggono i compiti.  (S-V)
           The teacher corrects/*correct the homework.PL
      f.   Io e il bambino mangiamo/*mangio il pane. (CumS-V)
           [I and [the child]]1P.PG eat.1P.PL /*.1P.SG the bread

Minimal variations on these items include the introduction of "attractors" in the S-AP and S-V condition (i.e. *The teacher [of the students] corrects/*correct the homework.PL*, Franck et al. 2006). The performance on these agreement configurations is predicted to be coherent with the following complexity scale:

(5)      D-N < S-AP < V-S < S-V < S-PstPrt < CumS-V

The linear regression model obtained by fitting this scale with children's performance is more accurate than coarser scales (obtained, for instance, by relying on corpus evidence or on processing considerations), but only for older children (8-10 y.o. NC). Younger children perform significantly less well and, more importantly, do not discriminate between different agreement configurations according to the fine-grained complexity scale presented in (5). DF population patterns with younger children, with two relevant exceptions: DF significantly rate more often as acceptable ungrammatical items and they do not show significant attraction effects (e.g., Franck et al. 2004; 2006). This indicates that agreement is in fact a relevant dependency to assess grammatical competence through grammaticality judgments (Chesi et al., 2023).

### 2.5.2 *Argumental structure*

Another grammatical dependency considered is argument structure. From acquisition studies, we know that at the 1-word stage, which is generally well before age 2:0, verbs are essentially absent from children's productions (Gentner, 1982). Around age 2:0, when children start combining words in two-words utterances (Bates et al., 1995), predicates get productively introduced and the initial thematic structure is set up (Guasti, 2017). According to Ninio (1999), the first predicates used are prototypical predicates of the 'obtaining' (e.g. *want*, *take*, *get*, *bring*, *give*, etc.), 'creating' (e.g. *do*, *make*, *prepare*, etc.) and 'consumption/perception' (e.g. *eat*, *drink*, *see*, and *hear*) type. From this early set, we might conclude that both transitive (e.g., *eat*) and ditransitive (e.g., *give*) predicates are mastered. Notice however that both the subject/agent and the beneficiary are often either the speaker (i.e. the child) or the addressee in both transitive and di-transitive constructions, leaving a bare predicate-theme construction as the only overt production:

(6)      *CHI:give doggie (Adam, 2;3.04)
         *CHI:give paper pencil

This confirms that children often omit given information from early productions (Serratrice & Sorace, 2003). So, using a specific di-transitive predicate does neither prove that the thematic structure of that di-transitive predicate is fully in place, nor that extra thematic roles processing is costless.

A prominent theory on predicates meaning acquisition (and thematic roles accommodation) strongly builds on children's early sensitivity to syntactic cues (Gleitman & Gillette, 1995; Landau & Gleitman, 1985). Children are early sensitive to this kind of evidence and the subtle semantic difference between appar-

ently similar perception verbs is quickly learned (also by blind children, suggesting that the sensorial input is irrelevant in this context). This early sensitivity unfortunately represents a clear obstacle for deaf children that cannot always rely on phonetically weak particles such as prepositions and, in general, verbal particles (including auxiliaries). Their deficit with these elements, which are fundamental for the correct thematic bootstrapping, is attested in many studies (Chesi, 2006; Kluwin, 1982; Radelli, 1999). Focusing on minimal VP thematic difference, unaccusative predicates (the subject is the internal argument, also triggering past participle agreement, that is, it merges first with a lower functional agreement position as compared to unergative predicates) seem to behave differently with respect to unergative (and transitive, with their subject originated in a higher vP thematic position and moving next to inflected T to check subject-verb agreement) ones also from the acquisition perspective (Lorusso, 2018). Three factors must be assessed for disentangling thematic role licensing: first, children sometimes rely on the sub-standard argument introduction using specific prepositions to avoid intervention configurations[1]. Second, auxiliary selection is indicative of the predicate type (unaccusative "be" vs unergative/transitive "have", Sorace 2000), but passive diathesis in transitive verbs might induce confusion, and so probably does the ambiguity between the auxiliary "be" and the copula "be" for which we know children are early sensitive (Franchi, 2006).

We then decided to test deaf children's sensitivity to argument structure by considering some of these crucial factors. More precisely, a group of items in the COnVERSA battery tests the sensitivity to predicate structures by varying the argument role simply (i) adding/removing a preposition (unaccusative predicates) or (ii) removing the argument leaving a locative phrase only (obligatorily transitive predicates):

(7)  a.  Il libro è caduto   dal/*il tavolo                    (unaccusative)
         the book is fallen          from the/*the table
     b.  La mamma mette *(il piatto)       sul tavolo.        (transitive)
         the mom puts     the dish on the table
         *mom puts the dish on the table*

We expect the transitive predicate with an overly realized direct object to pay a fee for the extra adjunct (extra merge operation, that is: unaccusatives < transi-

---

1 This is the case of the "a-marked topics" in Italian: Al re il bambino lo pettina / To the king the child him.cl combs instead of (Belletti & Manetti, 2019). Notice that the prepositional marking of specific arguments is a grammatical strategy in some languages (Spanish) and what we call Differential Object Marking is a tendency early adopted in young children (Rodríguez-Mondoñedo, 2008).

tives), but we can also predict that for those who accept the unergative interpretation of (7).b, the cost of the two structure be the same.

The complex interaction between inflectional positions and thematic structure could not be exhausted by the previous contrasts. Two important phenomena involving inflection both with a degree of morphological independence (namely using overt morphology separated by the main predicate) and with close interaction with predicate type and diathesis choice is auxiliary selection in past participle and passive constructions.

Notice that as far as auxiliary selection is tested, the sensitivity to the auxiliary type should be verified independently of the past participle agreement with unaccusatives. For this reason, since specific items were used to verify agreement in these contexts (Chesi et al., 2023), the items in this block always employ default subjects (animate, singular, masculine) but we crucially included unaccusatives / unergative predicates opposition, (8).a vs (8).b and passive active diathesis (8).c vs (8).d

(8)     a.     Il bambino ha/*è dormito. (unergative)
               The child has/*is slept
        b.     Il treno *ha/è arrivato.      (unaccusative)
               The train *has/is arrived
               The train has arrived
        c.     Il maestro è stato/*ha ringraziato dal bambino. (passive)
               the teacher has been greeted by the child
        d.     Il dottore ha/*è stato visto il ragazzo.       (active)
               The doctor has seen the child

As far as complexity is concerned, apart from the internal (VP, unaccusative) vs external (vP, unergative) base position of the grammatical subject, no difference in terms of merge operations should justify an asymmetry in (8).a and (8).b (which is, unergative = unaccusative). On the other hand, passive constructions should involve more movement operations (Collins, 2005) than the active counterpart (active < passive), while the active sentences require an extra argument merge, with respect to the unergative/unaccusative counterparts. In the end:

(9)     unaccusative = unergative < active/transitive < passive

## 2.5.3 *Pronominalization*

A natural follow-up of the argument structure investigation is the pronominal domain. Cliticization, for instance, is the option adopted in many romance languages to reduce a (given/salient) DP argument to a weak (Cardinaletti & Starke, 1994) pronominal form as exemplified in (10).b below:

(10)  a.    Maria    mangia   la torta
            M.       eats     the cake

(11)  b.    Maria    la$_i$              mangia _$_i$
            M.       it.CL.F.SG          eats
            *M. eats it*

The accusative clitic *la* must be placed in a pre-verbal (proclitic) position when tensed verbs are merged (T), while it incorporates to the verbal root with infinitive predicates (e.g. "mangiar-la", to_eat-it.CL.F.SG). Clitic placement has been used to prove the sensitivity to the inflection of very young children that still produce sub-standard inflections and reduced clitic forms (Guasti, 1993). On the other hand, clitics require a complex derivation (Belletti, 1999) which is only partially related to their phonetic weakness: in the deaf children community, those particles often represent the last residual problem for otherwise rather proficient children (Chesi, 2006; Chesi et al., 2019a; Ghersi, 2017). Both normally developing children and children with specific language impairment (SLI) show problems with clitics in many romance languages (Jakubowicz et al., 1998).

We then decided to include a group of items to test the sensitivity of the children for the correct clitic form, both in terms of case (accusative vs dative) and gender/number agreement with the correct referent:

(12)  a.    La maestra prende il libro   e        lo/*gli legge.
            The teacher takes the book and      it.CL.ACC/*CL.DAT read
            *The teacher takes the book and read it*
      b.    La mamma chiama il bambino e gli/*lo regala un libro.
            The mom calls the child and to_him.CL.DAT/*CL.ACC offers a book
            *the mom calls the child and offers him a book*

These constructions are clearly much more difficult than the previous ones both in terms of minimal merge operations (two full-fledged sentences are conjoined) and height of functional layers required (both C and incorporation to T are required). Between the two constructions, the second is the one requesting more arguments (and cliticizing the oblique one), then our prediction is the following:

(13)    it/him.cl.acc < to_him.cl.dat

As a control, we also contrasted clitics, (14).a, with person agreement (involving high logophoric centers, Sigurdsson 2004) in the answer to questions in 1$^{st}$/2$^{nd}$ person (14).b, including also complex 1$^{st}$/2$^{nd}$ into 3$^{rd}$ person rotation (14).c:

(14)    a.    [child]:  Cosa fa la bambina con il piatto?
              what does the child (do) with the dish?
              [melix]: Lo/*la rompe.     (agree)
                    It.CL.M.SG brakes
              *what does the child do with the dish? She brakes it.*
        b.    [bambino]: Cosa fai?      [melix]: Mangio/*mangi.
              What (do you) do?                 (I) eat.1.SG/*(you) eat.2.SG
        c.    [bambino]: Dì alla mamma che ho fame.
              Say to mom that I'm hungry
              [melix]: Mamma, ha/*ho fame.
              Mom, he is / *I am hungry

Being the higher part of the CP layer (at least QP) activated in these sentences, we expect these constructions to appear more complex than the previous ones, possibly in this order (with person rotation harder than 1$^{st}$/2$^{nd}$ person agreement):

(15)    it/him.CL.ACC < to_him.CL.DAT < cl answer < 1$^{st}$/2$^{nd}$ person answer < 3$^{rd}$ person rotation

2.5.4 *Interrogative structure*
To better accommodate the complexity of the last items, we also included in our battery various kinds of non-local *wh-* dependencies, such as *wh-* questions of the *what/who* kind in which the argument (either subject or object) moves from the basic thematic structure to a highest relevant position into the left periphery (Q/Focus position, §0) and *wh*-adjuncts (*when* and *where*) have been tested[2]. For *wh*-adjuncts, we consider similar non-local dependency in which a lower functional IP layer is activated as an "adjunction site" and the higher *Q/focus* position qualifies as the final landing site. *Why* questions and *yes-no* questions are also been exploratively included in the test.

---

[2] Due to the huge variety of factors we decided to made a choice only focusing on less controversial wh- adjuncts such as *where*, *when* or *with what*, instead for instance of *how* to avoid interpretative confounds (Sæbø, 2016).

From acquisition literature, we know that for younger children interveners represent a relevant problem in all these configurations (Roeper & de Villiers, 1992), and, similarly, deaf children show non-orthodox patterns in *wh-* questions formation (Berent, 1996). An asymmetry between subject and object *wh-* questions is also observed in Italian, with object clearly more problematic than subject questions (Guasti et al., 2014). An asymmetry revealed in production and comprehension suggests that early studies on young children's sensitivity to most kinds of wh-questions were at least optimistic (del Puppo et al., 2016). We then decided to test different kinds of *wh-* questions by focusing on the height of their functional landing site and intervention of each dependency. Starting with the highest *wh-* item, which is base-generated in the left periphery according to (Rizzi, 2001) analysis, *why* questions have been considered (stage 3 in Friedmann et al. 2021) despite the intricacies related to this construction (Beltrame & Chesi, 2021), (16).b. We then considered *where/when* simple (adjunct) questions without intervener, and *who/what* simple subject/object (argument) questions, again without overt intervention (post-verbal subject, smuggling analysis, Belletti & Chesi 2014) (16).c-e (contrasting animate, d, and inanimate, e, objects). According to our complexity comparison, we do not have a reason to assume that argument and adjuncts questions differ in terms of difficulty (both targeting a Q/FocP position) when relevant intervention is absent, while *why* questions (targeting the higher *IntP*) might result minimally more complex. To preserve (and reduce) as much as possible the relevant pragmatic configuration of each *wh-* question, we decided to test the answers to the relevant target questions. The choice of the relevant answer in the pair should reveal the comprehension of the critical factors (Belletti, 2008). In the end, we have also included, exploratively, some yes-no questions (16).a, for which, in Italian, we mainly rely on intonational cues which are obviously absent in written questions.

(16)  a.  [child]: La bambina mangia?        [melix]: Sì/*[una torta].
          the child eats?                          Yes/*a cake
          *Does the child eat?*                    *Yes/*a cake*

     b.  [child]: Perché dorme?      [melix]: Perché [è tardi]/*no.
          why (he) sleeps?            Because (it) is late.
          *Why did he sleep?*         *Because [it's late]/*doesn't.*

     c.  [child]: Chi mangia? [melix]: [Mamma] mangia/*[la pasta].
          Who (he) eats?              Mom eats / *(she) eats pasta.
          *Who does eat?*             *Mom does / she eats pasta*

     d.  [child]: Chi salutano i ragazzi? / *Quali ragazzi saluta Gianni?
          Who greet the boys?        Which boy greets G.?
          *who do the boys great?*   *Which boys does G. greet?*

            [melix]: I ragazzi salutano Gianni.
            The boys greet G.
e.      [child]: Cosa hanno sporcato i bambini?
            What have dirtied the children?
            *What did the children get dirty?*
            I bambini hanno sporcato la tovaglia.
            *The children got the tablecloth dirty*

## 2.6    Dynamic modality as a decision tree

According to the complexity contrasts discussed so far, we formulated a decision tree based on the performance on each phenomenon. The basic idea is to use the discriminative accuracy revealed on a specific block to decide which block to present later: If the complexity scale is correct, we expect a failure on a specific level of a certain complexity to indicate the actual level of performance, that is, if the child does not pass a block of complexity 6, this is, at best, its score/performance (e.g. score 6, accuracy 76%[3]). This approach would dramatically reduce the administration time of the test (passing a block of complexity 5 will prevent the child from being prompted with a block of complexity < 5), possibly maintaining a solid and complete assessment. This would save time to children, families, and therapists and reduce frustration in one sense (too simple blocks will not be presented to children performing very well) or the other (too complex items will not be prompted to children with low performance). Normalizing our contrasts, including agreement data (Chesi et al., 2023), on a 1 to 10 scale, we obtained a complexity hierarchy that we used to build the diagram representing our final decision tree as reported in .

---

[3] COnVERSA synthetic score is produced by multiplying accuracy (e.g., 76%) for the normalized highest complexity score, that is, given 6 as the complexity score and 76 as accuracy result, COnVERSA synthetic score is 76*0.6, namely 45.
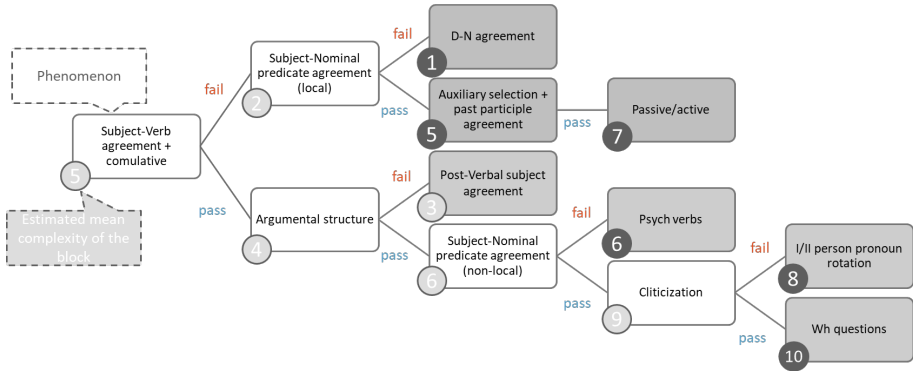
Figure 2. Decision tree used for the block choice to be administered in the dynamic modality

Sometimes the decision is not exactly numeric: For instance, argument structure will be tested (estimated normalized complexity = 4), even though this block follows a success of a block of mean complexity (5), subject-verb agreement + cumulative agreement. Due to the greater solidity (lower variation) in performance with agreement phenomena than with argument structure phenomena as revealed by a preliminary version of this test: the first agreement block might have been passed based on linear word order only considerations, hence argument structure processing must be verified afterward (which is more complex than simple subject-verb agreement, but possibly less complex than cumulative agreement). This is an important check for deaf competence, possibly revealing an interference between LIS (an SOV language without prepositions) and Italian (SVO with indirect object and passive agent marked by prepositions). We expect the dynamic modality to be successful if the accuracy in the complete modality and dynamic one is comparable. In the experimental part, we will concentrate on argument structure, pronominalization, and question formation, in the end considering also agreement (Chesi et al., 2023) for evaluating the correlation between complete and dynamic administrations modalities.

## 3.     Materials and methods

### 3.1     Participants

The control group (NC) was composed of 90 Italian normally developing children; 74 met our inclusion criteria and have been considered in this study (age

range=6-10, M=8.35, SD=1.19; Years of schooling range=1-5, M=3.02, SD=1.18; 40 female): being enrolled without delay in primary school and being monolingual. 60 children spoke a northern Italian variety, 9 a central one, and 9 a southern variety.

The deaf children group (DF) enrolled in the test consisted of 54 participants, but only 34 met the inclusion criteria (age range=7-16, M=10.58, SD=2.46; years of schooling range=1-8, M=4.64, SD=1.96, 15 female): (i) profound deafness, (ii) either cochlearly implanted or prothesized, (iii) either congenital or diagnosed before 3.y.o., (iv) sufficient independence in reading, (v) enrolled at least in primary school. 28 of them had cochlear implants (15 binaural, 13 monoaural), and 6 used external prostheses. 15 children spoke a northern Italian variety, 15 a central one, and 2 a southern variety. Informed consent was collected from all children's parents. The test has been approved by the University of Pavia, Department of Psychology Ethical Committee.

## 3.2    Materials and procedure

The following story, supported by comic-like pictures, was told to each child at the beginning of the test: a little alien, dubbed Melix, just arrived in Italy. He wanted to learn Italian, but as a beginner, he had many troubles producing correct utterances. The child should help him by telling him when the uttered item was "correct" or "wrong". Two warm-up items were presented for practice. The adults administering the test were asked to provide sufficient feedback for these items and verify the children's understanding of the task before the real test began. Upper chars display could have been activated for younger children who preferred that reading experience. The online test has been implemented using JSPsych libraries (De Leeuw, 2015). Answers and reading times were recorded. Smartphones and tablets (90% of the devices used) have been tested successfully with this platform.
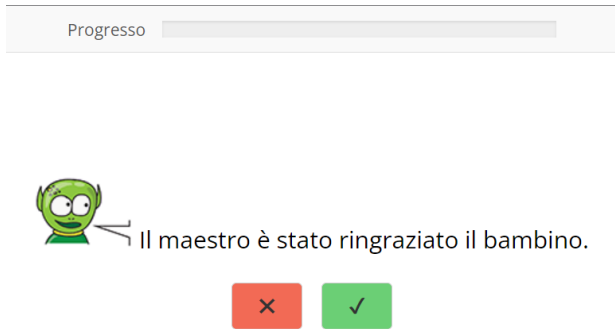
Figure 3. Sample item assessment through the online platform (*the teacher has been thanked \*(by) the child*).

The full test consisted of 240 experimental items. 112 targeted various agreement configurations (Chesi et al., 2023), 40 argument structures, 56 pronominal forms, and 32 questions formation as illustrated in §0.

Lexical items were controlled for elementary accessibility (Marconi, 1994) and sentence readability was above 80 in the GULPEASE index (Lucisano & Piemontese, 1988), which indicates full readability at the primary school level. The test was divided into 2 equivalent parts (A and B), ideally thought for longitudinal studies (same structures, different lexical items). Each part was administered in two sections, both with an equivalent number of items per phenomenon and an even number of wrong and correct items to be judged. After each session, an e-mail was received by the adult administering the test with the following session to be performed. We considered only sections completed within four weeks after the first administration. 254 distinct sections have been collected with NC (M=2.822, SD=2.031). 132 sections have been collected with DC (M=2.469, SD=0.915). Reaction time was also controlled: answers faster than 1000 ms or longer than 60000 ms have been excluded (less than 0.6% of the datapoints). On average, each item took about 6 seconds and a half to be processed (rt range=1018.96-59865.45, M=7032.14, SD=6447.83). Including warm-up and two pauses, we estimated an average of 8 minutes per session.

### 3.2.1 *Administration modalities*

The complete test consisting of 240 items, divided into 2 parts (A and B), each consisting of 2 sub-sections (sections 1 and 2) has been administered to all children. Each dependency configuration (discussed in detail in §0) is investigated at least four times in each section, which consisted of 60 items. Items were dynamically randomized at each administration.

The dynamic modality has been proposed to all children (either before or after the complete modality administration). This modality does not split any block and, consisting of a randomized cluster of phenomena dynamically chosen depending on the child's performance in each block, can vary from 42 to 72 items (as opposed to 120 items of a full complete part). Items within each block were fully dynamically randomized at each administration.

### 3.3    Statistics

The presence of main effects and interactions has been performed fitting generalized linear mixed models under the *R* environment (R Core Team, 2021) using *lme4* package version 1.1.24 (Bates et al., 2015b). Models have been constructed using a parsimonious approach (Bates et al., 2015a). Random structure always includes by subject and by item random intercept adjustments. Models considering Age and Group main effects also include random slope adjustments by subject, and by subject plus by item, respectively. Accuracy was considered a binomial dependent numeric variable (0=wrong, 1=correct), while reading time (rt) was a continuous numeric dependent variable. Five fixed factors were considered in all analyses: expected grammatical Correctness (ungrammatical, grammatical), 3 dependency type (3-levels factor, with sub-type nested factors), two continuous factors (Age, with decimal specification, also considered as a two levels factor, Age group; School, integer). Three more factors, two of them continuous and one categorical, were included in the DF group analyses only: Onset (integer, from 0, birth, to 5 y.o.), Hearing aid (Cochlear implant, external prosthesis), implantation age (before 1 to 5 y.o.). One global analysis and three distinct analyses are performed targeting each dependency type.

### 4.    Results

### 4.1    General and group-specific factors

Overall, the DF group performs significantly worse than NC ($\chi^2$(1)=25.239, $p$=0.0002). No trial order effect is obtained ($\chi^2$(1)= 2.264, $p$=0.1324): both DF and NF perform equally at the beginning and at the end of the experiment even though in both groups, children become significantly faster with later items ($\chi^2$(1)=65.308, $p$<0.0001). DF are 12% slower than NC (NC: 6871ms on average per item vs DF: 7717ms on average per item) and their performance is (only numerically) slightly less accurate at the end of the experiment. Italian variety spoken is not a significant factor either ($\chi^2$(7)=4.9759, p=0.6629).

### 4.1.1 *Correlation between administration modalities*

To evaluate the reliability of the different modalities and parts we checked by subject Pearson's correlation between sessions: we calculated the average performance per section and then the correlation between sections also including the dynamic modality (Table 1). Since most deaf subjects only completed part A, the correlation analysis is only performed between A subsections and the dynamic modality. The high correlation between parts A and B confirms that performance in those parts is in fact equivalent. A similar correlation in performance is observed between the two sections and between the overall performance in the A part and the Dynamic modality (Table 2). This result is not affected by the overall performance: both deaf children with poor performance and children with higher performance contribute to the correlation in a comparable way.

**Table 1.** By Subjects Pearsons' correlation between sections and administration modality of the test. A and B are the averages of A1-A2 and B1-B2 performance respectively.

|  | A-2 | B-1 | B-2 | A | B | Dynamic |
|---|---|---|---|---|---|---|
| A-1 | $r(44) = .44$ $p = .002$ | $r(37) = .57$ $p < .001$ | $r(29) = .36$ $p = .045$ | $r(44) = .80$ $p < .001$ | $r(29) = .49$ $p = .005$ | $r(23) = .23$ $p = .273$ |
| A-2 |  | $r(36) = .30$ $p = .064$ | $r(29) = .42$ $p = .018$ | $r(44) = .89$ $p < .001$ | $r(29) = .45$ $p = .012$ | $r(24) = .39$ $p = .047$ |
| B-1 |  |  | $r(31) = .49$ $p = .004$ | $r(34) = .48$ $p = .003$ | $r(31) = .76$ $p < .001$ | $r(23) = .24$ $p = .248$ |
| B-2 |  |  |  | $r(27) = .56$ $p = .002$ | $r(31) = .94$ $p < .001$ | $r(23) = .48$ $p = .015$ |
| A |  |  |  |  | $r(27) = .60$ $p < .002$ | $r(23) = .41$ $p = .042$ |
| B |  |  |  |  |  | $r(23) = .43$ $p = .032$ |

**Table 2**. By Subject Pearsons' correlation between sections and administration modality of the test in deaf subjects. A is the average of A1-A2 performance.

|  | A-2 | A | Dynamic |
|---|---|---|---|
| A-1 | $r(19) = .86$ $p < .001$ | $r(19) = .95$ $p < .001$ | $r(17) = .65$ $p = .002$ |
| A-2 |  | $r(19) = .98$ $p < .001$ | $r(14) = .44$ $p = .085$ |
| A |  |  | $r(14) = .48$ $p = .057$ |

### 4.1.2 *Age, school, and estimated complexity*
Both Age and (estimated) Complexity are significant predictors of the overall NC performance, but not for DF: in NC, accuracy increases with Age ($\chi^2$=7.1134, $p$=0.008, Figure 4.a) and decreases with Estimated Complexity ($\chi^2$= 20.1040, $p$<0.001, Figure 4.b). School, compared to Age, is a less significant predictor, although it is still significant in the hearing population ($\chi^2$=6.8466, $p$=0.033) but not in the deaf one ($\chi^2$=2.0012, $p$<0.3677).

Age x Complexity interaction is also strongly significant in the hearing children group ($\chi^2$=15.8114, $p$<0.001) suggesting that a complexity increase induces a more marked performance decrease in younger hearing children than in older hearing ones. Overall, estimated complexity is also a significant predictor of performance in the deaf group ($\chi^2$=7.9923, $p$=0.01839), but notice the variance.
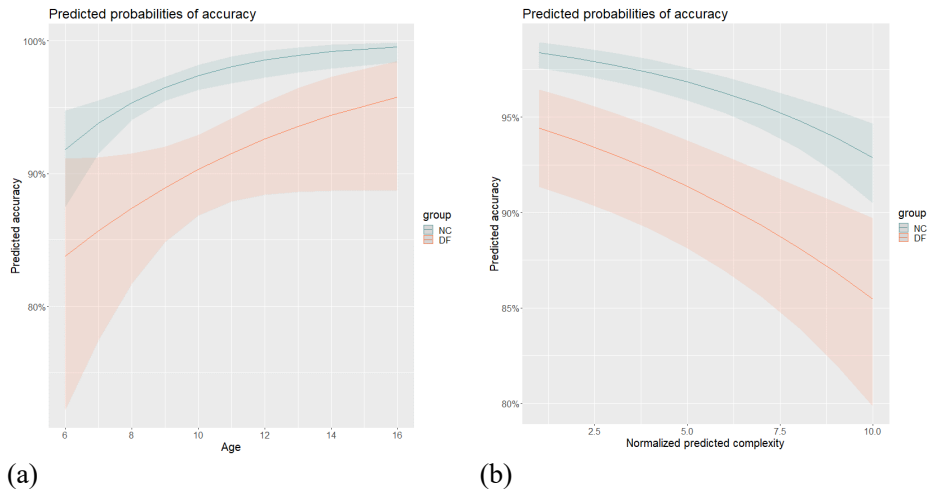


Figure 4. Overall predicted accuracy in DF and NC groups, based on Age X Group (a), and estimated Complexity X Group interaction. Shading indicates 95% Confidence Interval (CI). Data for NC older than 10 and DF younger than 6 are interpolated for convenience.

### 4.1.3 *Deaf-specific factors*
Diagnosis year ($\chi^2$(1)=0.0421, p=0.8374) is not a relevant predictor of deaf children's perception of grammaticality in the test. On the other hand, congenital deafness is a relevant factor ($\chi^2$(1)=6.2522, p=0.0124): non-congenital deaf performs systematically better (*estimate*=0.9943, *SE*=0.3741, *z*=2.658, p=0.0079). Also having a cochlear implant significantly correlates with a performance improvement in the test ($\chi^2$=9.4864, p=0.0087); the performance of children wear-

ing an external prosthesis is much worse than that of children with CI (*estimate*= -1.2488, *SE*=0.5193, *z*=-2.405, *p*=0.0162). A sub-analysis within the CI children implanted for which we know the exact implantation year (N=12) at age 1, 2, or after 3, revealed a degradation in performance with later implantation (*estimate*=-0.8409, *SE*=0.4474, *z*=-1.880, *p*=0.0602). A strongly significant interaction between estimated complexity and CI ($\chi^2(1)$=14.8781, *p*=0.0005) suggests that while cochlear-implanted children perceive estimated complexity similarly to the control group, those children who wear an external prosthesis lose completely this discriminative ability.

## 4.2 Specific phenomena

### 4.2.1 *Argument structure*

A clear group effect is obtained with NC performing significantly better than DF ($\chi^2(3)$=37.896, *p*<0.0001). Age is also a mildly significant factor ($\chi^2(3)$=7.5813, *p*=0.05551). Neither the contrast type nor nested features factors seem significant overall (Figure 5):
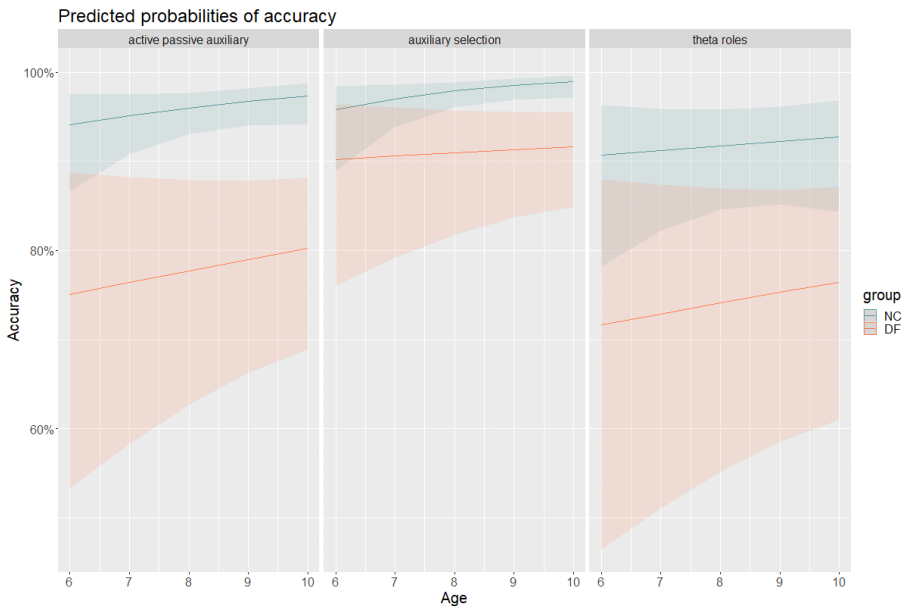


Figure 5. Predicted accuracy on argument structure judgments based on the three-way interactions Group X Contrast type X Age (a) and Group X Feature type X Age (b) (95% CI).

Digging into within-groups contrasts, NC performs significantly worse with thematic role substitutions/omissions than with auxiliary selection (*estimate*=-1.451, *SE*=0.453, *z*=-3.206, *p*=0.0038). Inspecting nested features factors, in fact, this effect seems to be driven only by the higher acceptance of the (ungrammatical) cases in which an omission is present (omission – substitution: *estimate*=-1.554, *SE*=0.610, *z*=-2.548, *p*=0.0529).

A similar pattern emerges in DF ($\chi^2$(2)=6.2329, *p*=0.0443), with hearing aid marginally interacting with the contrast type ($\chi^2$(3)=6.3184, *p*=0.0971, Figure 6.a) and strongly interacting with nested features factors ($\chi^2$(4)=14.679, *p*=0.0054, Figure 6.b):



(a)                                                     (b)
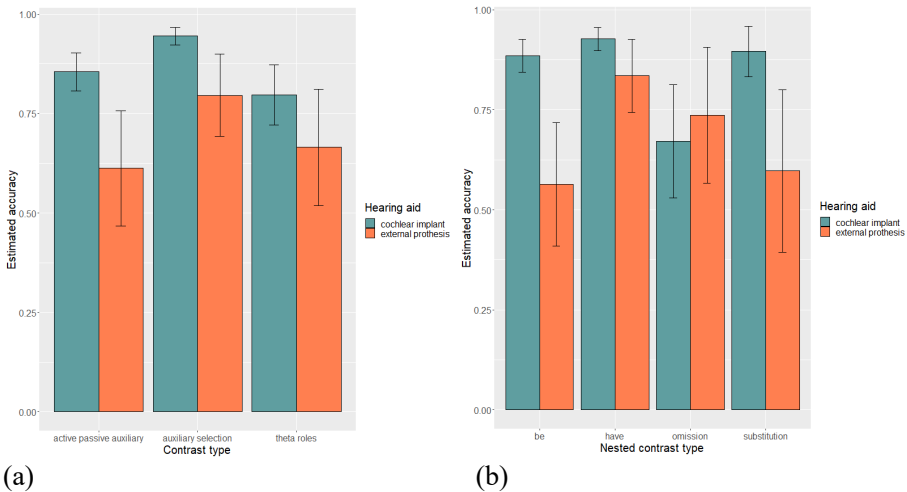
Figure 6. Predicted accuracy based on Hearing aid x Contrast type and Hearing aid X Feature type in DF. Error bars indicate Standard Errors (SE).

The lower performance of the DF group is especially clear in the auxiliary selection contrast in the passive diathesis, where non-cochlearly implanted children present major difficulties. Overall, DF present a significantly higher degree of acceptance for ungrammatical sentence in all conditions, while NC only present this preference in case of direct object omission.

### 4.2.2 *Pronouns*

A huge group effect is again observed ($\chi^2$(5)=34.098, p<0.0001) but not global age effect ($\chi^2$(3)=1.7529, p<0.6252), even though a significant group x age ($\chi^2$(2)=23.8566, p<0.0001) and two three-way interactions are obtained: group x

age x contrast type ($\chi^2(6)$=17.423, $p$=0.0078, Figure 7.a) and group x age x feature type ($\chi^2(6)$=130.143, $p$<0.0001, Figure 7.b):



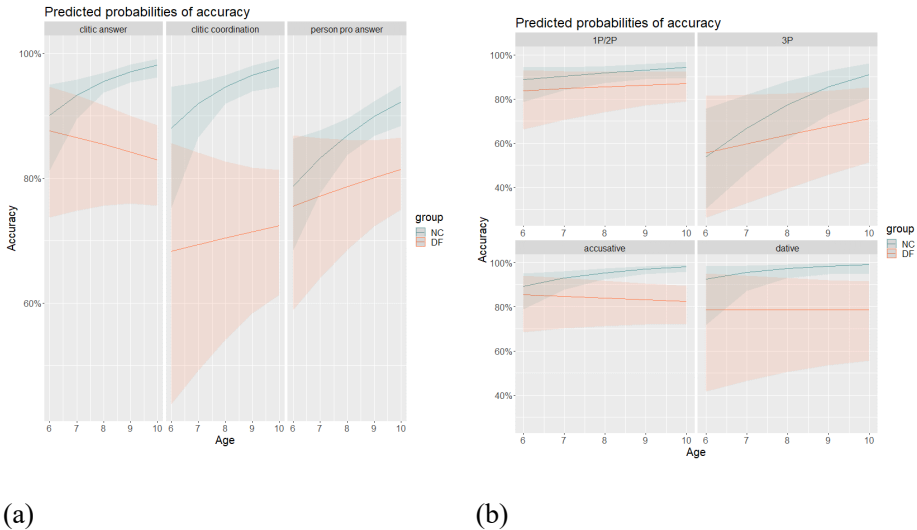(a)                                              (b)

Figure 7. Predicted accuracy on pronominal judgments based on the three-way interactions Group X Contrast type X Age (a) and Group X Feature type X Age (b) (95% CI).

Age becomes a significant factor within NC ($\chi^2(2)$=9.8749, $p$=0.0072). Younger (6-7 y.o.) children perform significantly worse with respect to the older ones (8-10 y.o.) on judging the correct clitic form in question answering (*estimate*=-1.191, *SE*=0.381, *z*=-3.130, *p*=0.0216) and, marginally, in person rotation answering (*estimate*=-0.740, *SE*=0.283, *z*=-2.614, *p*=0.0938). NC also present a bias in accepting ungrammatical items, but only in person rotation answering (*estimate*=-1.154, *SE*=0.309, *z*=-3.735, *p*=0.0026). Digging into feature contrasts, a clear improvement is observed both with accusative clitics (*estimate*=-1.134, *SE*=0.350, *z*=-3.237, *p*=0.0266) and with 3rd person rotation in (*estimate*=-1.194, *SE*=0.350, *z*=-3.410, *p*=0.0150).

The DF group, on the other hand, again shows an indiscriminate bias in accepting ungrammatical sentences independently of the contrast or feature type ($\chi^2(3)$=16.961, $p$=0.0007). A mild congenital ($\chi^2(1)$=3.3354, $p$=0.0678) and hearing aid ($\chi^2(1)$=3.1169, $p$=0.0775) effects are observed, with a mild interaction between Hearing aid and Contrast type ($\chi^2(3)$=6.4861, $p$=0.09021, Figure 8.a). Hearing aid X Feature type interaction is not significant (Figure 8.b).

(a)                                                    (b)
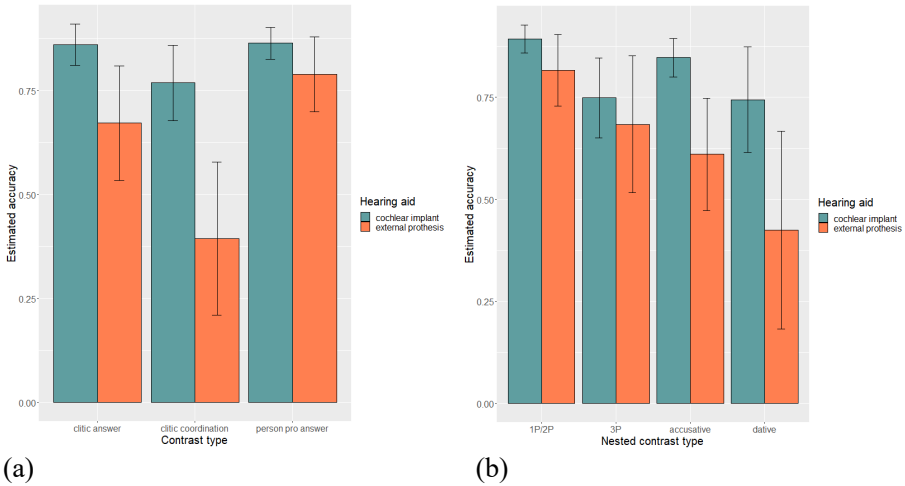
Figure 8. Predicted accuracy based on the two-way Hearing aid X Contrast type interaction (a) and Hearing aid X Feature type interaction (b) with SE.

### 4.2.3 *Interrogative structures*

A main group ($\chi^2(5)$=11.168, *p*=0.0481) and Age ($\chi^2(3)$=9.8575, *p*=0.0198, essentially driven by the NC control group, Figure 9) effects are observed. Both a two-way Contrast type X Group ($\chi^2(5)$=15.5349, *p*=0.0083) and a three-way interaction Contrast Type X Age X Group ($\chi^2(4)$=9.0443, *p*=0.060, Figure 9.a; notice that the model presents a minor convergence warning due to huge performance variance in yes-no questions with DF; in this group, in this condition, the prediction is probably inaccurate).
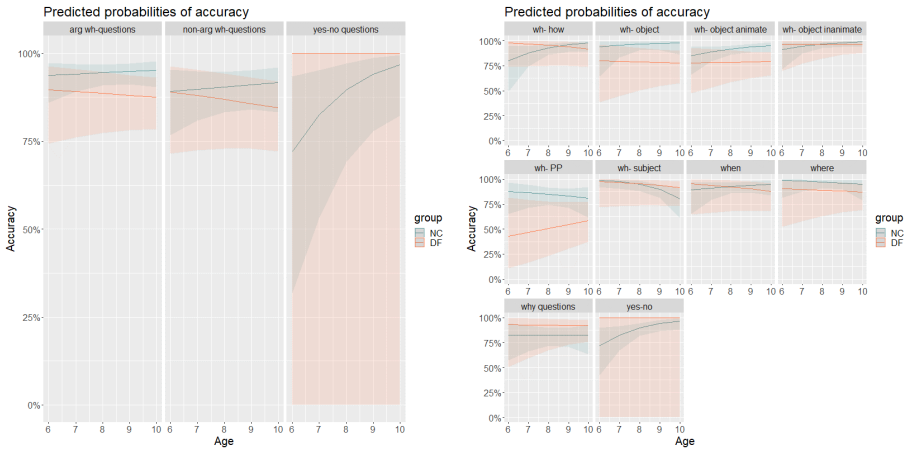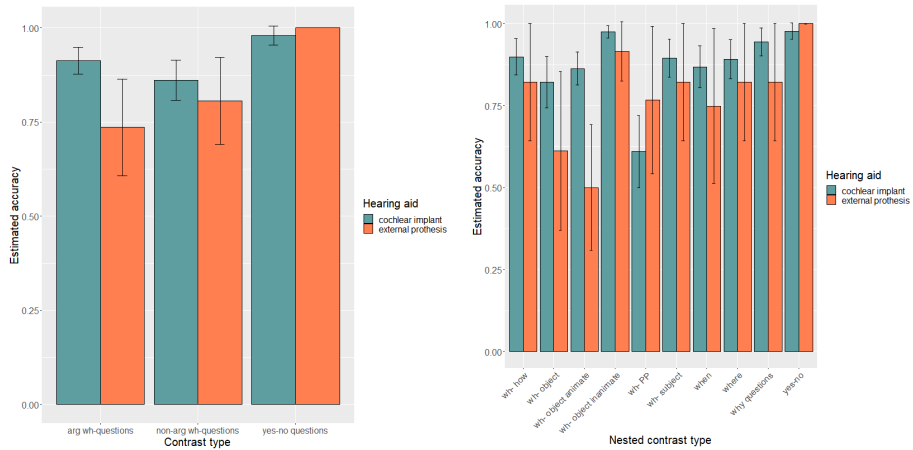
(a)                                                    (b)

Figure 9. Predicted accuracy on question judgments based on three-way interactions Group X Contrast type X Age (a), Group X Feature type X Age (b) (95% CI).

The NC group does not show any significant bias in accepting ungrammatical items and the only relevant contrasts obtained are between object *wh-* questions and *why* questions with the best performance recorded with the first dependencies (*estimate*=1.8530, *SE*=0.576, *z*=3.214, *p*=0.0428), *where* questions again against *why* questions (*estimate*=1.8490, *SE*=0.576, *z*=3.207, *p*=0.0438) and *wh-object question*, with inanimate object (prototypical questions) - *wh-* questions with PP adjuncts (*estimate*=1.7024, *SE*=0.437, *z*=3.894, *p*=0.0039).

DF, as usual, present a bias towards ungrammatical acceptance in argumental wh-questions (*estimate*=-1.4569, *SE*=0.465, *z*=-3.130, *p*=0.0216). The huge variance in performance essentially erases any significant contrast, with the relevant exception of the prototypical *wh-* object question (inanimate object) vs *wh-*adjunct introduced by a preposition, the first being significantly more accurately judged than the second (*estimate*=2.8767 0.743, *SE*=3.871, *z*=0.0043).

(a)                                                (b)

Figure 10. Predicted accuracy on question judgments based on two-way Hearing aid X Contrast type interaction (a) and Hearing aid X Feature type interaction (b) with SE.

## 5.     Discussion

### 5.1     Administration modalities, experiment length, and complexity

A relevant correlation is observed both between parts A and B and between both parts and the Dynamic modality. This confirms that: (i) the Dynamic administration modality is sufficiently reliable and can be safely used to assess the implicit grammatical competence both in NC and DF, (ii) the complexity scale is sound and predictive of the overall performance, (iii) the two parts, being equivalent, can be used in longitudinal studies that aim at assessing rigorously the competence level of tested children before and after a specific logopedic activity always using different lexical items. The fact that the order of the items is not a significant predictor of performance within groups indicates that the length of each section/modality is appropriate for the age range tested both in NC and DF.

### 5.2     Age and school

Both age and schooling contribute to performance in hearing children but not in deaf children. This confirms that competence in deaf children is greatly independent of age (Chesi, 2006). Age better explains hearing children's perfor-

mance than schooling. This is probably due to the finer granularity of distinctions for the Age dimension (decimal) than for Schooling (integer).

## 5.3    Deaf specific factors

The heterogeneity of the DF group partially allowed us to assess the specific contribution of various factors in this population of profoundly deaf children with different ages, schooling, and logopedic histories. What clearly seems to matter as a significant factor is the hearing aid adopted: cochlear implanted children greatly outperform deaf children using an external prosthesis. Their performance is not only higher in all contrasts, but also their sensitivity to featural opposition is always more coherent with the younger 6-7 y.o. NC children. The utility of an (early) CI confirms the previous studies performed on Italian (Guasti et al., 2014) as well as other languages (Friedmann & Szterman, 2006).

## 5.4    Argument structure

The prediction scale in (9), repeated below for convenience, is only numerically supported by the NC population.

(9)    unaccusative = unergative < active/transititive < passive

The clear contrast in performance, observed with transitive predicates (low performance, high variance) as compared to unaccusative/unergative predicates, is totally related to the acceptance as an unergative of the transitive predicate (e.g., *il nonno mette sul tavolo / Grandpa put on the table*). Moreover the significant variance in the recorded data does not decrease with age, indicating a solid tendency to be further investigated. A plausible hypothesis is that argument dropping (c.f. topic-drop proposed for clitics, Chesi 2006) is considered a less severe violation compared to auxiliary wrong selection or introduction of the oblique argument (e.g. *il signore passeggia la strada / the man walk the road*).

The absence of other significant differences in performance with respect to contrasts is not surprising: on the one hand, the numeric trend fits well with the estimated complexity even though no significant statistical distinctions can be found in terms of estimated marginal means, on the other, the "almost at-ceiling" performance since the beginning (even though a general Age effect is observed) indicates that both passive and active constructions (at least as far as auxiliary selection is concerned, that is, T-related phenomena) are well mastered at 6 y.o. The prediction that passive constructions are more problematic than simple auxiliary selection is borne out: hearing children correctly individuate the

appropriate auxiliary before mastering the auxiliary in passive diathesis (the contrast is stronger in the younger population than in the older one). This is also partially true in deaf children with Cochlear Implants that however perform worse on both tasks compared to the hearing children in the control group. This contrast is even more marked in DF with an external prosthesis (Figure 6). To be noticed the remarkable drop in performance with the nested feature "be" in non-implanted children, which is mostly driven by the acceptance of the "have" auxiliary in passive constructions. In this case, the complexity scale predicts correctly significant contrasts found in the DF population. This suggests that the absence of significant differences in NC (lower functional layer involved, phenomena fully mastered at the age under scrutiny) does not obscure the utility of the scale that not only produces the best-fitting regression model but also indicates probable contrasts in a less linguistically mature population.

## 5.5    Person concord and cliticization

The prediction we made in this domain was the following one:

(15)    it/him.CL.ACC < to_him.CL.DAT < cl answer < $1^{st}/2^{nd}$ person answer < $3^{rd}$ person rotation

This scale is fully supported both by the perfect regression fit and by the relevant contrasts obtained in NC. Age improvement in this domain is clear in the NC population, as well as the difficulty in answering correctly by selecting the correct person ($1^{st}/2^{nd}$ person) or gender/number. Performance variance is also significantly reduced in older children, confirming that this type of dependencies, involving both C and T domains, requires more time to be mastered. The fact that no significant contrast is observed between accusative and dative clitics in both groups is however compensated by the greater variance in performance with datives in both populations (especially in younger NC), thus confirming that the complexity scale is on the right track.

DF group performs as expected according to the literature: clitics and person agreement/rotation are in fact among the last domains resisting maturation (Chesi, 2006; Chesi et al., 2019b; Musola, 2006). This is clear both by looking at the lower performance and the high variance recorded with these items. Overall, DF again show a significant bias in accepting ungrammatical sentences. Here, again, a lower performance of the non-implanted children is observed. The expected accusative < dative clitic difference is on the other hand confirmed.

## 5.6 Interrogative structure

The cluster of phenomena included here was necessary also for interpreting the results obtained in the pronominal domain, where agreement concord was elicited through question-answers pairs. The main expectation was that *why* questions, involving the highest functional node in the CP domain, should have been the hardest configuration to be mastered. This prediction is borne out in NC, but not in DF: NC performs significantly better with object wh-questions, in which the subject was post-verbal than with why questions. This, on the one hand, confirms our prediction, on the other, suggests that no intervention was present as suggested following the smuggling analysis of the post-verbal subject derivation (Belletti & Chesi, 2014). Notice that, in this configuration, DF children perform quite badly, especially non-cochlear implanted children. This suggests that, in these children, the "smuggling" derivation to avoid intervention is not a suitable option. Overall, DF again present an acceptance bias with ungrammatical solutions. Notice that the huge variability in performance with yes-no questions does not allow us to rely much on the predicted accuracy in this domain for the DF population. A similar variance is observed in the younger population, suggesting that these kinds of questions, requiring a clear prosodic contour to be correctly interpreted as questions, are in fact problematic. An alternative interpretation of these constructions as declaratives would favor an interpretation of the answer as a completion of the previous sentence: *La bambina mangia… una torta* / *The child eats… a cake*. This explains the high acceptance revealed both in DF and in younger NC of these items.

Overall, also in this case, the complexity scale supports the major contrast and correctly predicts this as the hardest domain: it involves the highest functional positions, and it might produce intervention effects in those children that are not sufficiently mature to perform a "smuggling" derivation.

## 6.    Conclusion

In this work, we demonstrated how the precise assessment of linguistic competence in deaf children can be performed using simple grammaticality judgments of written sentences forming minimal pairs in which a single feature variation induces a clearly detectable ungrammaticality in the adult grammar. Adopting this approach, we mapped a relevant set of structural aspects whose mastering was attested in the literature for hearing children before age 6. In our test we observed an overall good performance on the items used in the hearing children population ranging from 6 to 10 y.o., but we also noticed an improvement in

many phenomena which are not at ceiling level as one might expect. The problems revealed are systematically correlated with the complexity metrics we associated with each item: younger children perform at ceiling on simpler items (D – N agreement, subject – nominal predicate agreement) but not with more complex ones (e.g., clitics, object-drop licensing, and certain interrogative constructions).

Similarly, we can appreciate a sensitivity for the complexity metrics in deaf children which is complementary to the one observed in agreement dependencies (Chesi et al., 2023): while DF children, as younger NC children were less sensitive to finer discriminations (supporting Grillo's 2008 original idea), in argument structure, pronominal concord, and question formation we observed the opposite: NC performing at ceiling with certain contrasts, only numerically support the complexity metrics, while in DF, performing worse on these contrasts, significant differences emerge, on the one hand indicating that the expected analysis was problematic for these children, on the other, suggesting that a simpler strategy adopted by mature children is not available for DF.

This study also confirms the fragilities of deaf children discussed in the literature (pronominal forms, non-local dependency formation) and on the utility of a cochlear implant. Those children wearing an external prosthesis perform significantly worse (and equally bad) on most items with mid-high complexity on our scale. Moreover, in deaf children, Age is not a significant factor: only occasionally older deaf children perform better than younger ones (and the opposite is also true, sometimes).

Focusing on the contrasts tested in COnVERSA, worth to be highlighted is the fact that the complexity metrics is built simply by comparing minimal differences between phenomena both in terms of height of the functional projection involved and locality of the dependency (expressed in terms of interveners). This intuition (Chesi & Canal, 2019; Friedmann et al., 2009, 2021) seems to produce solid generalizations useful to guide both the phenomena to be tested dynamically and the eventual logopedic support directed at deaf children with specific fragilities identified by a certain level of complexity.

## Acknowledgments

# References

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. 2015a. Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. 2015b. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1): 1–48.

Bates, E., Dale, P. S., Thal, D., & others. 1995. Individual differences and their implications for theories of language development. *The handbook of child language*, 30: 96–151.

Belletti, A. 1999. Italian/Romance clitics: Structure and derivation. *Empirical approaches to language typology*, 543–580.

Belletti, A. 2004. Aspects of the low IP area. In L. Rizzi (Ed.), *The structure of CP and IP. The cartography of syntactic structures* (Vol. 2, pp. 16–51). Oxford (UK): Oxford University Press.

Belletti, A. 2008. Answering strategies: New information subjects and the nature of clefts. In A. Belletti (Ed.), *Structures and Strategies* (pp. 242–265). Routledge.

Belletti, A. 2017. (Past) participle agreement. *The Wiley Blackwell Companion to Syntax, Second Edition*, 1–29.

Belletti, A., & Chesi, C. 2014. A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production. *RGG. RIVISTA DI GRAMMATICA GENERATIVA*, 36: 1–28.

Belletti, A., & Manetti, C. 2019. Topics and passives in Italian-speaking children and adults. *Language Acquisition*, 26(2): 153–182.

Belletti, A., & Rizzi, L. 1988. Psych-verbs and θ-theory. *Natural Language & Linguistic Theory*, 6(3): 291–352.

Beltrame, F., & Chesi, C. 2021. Why-questions and focus in Italian. *The Linguistic Review*, 38(4): 687–726.

Berent, G. P. 1996. Learnability constraints on deaf learners' acquisition of English wh-questions. *Journal of Speech, Language, and Hearing Research*, 39(3): 625–642.

Bertone, C., Cardinaletti, A., Grosselle, S., & Volpato, F. 2011. Le abilità di comprensione dell'italiano in sei adolescenti sordi segnanti LIS. In E. Franchi & D. Musola (Eds.), *Acquisizione dell'italiano e sordità* (pp. 87–103). Venezia: Libreria Editrice Cafoscarina.

Bishop, D. V. 2009. *TROG 2: Test for Reception of Grammar - Version 2* (II.). Firenze: Giunti Editore OS.

Cardinaletti, A. (Ed.). 2018. *Test linguistici accessibili per studenti sordi e con Dsa: Pari opportunità per l'accesso all'Università*. Franco Angeli.

Cardinaletti, A., & Starke, M. 1994. The typology of structural deficiency: A case study of the three classes of pronouns. In H. van Riemsdijk (Ed.), *Clitics in the languages of Europe* (pp. 145–234). Berlin: Mouton de Gruyter.

Cecchetto, C., Di Domenico, A., Garraffa, M., & Papagno, C. 2012. *Comprendo. Batteria per la Comprensione di frasi negli adulti*. Raffaele Cortina.

Chesi, C. 2006. *Il linguaggio verbale non-standard dei bambini sordi*. ROMA: EDIZIONI UNIVERSITARIE ROMANE.

Chesi, C., & Canal, P. 2019. Person Features and Lexical Restrictions in Italian Clefts. *FRONTIERS IN PSYCHOLOGY*.

Chesi, C., Ghersi, G., Musella, V., & Musola, D. 2023. On agreement complexity in hearing and deaf children's oral language acquisition: testing minimal pairs in Italian. *Under review*.

Chesi, C., Ghersi, G., & Musola, D. 2019a. Minimal pairs as triggers for clitics acquisition in hearing-impaired children. *GALA 14 Generative Approaches to Language Acquisition* (pp. 8–9). Milano (IT): University of Milano-Bicocca.

Chesi, C., Ghersi, G., & Musola, D. 2019b. L'acquisizione dei pronomi clitici nei sordi: evidenze a favore dell'utilità dell'esposizione a coppie minime. *STUDI E SAGGI LINGUISTICI*, 57: 17–70.

Chilosi, A. M., Cipriani, P., Giorgi, A., Fazzi, B., & Pfanner, L. 2006. *TCGB: Test di comprensione grammaticale per bambini* (II.). Edizioni del Cerro.

Chomsky, N. 1957. *Syntactic structures*. Berlin: Walter de Gruyter.

Chomsky, N. 1981. *Lectures on government and binding: The Pisa lectures*. Walter de Gruyter.

Chomsky, N. 1995. *The minimalist program*. Cambridge, MA: MIT press.

Chomsky, N. 2001. Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). Cambridge (MA): MIT Press.

Cinque, G. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford (UK): Oxford University Press.

Cinque, G. 2002. *Functional Structure in DP and IP: The Cartography of Syntactic Structures, Volume 1*. Oxford University Press.

Collins, C. 2005. A smuggling approach to raising in English. *Linguistic Inquiry*, 36(2): 289–298.

Curtiss, S. 1978. *Genie: A Psycholinguistic Study of a Modern-day" wild Child"*. New York: Academic.

De Leeuw, J. R. 2015. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, 47(1): 1–12.

del Puppo, G., Pivi, M., & Cardinaletti, A. 2016. 5 Elicited production of who-questions by school-aged Italian-speaking children. *Acquisition of Romance Languages*, 121–140.

Franchi, E. 2006. Patterns of copula omission in Italian child language. *LANGUAGE ACQUISITION AND LANGUAGE DISORDERS*, 41: 135.

Franchi, E., & Musola, D. 2011. La logogenia come strumento di indagine dell'autonomia linguistica dei sordi in italiano: metodo e primi risultati. *Franchi Elisa and Musola*

*Debora (eds.), Acquisizione dell'italiano e sordità, Venezia: Libreria Editrice Cafoscarina, 2011, pp. 47-69*. Venezia, Libreria Editrice Cafoscarina.

Franchi, E., & Musola, D. 2012. Percorsi di Logogenia/1. Strumenti per l'arricchimento del lessico con il bambino sordo. *Venezia: Cafoscarina*.

Franck, J., Cronel-Ohayon, S., Chillier, L., Frauenfelder, U. H., Hamann, C., Rizzi, L., & Zesiger, P. 2004. Normal and pathological development of subject–verb agreement in speech production: a study on French children. *Journal of Neurolinguistics*, 17(2–3): 147–180.

Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. 2006. Agreement and movement: A syntactic analysis of attraction. *Cognition*, 101(1): 173–216.

Friedmann, N., Belletti, A., & Rizzi, L. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, 119(1): 67–88.

Friedmann, N., Belletti, A., & Rizzi, L. 2021. Growing trees: The acquisition of the left periphery. *Glossa: a journal of general linguistics*, 6(1).

Friedmann, N., & Szterman, R. 2006. Syntactic movement in orally trained children with hearing impairment. *Journal of Deaf Studies and Deaf Education*, 11(1): 56–75.

Gentner, D. 1982. Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Center for the Study of Reading Technical Report; no. 257*.

Ghersi, G. 2017. *L'acquisizione del linguaggio in contesti speciali: il caso dei clitici nella sordità*. IUSS Pavia, Pavia.

Gleitman, L. R., & Gillette, J. 1995. The handbook of child language. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 413–427). Oxford (UK): Blackwell.

Gordon, P. 1996. The truth-value judgment task. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax*. Cambridge (MA): MIT Press.

Gordon, P., Hendrick, R., & Johnson, M. 2004. Effects of noun phrase type on sentence complexity. *Journal of memory and Language*, 51(1): 97–114.

Grillo, N. 2008. *Generalized minimality: Syntactic underspecification in Broca's aphasia*. LOT.

Guasti, M. T. 1993. Verb syntax in Italian child grammar: Finite and nonfinite verbs. *Language acquisition*, 3(1): 1–40.

Guasti, M. T. 2017. *Language acquisition: The growth of grammar*. MIT press.

Guasti, M. T., Papagno, C., Vernice, M., Cecchetto, C., Giuliani, A., & Burdo, S. 2014. The effect of language structure on linguistic strengths and weaknesses in children with cochlear implants: Evidence from Italian. *Applied Psycholinguistics*, 35(4): 739–764.

Haegeman, L. 1996. Root infinitives, clitics and truncated structures. *Language acquisition and language disorders*, 14: 271–308.

Hale, K., & Keyser, S. J. 1993. On argument structure and the lexical expression of syntactic relations. *1993*, 53–109.

Hoekstra, T., & Hyams, N. 1998. Aspects of root infinitives. *Lingua*, 106(1–4): 81–112.

Jakubowicz, C., Nash, L., Rigaut, C., & Gerard, C.-L. 1998. Determiners and clitic pronouns in French-speaking children with SLI. *Language acquisition*, 7(2–4): 113–160.

Kluwin, T. N. 1982. Deaf adolescents' comprehension of English prepositions. *American Annals of the Deaf*, 852–859.

Landau, B., & Gleitman, L. R. 1985. *Language and experience: Evidence from the blind child* (Vol. 8). Harvard University Press.

Larson, R. K. 1988. On the double object construction. *Linguistic inquiry*, 19(3): 335–391.

Lewis, R. L., & Vasishth, S. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, 29(3): 375–419.

Lorusso, P. 2018. *The Acquisition of Verbs at the Syntax-semantics Interface: Early Predicates*. Cambridge Scholars Publishing.

Lucisano, P., & Piemontese, M. E. 1988. GULPEASE: una formula per la predizione della difficoltà dei testi in lingua italiana. *Scuola e città*, 3(31): 110–124.

Marconi, L. 1994. *Lessico elementare: Dati statistici sull'italiano scritto e letto dai bambini delle elementari*. Zanichelli.

McDaniel, D., McKee, C., & Cairns, H. S. (Eds.). 1996. *Methods for assessing children's syntax*. Language, speech, and communication. Cambridge, Mass: MIT Press.

Moscati, V., & Rizzi, L. 2014. Agreement configurations in language development: a movement-based complexity metric. *Lingua*, 140: 67–82.

Musola, D. 2006. *Né io né tu: un percorso di logogenia nell'analisi del tratto di persona: il caso di un'adolescente sorda profonda preverbale non segnante*. PhD Thesis. Università Ca' Foscari Venezia, Venezia.

Ninio, A. 1999. Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of child language*, 26(3): 619–653.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Radelli, B. 1999. *Nicola non vuole le virgole: dialoghi con i sordi: introduzione alla logogenia*. Decibel.

Rizzi, L. 2001. On the position "Int (errogative)" in the left periphery of the clause. *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*, 59: 287–296.

Rizzi, L. 2004. *The Structure of CP and IP: The Cartography of Syntactic Structures Volume 2* (Vol. 2). Oxford University Press on Demand.

Rodríguez-Mondoñedo, M. 2008. The acquisition of differential object marking in Spanish. *Probus*, 20(1): 111–145.

Roeper, T., & de Villiers, J. 1992. Ordered decisions in the acquisition of wh-questions. *Theoretical issues in language acquisition: Continuity and change in development*, 191–236.

Rustioni, D., & Lancaster, M. 2007. *PVCL: Prove di Valutazione della Comprensione Linguistica*. Firenze: Edizioni Giunti OS.

Sæbø, K. J. 2016. "How" questions and the manner–method distinction. *Synthese*, 193(10): 3169–3194.

Schütze, C. T. 2016. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Language Science Press.

Sigurdsson, H. A. 2004. The syntax of person, tense and speech features. *Italian journal of linguistics*, 16: 219–251.

Sorace, A. 2000. Gradients in auxiliary selection with intransitive verbs. *Language*, 859–890.

Volpato, F. 2010. *The acquisition of relative clauses and phi-features: evidence from hearing and hearing-impaired populations*. Università Ca' Foscari.

Wechsler, D. 2014. Wechsler Intelligence Scale for Children-IV. (O. Arturo, P. Lina, & P. Laura, Trans.)*Florence, IT: Giunti OS*.