



Universidad Autónoma
de Madrid

Biblos-e Archivo
Repositorio Institucional UAM

Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:

This is an **author produced version** of a paper published in:

Language, Cognition and Neuroscience 30.10 (2015): 1291 – 1305

DOI: <http://dx.doi.org/10.1080/23273798.2015.1066509>

Copyright: © 2015 Taylor & Francis

El acceso a la versión del editor puede requerir la suscripción del recurso

Access to the published version may require subscription

Gesture and speech combinations beyond two-word stage in an experimental task

Eva Murillo^a, Nieves Galera^b and Marta Casla^b

^aFacultad de Psicología, Departamento de Psicología Básica II, Universidad Complutense de Madrid, Campus de Somosaguas, Pozuelo de Alarcón 28223, Madrid, Spain; ^bFacultad de Psicología, Departamento de Psicología Evolutiva y de la Educación, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain

Eva Murillo <http://orcid.org/0000-0002-5978-3260>

Marta Casla <https://orcid.org/0000-0002-4723-773X>

(Received 8 July 2014; accepted 16 June 2015)

Abstract

Gestures and speech combinations have a crucial role on early lexical and syntactic development. Nonetheless, little is known about the role of these combinations on language learning beyond the two-word stage. Our aim is to explore how children combine gestures and speech when they start to master syntactic rules. Thirty Spanish children (aged 24–35 months) participated in a task with a “find the odd one” game structure. The complexity of the target picture increased in terms of the relationships between the elements depicted in each image. Children coordinate gestures and words preferably as their main communicative resource, regardless the complexity of the message to convey or their linguistic development. They distribute the semantic load between speech and gesture depending on message complexity. Among all types of gesture–speech combinations produced, reinforcing combinations were related to lexical and syntactic development, and supplementary combinations were related to lexical development.

Keywords: gestures; language development; multimodality; early syntactic development; gesture–speech coordination

Gesture and speech combinations beyond two-word stage in an experimental task

Vocal and motor elements are part of a single communication system from the early stages of language development. The coordination of these elements can be considered as an ontogenetic precursor of the speech–gesture integration seen in adult language (Iverson & Thelen, 1999). In fact, even before children are able to produce words, they coordinate effectively gestures and vocalisations in their communicative attempts. At the end of the first year, children start to produce communicative gestures addressed to adults, such as pointing or reaching, and these gestures are very often accompanied by vocalisations. In particular, nearly 70% of the gestures produced by the first year are coordinated with vocalisations (Cochet & Vauclair, 2010; Dobrich & Scarborough, 1984; Franco & Butterworth, 1996; Iverson & Goldin-Meadow, 2005; Leavens & Hopkins, 1999; Leung & Rheingold, 1981). Around the first year, gestures and vocalisations not only co-occur, but they are also temporally synchronised, showing an adult-like pattern: the gesture onset precedes speech onset, infants align more closely the beginning of the gesture stroke with the speech onset and with the beginning of the accented syllable, and the gesture apex precedes the end of the accented syllable (Esteve-Gibert & Prieto, 2014). Importantly, this coordination seems to have a role on subsequent linguistic development. The coordinate use of gestures, vocalisations and social gaze at 12 months of age is a strong predictor of lexical development at 15 months (Murillo & Belinchón, 2012; Wu & Gros-Louis, 2014). Also infants who coordinate pointing and speech at 12 months show better vocabulary abilities at 18 months

of age (Igualada, Bosch, & Prieto, 2015). Child gesture uses at 14 months of age are predictors of vocabulary size even at 42 months (Rowe, Özçalışkan, & Goldin-Meadow, 2008).

The predictive value of gesture and vocal coordination on subsequent linguistic achievements is not restricted to the period of transition to first words; on the contrary, the role of gestural and vocal coordination evolves through language development. During the one-word period, Capirci, Iverson, Pizzuto, and Volterra (1996) found that the use of gestures and their coordination with words are related with the total vocal production at 20 months of age. The coordination of gestures and words that convey different information is related to lexical development (Morford & Goldin-Meadow, 1992) and predates and predicts the onset of two-word combination (Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005). The ability to combine two communicative elements in a single sentence is considered as the first step to syntactic development. In this sense, gesture with word coordination enables the children to communicate two pieces of information within a single utterance. In this transition to two- word constructions, children produce sentence structures by means of gesture–speech combinations before they produce the same constructions entirely within speech (Özçalışkan & Goldin-Meadow, 2005, 2009). In addition, by 18 months, the number of gesture–speech combinations in which gesture conveyed one idea and speech another one is a strong predictor of verbal sentence complexity at 42 months (Rowe & Goldin-Meadow, 2009).

It seems, thus, that speech and gestures combination plays a role not only on early lexical development, but also on the first steps of syntax development. Nevertheless, little is known about how children use speech–gesture combinations beyond the two-word stage. After this period, some studies address the role of speech and gesture coordination in school-age

children's problem-solving tasks: conservation problems (Church & Goldin-Meadow, 1986), mathematical equivalence problems (Alibali & Goldin-Meadow, 1993; Perry, Church, & Goldin-Meadow, 1988) and balance problems (Pine, Lufkin, & Messer, 2004). However, there is scarce evidence about how children coordinate gesture and speech when they are starting to master syntactic rules. How are children's gestures integrated in their communicative attempts when they are moving beyond the two-word stage?. Considering that the use of gestures lightens cognitive load (Cook, Yip, & Goldin-Meadow, 2012; Ping & Goldin-Meadow, 2010), which is the role of gestures when communicative requirements challenge children's linguistic abilities? Will the type of speech-gesture coordination vary depending on the linguistic complexity of the message to be expressed?

Stefanini, Bello, Caselli, Iverson, and Volterra (2009) explored the use of speech and gestures from 2 to 7 years of age using a naming task. At the beginning of the period of study, almost half of the answers from the children included speech-gesture combinations. However, the use of gestures decreased along this period, and only 10% of the answers of the oldest group (mean age 6; 4), included multimodal coordination. Other studies employing a similar naming task have shown the same trend: although gesture and speech coordination did not disappear from children answers, the use of co-speech gestures decreases after 2 years of age (Huttunen, Pine, Thurnham, & Khan, 2013). It is possible that a picture naming task is difficult for two-year-olds, but it might not be challenging considering older children's cognitive and linguistic abilities. In addition, it should be noted that picture labelling might not require word-combinations.

Some authors have also reported that children rely on gestures to produce the first instance of specific linguistic constructions. But once the children are able to express these constructions entirely with speech and incorporated to their repertoires, they do not use the

gesture to flesh out these constructions (Özçalışkan & Goldin-Meadow, 2009). Mayberry and Nicoladis (2000) claimed that young children begin to produce new kinds of gestures when they start to produce sentence-like utterances (gestures that adults produce in their speech such as iconics and beats). These studies are based on naturalistic observations. The question that arises is whether children will use gesture and speech coordination as the main communicative resource after two years in a linguistically challenging situation.

Furthermore, previous studies focused on the relationships between the information conveyed by gesture and speech when they appear together. Özçalışkan and Goldin-Meadow (2005) classified gesture–speech combinations depending on the relationship of the information conveyed by each communicative element. They considered a *reinforcing* relation when gesture conveyed information that was redundant with speech (e.g. “car” + point at a car). The relation was *disambiguating* when gesture clarified the referent of a pronominal, demonstrative or deictic (e.g. “this” + point at toy). Finally, the relation was considered *supplementary* when gesture added semantic information to the message conveyed in speech (e.g. “daddy” + point at a car).

Several studies have emphasised the role of supplementary coordination on syntactic development. As mentioned above, supplementary coordination at the end of the second year predicts the onset of two-word combinations (Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005), as well as verbal syntactic complexity at 42 months (Rowe & Goldin-Meadow, 2009). According to these authors, by means of supplementary coordination children are able to convey two different pieces of semantic information, that is, sentence-like meanings. Other kinds of combinations such as disambiguating or reinforcing ones do not seem to have a crucial role on syntactic development. However, besides supplementary coordination, other kinds of combinations might coordinate two elements

with different meanings. Combinations of pointing gestures with object labels (e.g. POINT + “car”) are usually interpreted as combinations of elements that convey the same meaning. This interpretation leads to consider deictic gestures and referential words (object labels) as equivalent. However, unlike referential words, deictic gestures do not have a stable referent, and their meaning is context-dependent. During the transition to two-word coordination, deictic elements seem to be mainly conveyed by gestural means, whereas referential elements appear to be linked to verbal means (Pizzuto & Capobianco, 2005). Therefore, we would not consider that a deictic gesture with a referential word convey the same meaning, but two different pieces of information. In this sense, very recent studies are starting to consider the role of these combinations on subsequent linguistics achievements. Cartmill, Hunsicker, and Goldin-Meadow (2014) have found that the age at which children start producing pointing + nouns combinations predicts the onset age for determiner + noun constructions. It is possible, thus, that the information conveyed by means of combinations traditionally considered as reinforcing is not redundant, and can also be relevant on syntactic development.

Considering previous research, our hypotheses can be summarised as follows: (1) despite children’s ability to combine two verbal elements in their constructions, gestural–verbal coordination will still be the main communicative resource when facing a linguistically challenging task; (2) the relationship between the information conveyed by gesture and by verbal elements will change depending on the complexity of the message to convey. That means that the type of gesture–word coordination used by children will be different depending on the complexity of the information that they are trying to transmit; (3) reinforcing combinations

imply the transmission of two pieces of information within the same communicative attempt, therefore they will be positively related to syntactic development.

To test these hypotheses, we designed an experimental task based on the task used by Vilain, Vilain, and Clarke (2013). This task consists in a “find the odd one” game, in which the syntactic structure of the target names is modified. The structure of the task goes a step further of naming tasks used in previous studies (Huttunen et al., 2013; Stefanini et al., 2009), with the aim of eliciting not only object labelling, but the expression of different relations between elements.

Method

Participants

Thirty children (15 girls) aged 24–35 months ($M = 30.73$; $SD = 2.78$) participated in this study. They were contacted by means of day care facilities, and informed consent was obtained from parents upon enrollment. The initial sample consisted of 48 children, of which 13 were excluded due to fussiness or lack of cooperation, and 5 due to developmental or sensory difficulties.

All the participants but two came from monolingual Spanish-speaking homes. Two participants came from bilingual environments, in which one of the parents was a native speaker of Arabic and French. We decided to include them in the sample because their communicative developmental inventory (CDI) scores did not differ from the scores of the monolingual participants.

Parents, who voluntarily agreed to participate, completed the Spanish version of MacArthur CDI from Spain (López-Ornat et al., 2005). Once all data were collected, a brief

report about the CDI results of each child was provided to parents.

Stimuli, design and procedure

The task designed was a “find the odd one” game, based on the task described by Vilain et al. (2013). Each item included five familiar objects pictures, four of them were identical and one, the target picture, was different. The location of the target picture was randomised. All the pictures included in the task depicted familiar objects, and their corresponding verbal labels were all high frequency words for Spanish-speaking children, according to Spanish version of MacArthur CDI (López-Ornat et al., 2005). Trials differed on the basis of the distinguishing element of the target picture: the complexity increased in terms of the relationships between elements depicted in each image. In the first three trials, the target picture was different from others in terms of identity (e.g. a cow vs. four dogs). In the next three trials, the difference was based on colour or size, so it was necessary to convey two elements in order to identify the target. In the trials 7 and 8, the difference was marked by an additional element placed over or under a chair or a table. In the next two trials, all the pictures included two objects, a main object (a table or a chair) and other object placed under or over it. The difference of the target picture was the colour of the element placed on the main object (trial 9) or the location of this object (on the table vs. under the table) for the trial 10. The last two items had a similar structure, but the object placed on the chair or under the table was different in terms of two characteristics: colour and location for trial 11 and colour and size for trial 12. This way, the expected answer, or, in other words, the information to be conveyed to verbally spot the target object, was increasing in complexity. We did not expect to parallel specific syntactic structures, but to increase the complexity of

the information to be conveyed. We classified the items as easy, medium or difficult according to their complexity: we included in the “easy” category the first three items in which the verbal response to spot the target is only a name; in the “medium” complexity category, we included items from 4 to 8, in which the difference was based on a characteristic (colour or size) or on the presence of an object above or below the main element. In this category, the verbal answer (in the case of a perfect verbal response) was defined by two semantic elements. In the “difficult” category, we included the items from 9 to 12. In these items the difference of the target was based on two or more characteristics (colour, size and spatial relationship).

The pictures were presented in a laptop using Power Point presentation. In addition to the 12 test trials, two initial training trials were included so it was made sure that children could follow the task. These trials were not considered in further coding and analyses.

After a warm up period of 5 min play with bubbles, the child was asked to sit down in a chair in front of a table. The laptop was placed on the table at 60 cm from the child, to prevent her from touching the screen. A digital camcorder was used to videotape the sessions, and was placed behind the laptop. The items were presented in the same order for all the children, increasing in difficulty in order to achieve and maintain their interest in the task.

The experimenter showed the first training trial, and explained the task to the child. Then, the second training trial was showed, and the experimenter asked *Y ahora, ¿cuál es distinto?* [*And, now, which one is different?*]. If the child answered correctly, the experimenter continued showing the first test trial. If the child did not show evidence of understanding the task, the experimenter explained it again using the two training trials. After a correct answer in the second training trial, the experimenter continued with the test trials. We considered an answer as correct if children pointed or described verbally the target

objet (named the object, the distinctive feature or the special location of the target). For each test trial, the experimenter asked the same question: *Mira, ¿cuál es distinto?* [*Look, which one is different?*]. If the child did not respond, or the answer was wrong, the experimenter insisted *Mira, ¿cuál es diferente?* [*Look, which one is different?*] After this second chance, the experimenter showed the next item.

Coding

All the answers to the task were coded using ELAN annotation tool <http://tla.mpi.nl/tools/tla-tools/elan/> (Lausberg & Sloetjes, 2009). We did not consider comments or gestures used by children that were not an answer to the experimenter's question. We coded as responses all the utterances produced to describe each item in reply to the question, therefore it was possible to obtain more than a single response for each item.

The answers were first coded in terms of the modality of response (Verbal, Gestural or Verbal + Gestural). They were classified as verbal (V), when the answer consisted on spoken language or vocalisations and did not include any gesture; gesture (G) when the answer was a deictic, symbolic or conventional gesture without any vocal accompaniment; verbal–gestural coordination (V + G) when the answer includes a gesture and a verbal response with some time overlap.

The type of gesture (pointing, reaching, conventional, symbolic or others) was coded following Murillo and Belinchón (2012) criteria (see Table 1). The hand used (left, right or both) when producing a gesture was also coded.

When the response produced by the child included verbal–gestural coordination, the type of relation between both elements was coded following Goldin-Meadow and

Butcher's (2003) and Özçalışkan and Goldin-Meadow's (2005) classification, (supplementary, reinforcing or disambiguating). Gesture–speech combinations were classified as reinforcing when gesture conveyed information that was redundant with speech (e.g. “dog” + point at dog). A disambiguating relation was coded when gesture clarified the referent of a pronominal (e.g. “him” + point at teddy bear), demonstrative (e.g. “that one” + point at fish), or deictic (e.g. “there” + point at table). A supplementary relation was coded when gesture added semantic information to the message conveyed in speech (e.g. “blue” + point to fish). More details about this classification are provided in the results section.

Two independent observers coded the 10% of the recordings. Kappa indexes for interobserver agreement were .71 for modality of response and .72 for type of gesture–speech combination and .87 for type of gesture.

Children were classified according to their raw score on McArthur CDI's vocabulary subtest (Min. = 107; Máx. = 561; Mean = 375.70; S.D. = 137.194). The quartiles of the distribution were obtained and used to assign the children to three different groups. Children under the percentile 25 were included in group Low, children between percentiles 25 and 75 were included in group Medium and children above percentile 75 were included in group High.

The sample was also divided in three groups according to the raw score on the morphosyntactic subtest from the McArthur CDI (Min. = 21; Max. = 102; Mean = 71.07; S.D. = 25.746). As for vocabulary development, the quartiles of the raw scores were obtained. Children under the percentile 25 were included in group Low, children between percentiles 25 and 75 were included in group Medium and children above percentile 75 were included in group High.

INSERT TABLE 1 ABOUT HERE

Results

A series of binomial tests were carried to test whether children were performing above chance as a task comprehension indicator (see [Table 2](#)).

INSERT TABLE 2 ABOUT HERE

There were significantly more correct answers than wrong ones in the task, and in each particular item, except for the last three items. This is not surprising considering that the last three items included the more complex elements configuration. It is important to note that the task is perceptually easy: it is not difficult to find the target object. Nevertheless, the complexity of the linguistic message needed to verbally spot the target object, increases along the task.

A total of 688 responses were obtained among all participants ($M = 23.20$; $SD = 6.47$; $Min. = 8$; $Máx. = 39$). Most utterances coded, 446 (64.8%), involved speech and gesture coordination (V + G) ($M = 14.86$; $SD = 6.99$), whilst 111 (16.13%) were gestures alone (G) ($M = 3.70$; $SD = 5.28$), and 131 (19%) were only verbal (V) ($M = 4.36$; $SD = 4.73$).

[Table 3](#) shows the number of verbal utterances produced by children, the maximum and minimum elements per utterances and the mean length of utterance (MLU) for each participant.

We obtained a positive and significant correlation between the MLU obtained in the experimental task and the MLU provided for parents by means of CDI ($r = .483$; $p = .017$).

INSERT TABLE 3 ABOUT HERE

The gestures observed were mainly pointing gestures with the right hand ($n = 431$), pointing with the left hand ($n = 91$), pointing with both hands ($n = 4$) or with the head ($n = 1$). We observed also reaching with both hands ($n = 3$), symbolic and conventional gestures ($n = 4$) and undefined gestures classified in the category “others” ($n = 23$).

Modality of response and task performance

In order to test whether response modality had an effect on task performance, we conducted a repeated measures ANOVA, with the proportion of right answers as dependent variable and the modality of response (G, V and V + G) as factor. Table 4 shows the mean proportion of right answers according to the modality of response.

We found no differences on the proportion of correct answers when producing V + G, V or G responses, $F(2, 22) = 1.658$, $p = .213$.

In the next analyses we only considered the data of the correct answers ($n = 447$). These data included 80 verbal answers, 66 gestural answers and 301 verbal–gestural combinations.

Modality of response depending on message complexity, vocabulary level and syntactic level

To explore the influence of message’s complexity on the modality of response chosen by children, we compared the frequency of each answer modality for the items classified as easy, medium and difficult. As complexity groups did not have the same number of items, we obtained the rate by item, dividing the frequency of response of every group of complexity by the number of items that compose it. We did so in all the analyses involving the message complexity variable.

Figure 1 shows the mean rate of responses for each modality considering the complexity

of the message.

In order to explore the influence of the message complexity on the modality of response we conducted a repeated measures ANOVA. Frequency per item was the dependent variable and the modality of response (V, G, or V + G) and the message complexity (easy, medium, difficult) were the factors. As can be seen in [Figure 1](#), results show a main effect of the modality of response, $F(2, 58) = 21.179$; $p > .001$; $\eta^2 = .422$, but no effect of the complexity of the message and no interaction were found.

It seems, therefore, that gestural–verbal coordination is the main communicative resource employed to solve the task, regardless the complexity of the message to convey. It could be that the coordination of speech and gesture depends on children’s linguistic abilities. To investigate this point, we conducted a repeated measures ANOVA, with vocabulary level (low, medium and high) as between-subjects factor and modality of response (V, G, V + G) as within-subjects factor. Frequency of response was the dependent variable. [Figure 2](#) shows mean frequencies by modality of response for each vocabulary level group.

We obviously found a main effect of the modality of response, but no effect of vocabulary group, $F(2, 27) = 1.260$, $p = .30$, neither the interaction, $F(2, 54) = 1.803$, $p = .142$, were found.

The same analyses as above were performed to investigate whether the modality of response is mediated by children’s syntactic development. [Figure 3](#) shows mean frequencies by modality of response for each morphosyntactic level group.

We conducted a repeated measures ANOVA taking modality of response (G, V, V + G) as within-subjects factor, morphosyntactic development (low, medium and high) as between-subjects factor, and frequency per item as dependent variable. We found no effect of morphosyntactic group, $F(2, 27) = 1.49$, $p = .242$, and no interaction effect between

modality of response and syntactic development, $F(2, 27) = .634, p = .538$.

INSERT FIGURE 2 ABOUT HERE

Results analysed so far show that 2;6 to 3;6 year olds do produce speech–gesture combinations regardless of their level of vocabulary and syntax. Nevertheless, it could be that the kind of coordination varies depending on the complexity of the information to convey. To investigate this, we conducted three types of analysis. The first one focuses on the semantic elements that children use in order to describe the target. The second one focuses on the type of linguistic constructions that children produce with the gestures when the message increases its complexity. The third analysis addresses the type of relationship between the information transmitted through speech and the information transmitted through gesture. In these analyses we only considered children’s responses composed by gestures and speech combination.

Semantic elements on gesture–speech combination

To analyse how children use different semantic elements in the combinations of gestures and speech, we determined what semantic elements would define the target in the case of a perfect verbal response. [Table 5](#) shows the semantic elements defined for each item.

INSERT FIGURE 3 ABOUT HERE

INSERT TABLE 5 ABOUT HERE

Considering this classification, we scored the responses of children by giving one point for each semantic element explicitly produced in speech in every trial. We scored also as one if children mentioned the spatial location of the target; for example a sentence-like *el coche que está arriba* (*the car that is up there*) had a score of two (*coche* + *arriba*; *car* + *up there*). In these analyses we only considered children's responses composed by gestures and speech combination.

We divided the score by the number of semantic elements expected in each item. Thus, we obtained a semantic coefficient that allowed us to analyse the semantic load transmitted verbally. A coefficient of one means that the target is perfectly described verbally, and therefore the gesture accompanies or reinforces the verbal message. On the contrary, a coefficient near zero means that semantic meaning transmission is largely based on the gestural component (although there are verbal elements in the communicative attempt).

Then we conducted a repeated measures ANOVA taking the semantic coefficient as the dependent variable and the complexity of the message (easy, medium and difficult) as factor. We found an effect of the message complexity on the semantic coefficient, $F(2, 38) = 16.415$; $p < .001$; $\eta^2 = .463$, showing that semantic coefficient is lower with difficult items than with the easy (.11 vs. .54, $p < .001$) and the medium items (.11 vs. .31, $p < .001$) (see [Figure 4](#)).

The differences found between easy and medium items did not reach statistical significance (.31 vs. .54, $p = .064$). It seems, thus, that in the gesture–speech combinations analysed, the semantic load transmitted verbally decreases when the complexity of the message increases. Labelling the target in the easy trials is really easy for the children of this sample. However, when they have to describe several semantic elements, they rely an

important part of the message to the gestural component. Considering this, the next question that arises is which semantic elements are used and how are they combined when the message to convey is more and more challenging. To explore this question we analysed children's linguistic constructions produced with gestures as shown below.

INSERT FIGURE 4 ABOUT HERE

Type of linguistic constructions depending on the complexity of the message to convey

We classified children's responses according to the number and type of verbal elements produced as is shown in [Table 6](#).

INSERT TABLE 6 ABOUT HERE

To investigate whether specific linguistic constructions are coordinated with gestures depending on the complexity of the message, we conducted a repeated measures ANOVA with complexity (easy, medium and high) and type of linguistic construction (deictic, name, qualifier, deictic + name, two-elements, more than two elements) as factors. The frequency per item was the dependent variable. Results show a main effect of the type of construction (sphericity was not verified, so Greenhouse-Geisser correction was applied) $F(1.948, 56.479) = 7.425, p = .001; \eta^2 = .204$. As can be seen in [Table 7](#), "Name" constructions are more frequent than "qualifier" ($p = .003$), "deictic + name" ($p = .001$), "two-elements" constructions ($p = .002$) and "more than two elements" constructions ($p = .002$).

INSERT TABLE 7 ABOUT HERE

We also found an interaction effect between the type of construction and the complexity of the message (Greenhouse-Geisser correction was also applied to account for sphericity assumption) $F(4.677, 135.646) = 7.714, p < .001, \eta^2 = .210$. Results show that "name" constructions are more frequently produced in the easy items than in the medium ($p = .041$)

and in the difficult trials ($p < .001$). “Qualifier” constructions are less frequently found in the easy items than in the medium ($p = .002$) or in the difficult ones ($p = .026$). This is not surprising considering that in the easy items no qualifier is needed to describe the target object.

Comparing the different constructions together, we can see that in the easy items, “name” constructions are more frequent than “qualifier” ($p < .001$), “deictic + name” ($p < .001$), “two-elements” ($p < .001$) and “more than two elements” ($p < .001$) constructions.

In the medium items we found no differences between the types of constructions. In the difficult items “deictic” constructions are more frequent than “deictic + name” ($p = .048$) and “more than two elements” constructions ($p = .022$).

Type of gesture–word coordination depending on the complexity of the information to convey

Our last analysis was focused on the type of relationship between the information conveyed by gestures and speech. Based on Goldin-Meadow and Butcher’s (2003) and Özçalışkan and Goldin-Meadow’s (2005) classifications, we coded the responses as follows: we considered a disambiguating combination when gesture clarified the referent of a pronominal, (e.g. “him” + point at teddy bear), demonstrative (e.g. “that one” + point at fish), or deictic (e.g. “there” + point at table). We considered the gesture– speech coordination as reinforcing when the children labelled the whole object or one of the elements depicted in the item while pointing it (e.g. “the dog” + point; “the ball” + point). We considered as supplementary those combinations in which gesture added semantic information to the message conveyed in speech (an action, an attribute or an additional argument) (e.g. “that one that has a ball” + point; “the big one is different” + point). We included in this category those responses in which children named the distinctive element of the target without labelling the object

(e.g. “blue” + point). Some examples of answer’s classifications can be found in [Table 8](#).

INSERT TABLE 8 ABOUT HERE

We conducted a repeated measures ANOVA taking as factors the complexity of the message (easy, medium and difficult) and the type of coordination (disambiguating, reinforcing, supplementary). The frequency per item was the dependent variable. We found no main effect of complexity or type of combination on the rate of response, but we found an interaction effect between the factors, $F(4,112) = 5.690$, $p < .001$; $\eta^2 = .169$. As can be seen in [Figure 5](#), there are no differences on the distributions of disambiguating combinations according to the complexity of the message.

INSERT FIGURE 5 ABOUT HERE

However, reinforcing combinations are more frequent in the easy trials than in those of medium complexity (.425 vs. .221; $p = .046$) or in the difficult items (.425 vs. .155; $p = .004$). By contrast, supplementary combinations are more frequent in the items of medium complexity than in the easy items (.352 vs. .069; $p = .001$). There is no differences on supplementary combination frequency neither between medium and difficult items (.190 vs. .352; $p = .178$) nor between easy and difficult ones (.069 vs. .190; $p = .211$).

Relationship between linguistic development and type of gesture–speech coordination

In order to test whether specific types of gesture–speech combinations are related with linguistic development, the frequencies of supplementary, reinforcing and disambiguating combinations were correlated with vocabulary and morphosyntactic CDI raw scores. [Table 9](#) shows the values of these correlations, and [Figures 6–8](#) the scatter plots of each of them.

As can be seen in [Table 9](#), reinforcing combinations were strongly related to the vocabulary and morphosyntactic measures provided by parents. Supplementary combinations appeared also positively related to vocabulary development, but not with morphosyntactic measures.

INSERT TABLE 9 ABOUT HERE

Discussion

The task that children faced in the present study could be successfully disentangled without speech production. Similarly, it could be solved only by means of speech. However, children in our sample chose gestural–verbal combinations as the main communicative resource in their attempts to identify the target picture. Multimodal responses were preferably used regardless the complexity of the message to convey and child’s vocabulary and morphosyntactic levels. This result supports approaches that conceive gesture and speech as an integrated system during the course of language development, as observed in adult language (Kendon, [2004](#); McNeill, [1992](#)). The results obtained in this study did not support the idea of a general decrease in the use of gestures when the children start to produce two-word sentences as previously suggested (Huttunen et al., [2013](#); Stefanini et al., [2009](#)). On the contrary, it seems that gesture–speech coordination emerges as the main communicative resource when the task is linguistically challenging, even when children are able to combine words in their spontaneous productions. Children produced gesture–speech combinations facing all levels of complexity. If gestures had a mere supportive role, or served to express what child is not able to put in words, children would have produced them more frequently in order to describe items with higher levels of complexity. However, this not seems to be

the case. Gesture and speech are part of a communicative and linguistic system that evolves depending on its own characteristics and the context demands. Previous studies have highlighted the role of gesture–speech combinations in fleshing out new linguistic constructions and the decrease of those combinations when children are able to express the same meanings only by means of words (Özçalışkan & Goldin-Meadow, 2009). However, we found gesture–speech combinations from the simplest to the most challenging items. This means that children in our study have used gesture–speech combination when the task required simple constructions that they already had in their repertoire and when the task required more complex constructions.

Our results show that the use of gesture–speech combinations is not directly related with linguistic competence (vocabulary or morphosyntactic development), as can be seen in Figures 2 and 3.

It seems, nevertheless, that the role of these combinations changes depending on the complexity of the message to convey. In our task, when the message is easy to convey, gesture seems to have a reinforcing role with respect to speech. By contrast, as can be seen in Figure 4, gesture and speech share the semantic load when the message becomes more and more complex. We found a wider variety of linguistic constructions with the items of medium complexity. In these items we also found more supplementary combinations than in the easy items, conveying different parts of the message by means of gesture and words. In the difficult items, it seems that the message is too complex to be expressed only by verbal means, since they rely on gesture to transmit the referential load of the message. This suggests that the semantic load leans to the gesture when the complexity of the message increases, that is, when more complex linguistic constructions are needed. This is consistent

with Özçalışkan and Goldin-Meadow's (2009) findings, but adds a nuance: the use of gestures do not decrease with constructions that children have incorporated in their verbal repertoire, but rather they adjust the kind of combinations that employ, and the semantic load of the gesture increases with complexity. This suggests that children can flexibly adjust the use of gestures and speech to the complexity of the relations to be expressed.

Our findings suggest that reinforcing combinations are not redundant. That is, there are two different pieces of information coordination. In our view, pointing gestures to objects are not equivalent to the verbal labels of the objects. Pointing to an object while saying its name, implies the coordination of two elements. This kind of coordination happens to be positively related with syntactic development taken from independent morphosyntactic measures. These findings are in line with recent studies that start to emphasise the role of pointing + name coordination on early syntactic development (Cartmill et al., 2014).

By contrast, supplementary combinations appear related with vocabulary development, but not with morphosyntax development as one would initially have expected. It might be necessary a certain level of vocabulary development to start combining two meanings in a single multimodal communicative attempt. It is possible than in later stages of grammar development, we could find a clearer relationship between supplementary combinations and morphosyntax development. Our data do not allow us to draw any conclusion at this regard. Further research is needed to clarify this point.

Summarising, our results suggest that children that are moving beyond the two-word stage integrate in our task speech and gesture as their main communicative resource, regardless the complexity of the message to transmit or their vocabulary or syntactic development. They flexibly adjust the type of information to convey by means of gestures and speech depending on the complexity of the message to be conveyed.

How a child combines gestures and speech can be an observable indicator of lexical and syntactic development, and provide us with clues about how these processes are taking place in the early stages of syntax development. For example, recent findings suggest that children with specific language impairment aged 5–10 years, rely more on gestures than their typically developing peers, but there is no difference in the frequency of non-redundant gesture–speech combinations (Mainela-Arnold, Alibali, Hostetter, & Evans, 2014). Our results suggest that gesture–speech combinations considered as redundant (i.e. the “reinforcing” combinations of our study) are also related to language development. Finding differences in gesture–speech combination at initial stages of syntax development may be useful for defining predictors of later delays or language disorders.

Although the task used in the present study goes a step further than previous labelling tasks, including the description of relationships between objects, it has some limitations to be considered. For example, the pictures are all static objects, so children do not have to describe movements or actions. The description of a movement or an action can elicit a wider variety of gestures and linguistic constructions. The absence of actions limits the use of verbs, complicating an analysis in terms of arguments and predicates as the one developed by Özçalışkan and Goldin-Meadow (2005, 2009). This fact should be taken into account in further research.

Our results complement those obtained in studies based on naturalistic observations, and provide support to the conception of gestural and vocal elements as continuously interacting during the course of language development (Iverson & Thelen, 1999). This interaction between gestural and vocal components may have a facilitating role not only on lexical, but also on syntactic development beyond the two-word stage. Further studies are needed to explore the role of these gesture–word combinations in stages in which children are able to

produce more complex sentences.

Acknowledgments

We thank all the staff and families from E.I Caperucita and E.I. Los Cedros for their invaluable cooperation and support.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Spanish Ministry of Science and Innovation under the Project PSI2013-44250-P.

References

- Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hand reveal about a child's state of mind. *Cognitive Psychology*, 25, 468–523. doi:10.1006/cogp.1993.1012
- Capirci, O., Iverson, J. M., Pizzuto, E., & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child Language*, 23, 645–673. doi:10.1017/S0305000900008989
- Cartmill, E. A., Hunsicker, D., & Goldin-Meadow, S. (2014). Pointing and naming are not redundant: Children use gesture to modify nouns before they modify nouns in speech. *Developmental Psychology*, 50, 1660–1666. doi:0.1037/a0036003
- Church, R., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23, 43–71.
- Cochet, H., & Vauclair, J. (2010). Pointing gestures produced by toddlers from 15 to 30 months:

Different functions, hand shapes and laterality patterns. *Infant Behavior and Development*, 33, 431–441. doi:10.1016/j.infbeh.2010.04.

009

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, 27, 594–610. doi:10.1080/01690965.2011.

567074

Dobrich, W., & Scarborough, H. S. (1984). Form and function in early communication: Language and pointing gestures. *Journal of Experimental Child Psychology*, 38, 475–490. doi:10.1016/0022-0965(84)90090-0

Esteve-Gibert, N., & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication*, 57, 301–316. doi:10.1016/j.specom.2013.06.006

Franco, F., & Butterworth, G. (1996). Pointing and social awareness: Declaring and requesting in the second year. *Journal of Child Language*, 23, 307–336. doi:10.1017/S0305000900008813

Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two- word speech in young children. In Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 85–107). Hillsdale, NJ: Erlbaum.

Huttunen, K. H., Pine, K. J., Thurnham, A. J., & Khan, C. (2013). The changing role of gesture in linguistic development: A developmental trajectory and a cross-cultural comparison between British and Finnish children. *Journal of Psycholinguistic Research*, 42, 81–101. doi:10.1007/s10936-012-9205-7

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to

- multimodal communicative strategies at 12 months. *Infant Behavior and Development*, 39, 42–52. doi:10.1016/j.infbeh.2015.02.004
- Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16, 367–371. doi:10.1111/j.0956-7976.2005.01542.x
- Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 6(11–12), 19–40.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. New York: Cambridge University Press.
- Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers*, 41, 841–849. doi:10.3758/BRM.41.3.591
- Leavens, D. A., & Hopkins, W. D. (1999). The whole-hand point: The structure and function of pointing from a comparative perspective. *Journal of Comparative Psychology*, 113, 417–425. doi:10.1037//0735-7036.113.4.417
- Leung, E. H., & Rheingold, H. L. (1981). Development of pointing as a social gesture. *Developmental Psychology*, 17, 215–220. doi:10.1037/0012-1649.17.2.215
- López-Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S. & Martínez, M. (2005). *MacArthur: Inventario de desarrollo comunicativo* [MacArthur: Communicative developmental inventory]. Madrid: TEA Ediciones.
- Mainela-Arnold, E., Alibali, M. W., Hostetter, A. B., & Evans, J. L. (2014). Gesture–speech integration in children with specific language impairment. *International Journal of Language & Communication Disorders*, 49, 761–770. doi:10.1111/1460-6984.12115

- Mayberry, R. I., & Nicoladis, E. (2000). Gesture reflects language development evidence from bilingual children. *Current Directions in Psychological Science*, 9(6), 192–196. doi:10.1111/1467-8721.00092
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- Morford, M., & Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language*, 19, 559–580. doi:10.1017/S0305000900011569
- Murillo, E., & Belinchón, M. (2012). Gestural-vocal coordination: Longitudinal changes and predictive value on early lexical development. *Gesture*, 12(1), 16–39. doi:10.1075/gest.12.1.02mur
- Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, 96 (3), B101–B113. doi:10.1016/j.cognition.2005.01.001
- Özçalışkan, Ş., & Goldin-Meadow, S. (2009). When gesture- speech combinations do and do not index linguistic change. *Language and Cognitive Processes*, 24(2), 190–217. doi:10.1080/01690960801956911
- Perry, M., Church, R., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, 3, 359–400. doi:10.1016/0885-2014(88)90021-4
- Pine, K. J., Lufkin, N., & Messer, D. (2004). More gestures than answers: Children learning about balance. *Developmental Psychology*, 40, 1059–1067. doi:10.1037/0012-1649.40.6.1059
- Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, 34, 602–619. doi:10.1111/j.1551-6709.2010.01102.x

- Pizzuto, E., & Capobianco, M. (2005). The link (and differences) between deixis and symbols in children's early gestural-vocal system. *Gesture*, 5(1–2), 179–199. doi:10.1075/gest.5.1.13piz
- Rowe, M., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language*, 28(2), 182–199. doi:10.1177/0142723707088310
- Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science*, 12, 182–187. doi:10.1111/j.1467-7687.2008.00764.x
- Stefanini, S., Bello, A., Caselli, M. C., Iverson, J. M., & Volterra, V. (2009). Co-speech gestures in a naming task: Developmental data. *Language and Cognitive Processes*, 24(2), 168–189. doi:10.1080/01690960802187755
- Vilain, C., Vilain, A., & Clarke, J. (2013). 'The pig with the pink hat': An experimental study on speech/gesture coordination during development. Proceedings of the TIGER Conference, Tilburg.
- Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language*, 34(1), 72–90. doi:10.1177/0142723714521925

Table 1. Gesture's definitions adapted from Murillo and Belinchón (2012).

Gesture Definition	
Point	Extension of the arm and index finger
Reach	Arm is extended, palm usually down, hand open and fingers straight
Symbolic	Gestures that symbolically represent objects, events, desires and conditions
Conventional etc.	Say hello or bye with the hands, clapping,

Table 2. Proportion of correct answers to each item and p values from the binomial tests.

Item	Target identification	
	Correct	p
1	.80	.001
2	.77	.005
3	.93	<.001
4	.80	.001
5	.77	.005
6	.83	<.001
7	.87	<.001
8	.70	.043
9	.77	.005
10	.63	.20
11	.53	.856
12	.43	.585

Table 3. Number of verbal utterances, MLU, maximum and minimum number of elements per utterance.

Participant	Verbal utterances (<i>N</i>)	MLU (SD)	Max.	Min.
1	16	1.50 (.73)	3	1
2	24	2.37 (1.43)	6	1
3	5	2 (0)	2	2
4	21	2 (1.22)	5	1
5	14	2.14 (.94)	4	1
6	15	2.13 (.74)	4	1
7	9	2.11 (1.53)	5	1
8	1	1 (0)	1	1
9	12	1.25 (.62)	3	1
10	18	2.5 (1.20)	5	1
11	3	3.66 (.57)	4	3
12	8	2.37 (1.68)	6	1
13	11	2.09 (.70)	3	1
14	9	1.44 (.52)	2	1
15	14	1.28 (.61)	3	1
16	10	2.20 (.91)	3	1
17	13	1.53 (.66)	3	1
18	16	1.81 (.40)	2	1
19	15	1.46 (.83)	4	1
20	8	1.50 (.75)	3	1
21	9	1.44 (.72)	3	1
22	26	2.26 (1.53)	6	1
23	8	4.25 (2.60)	10	2
24	14	1.64 (1.15)	5	1
25	27	1 (0)	1	1
26	3	2 (0)	2	2
27	9	2.33 (.86)	4	1
28	12	1.16 (.38)	2	1
29	1	1 (0)	1	1
30	7	1 (0)	1	1
Total	358	1.87 (1.16)	10	1

Table 4. Mean proportion and standard deviation of right answers depending on the modality of response.

Modality	Mean	SD
Gesture	.44	.34
Verbal	.54	.28
Gesture + verbal	.64	.22

Table 5. Semantic elements that would define the target in the case of a perfect verbal response.

Complex	Item	Semantic elements description	Number of semantic elements
Easy	1	Oso (<i>bear</i>)	1
	2	Pez (<i>fish</i>)	1
	3	Vaca (<i>cow</i>)	1
Medium	4	Coche + blanco (<i>car + white</i>)	2
	5	Pez + azul (<i>fish + blue</i>)	2
	6	Perro + grande (<i>dog + big</i>)	2
	7	Silla + pelota [encima] (<i>chair + ball [above]</i>)	2
	8	Mesa + tren [debajo] (<i>table + train [below]</i>)	2
Difficult	9	Silla + pelota + amarilla + encima (<i>chair + ball + yellow + on</i>)	4
	10	Mesa + tren + grande + encima (<i>table + train + big + on</i>)	4
	11	Mesa + tren + gris + encima (<i>table + train + gray + on</i>)	4
	12	Silla + tren + gris + pequeño (<i>chair + train + gray + small</i>)	4

Table 6. Linguistic constructions' classification according to the number and type of elements produced.

Categories	Description	Example
1. Deictic	Deictic	"Ése" (<i>this one</i>)
2. Name	Name	"El perro" (<i>the dog</i>)
3. Qualifier	Adjective/adverb	"El azul" (<i>the blue one</i>)
4. Deictic + name	Deictic + name	"Ese oso" (<i>that bear</i>)
5. Two-elements constructions	Name + adjective	"Pez azul" (<i>blue fish</i>)
	Name + name	"La silla de pelota" (<i>the chair of the ball</i>)
	Adv + deictic	"El distinto ése" (<i>that different</i>)
	Adjective + adjective	"El marrón pequeño" (<i>the little brown one</i>)
	Verb + adverb	"Es distinto" (<i>((it) is different)</i>)/"está debajo" (<i>((it) is below)</i>)
	Verb + name	"Tiene pelota" (<i>((it) has ball)</i>)
6. More than two elements constructions	Deictic + name + adjective	"Ese coche blanco" (<i>this white car</i>)
	Deictic + name + adverb	"Esos trenes muchos" (<i>this many trains</i>)
	Deictic + verb + adjective	"Éste es grande" (<i>this is big</i>)
	Deictic + name + name	"Éste tren en una mesa" (<i>this train on a chair</i>)
	Four elements or more	"Ésa es la pelota amarilla y está en la silla" (<i>this is the yellow ball and is in the chair</i>)

Table 7. Mean rate of linguistic constructions depending on message complexity.

Linguistic construction	Message complexity					
	Easy		Medium		Difficult	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Deictic	.30	.58	.29	.52	.35	.47
Name	.53	.44	.31	.37	.06	.13
Qualifier	.00	.00	.17	.24	.13	.24
Deictic + name	.02	.08	.09	.23	.08	.18
Two-elements	.07	.16	.11	.18	.12	.25
More than two elements	.03	.10	.09	.16	.08	.21

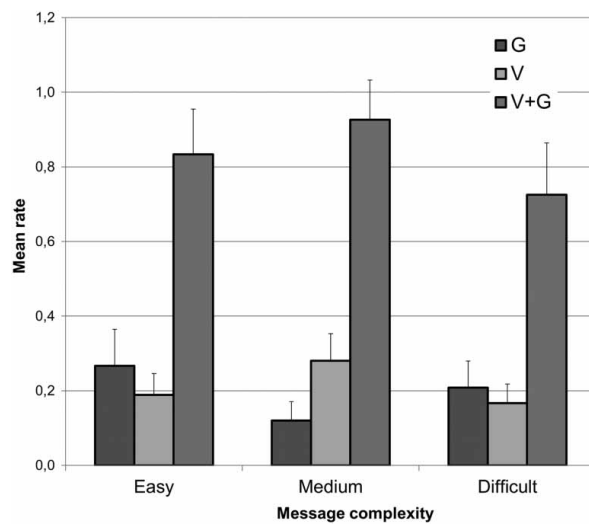


Figure 1. Mean rate of modality of response by message's complexity.

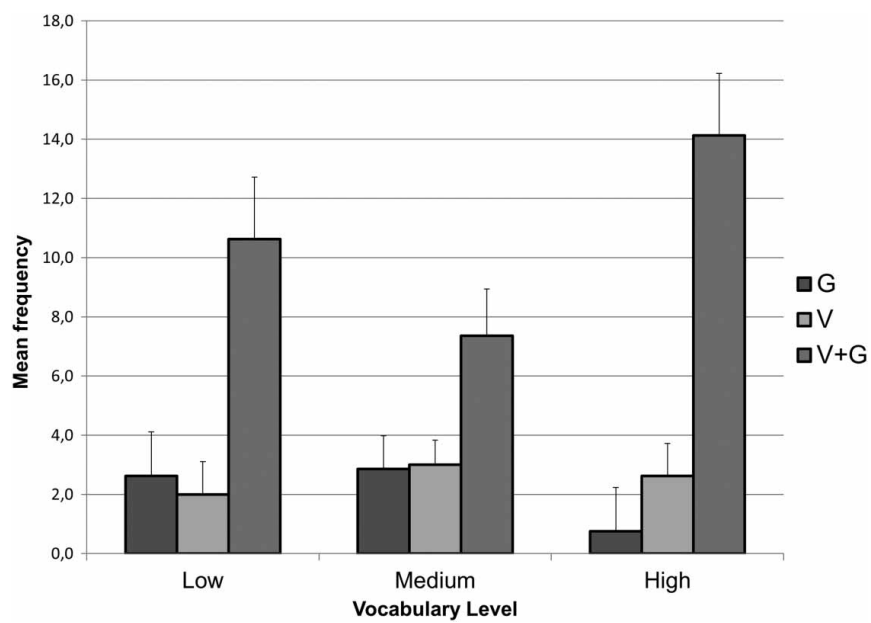


Figure 2. Mean frequency of modality of response for each level of vocabulary development (Spanish version of CDI).

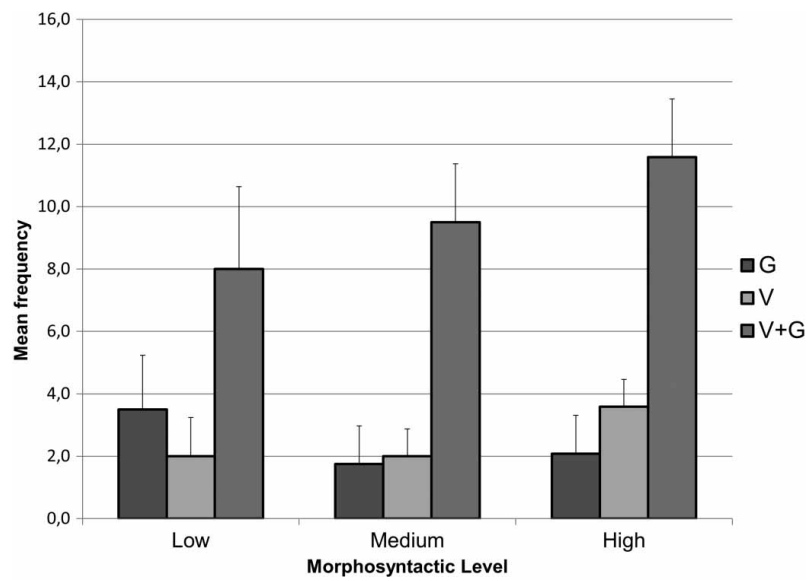


Figure 3. Mean frequency of modality of response for each level of syntactic development (Spanish version of CDI).

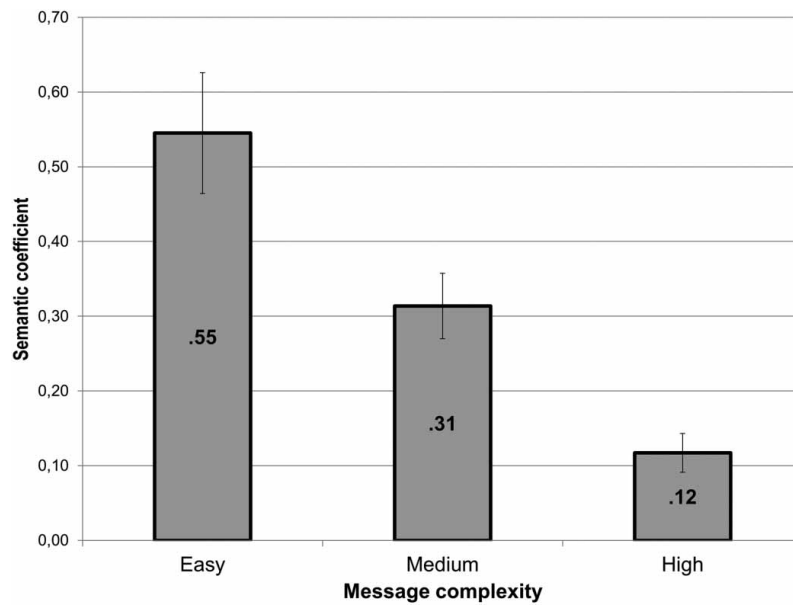


Figure 4. Semantic coefficient according to message complexity.

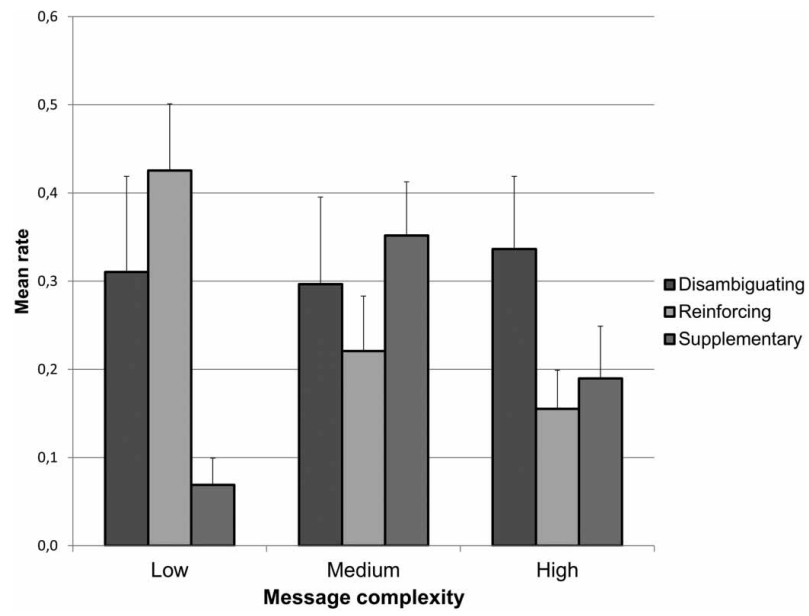


Figure 5. Mean rate of each type of $V + G$ coordination depending on message complexity.

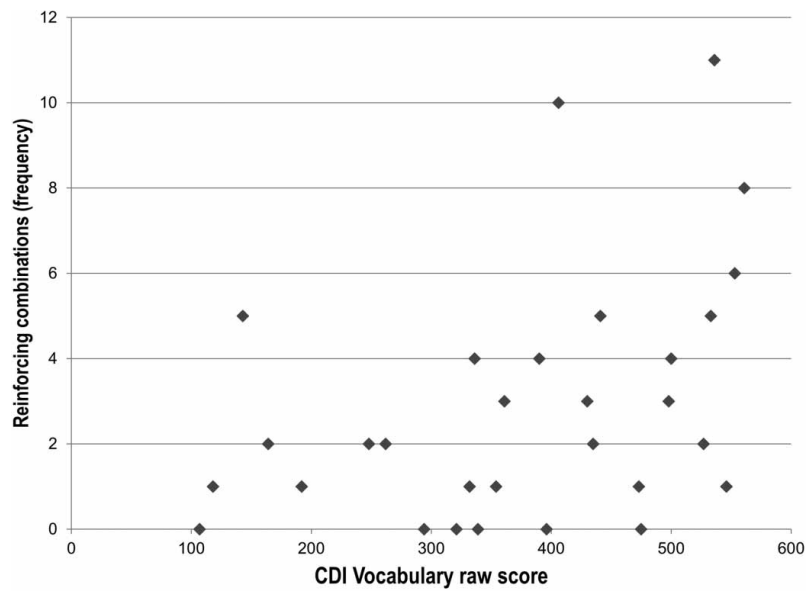


Figure 6. Scatter plot of vocabulary CDI scores by reinforcing combinations frequency.

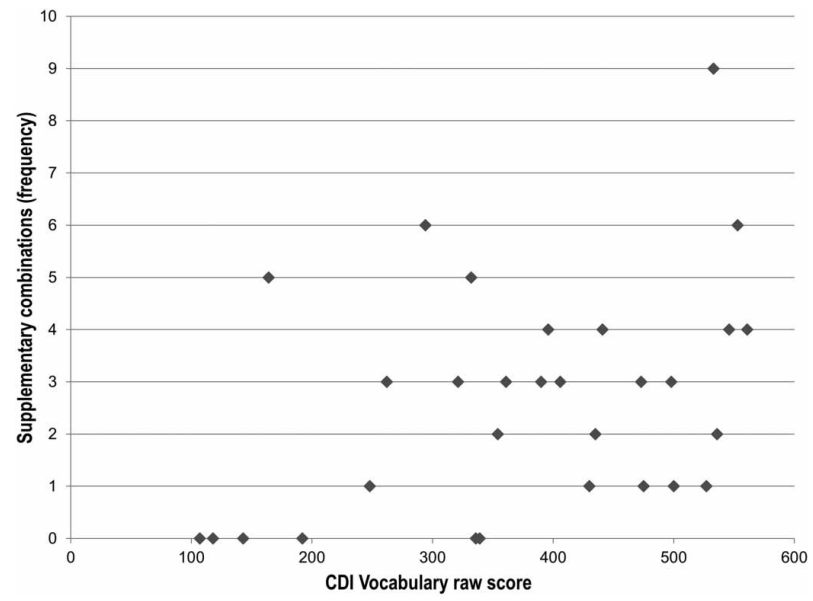


Figure 7. Scatter plot of vocabulary CDI scores by supplementary combinations frequency.

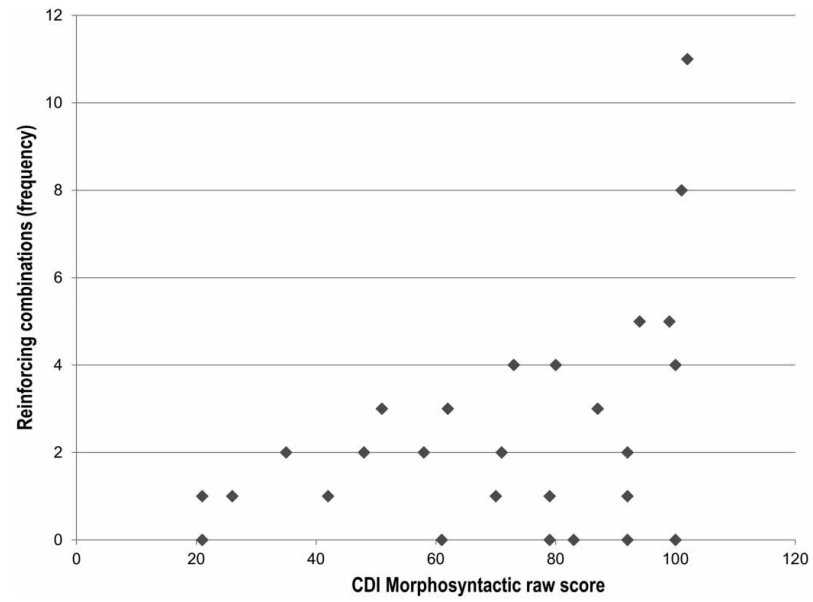


Figure 8. Scatter plot of morphosyntactic CDI scores by reinforcing combinations frequency

