Revisiting evidence for lexicalized word order in young children

FRANCK, Julie, LASSOTTA, Romy

Abstract

One major controversy in the field of language development concerns the nature of children’s early grammatical knowledge. This paper focuses on the early representation of word order. It questions the validity of the results obtained with the Weird Word Order methodology (Akhtar, 1999) in which children are presented with ungrammatical sentences. These results have previously been considered as major evidence for the constructivist, usage-based approach to word order development according to which young children initially encode word order as a verb-specific lexical property which only slowly develops into abstract knowledge at age 3 or 4 (e.g., Abbot-Smith et al., 2001; Matthews et al., 2005, 2007). The critical review presented here addresses various problems with the results and their interpretation. The discussion questions the relationship between theory and data as well as methodological issues related to the number of observations and the discarding of data not missing at random. It is argued that the data not only fail to support the constructivist account, but they actually bring evidence for the alternative […]

Reference


DOI: 10.1016/j.lingua.2011.11.010
Review

Revisiting evidence for lexicalized word order in young children

Julie Franck, Romy Lassotta

University of Geneva, Switzerland

A B S T R A C T

One major controversy in the field of language development concerns the nature of children's early grammatical knowledge. This paper focuses on the early representation of word order. It questions the validity of the results obtained with the Weird Word Order methodology (Akhtar, 1999) in which children are presented with ungrammatical sentences. These results have previously been considered as major evidence for the constructivist, usage-based approach to word order development according to which young children initially encode word order as a verb-specific lexical property which only slowly develops into abstract knowledge at age 3 or 4 (e.g., Abbot-Smith et al., 2001; Matthews et al., 2005, 2007). The critical review presented here addresses various problems with the results and their interpretation. The discussion questions the relationship between theory and data as well as methodological issues related to the small number of observations and the discarding of data not missing at random. It is argued that the data not only fail to support the constructivist account, but they actually bring evidence for the alternative hypothesis according to which children, from early on, represent word order abstractly.

© 2011 Published by Elsevier B.V.
1. Introduction

Language acquisition is not as much of a challenge for the child who is learning to speak without any apparent effort, as it is for the researcher who is trying to discover the nature of this learning process. Despite the growing body of both empirical and theoretical evidence concerning the processes involved in the acquisition of phonological and lexical representations, the acquisition of grammar remains highly controversial. Controversies concern the format of the early grammatical representations that the child is acquiring (whether it is a set of abstract symbolic rules or an accumulation of individual exemplars of particular structures), the role of innate constraints in grammar induction from the input (whether induction is filtered out by internal constraints or operates solely on the basis of the input), and, if such constraints are acknowledged, the nature of these constraints (whether they are specific to language or domain-general, in the form of cognitive or perceptual constraints). This paper focuses on the controversy about the format by which young children represent one major property of the grammar of their language: word order.

2. The debate

The basic structure of syntactic constituents is made of a head, which constitutes the principal element (e.g., the verb in the verb phrase), its complements (e.g., the object) and specifiers (the subject; Chomsky, 1981). The order of the head and the complement is generally fixed for a given category in a language (i.e., a language has Verb–Object or Object–Verb order). Moreover, a language tends to choose a single order across categories (Greenberg, 1963), even though coherence holds only statistically, not absolutely across languages. In the Principles and Parameters model, this aspect of variation in word order is captured by the ‘head direction parameter’ (e.g., Chomsky, 1995).

When do children develop knowledge of the word order of their language? Acquisition research has shown that word order is already represented when children start to combine words into sentences. Analyses of early sentences, arising around 18–24 months, show that word order mistakes are hardly ever made (e.g., Bloom, 1991; Brown, 1973; Lust and Wakayama, 1981; Tomasello, 1992). But evidence suggests that this knowledge is actually encoded much earlier. Head direction correlates with cues in the speech signal that infants are sensitive to and which were found to play a critical role in the bootstrap of word order. First, head direction correlates with rhythmic properties of the language due to the fact that within phonological phrases, prominence always falls on the complement. Hence, while prominence falls on the right edge of constituents in head-complement languages, it falls on the left in complement-head languages (Nespore and Vogel, 1986). Babies as early as 3 months old were found to rely on this cue to discriminate between languages with different head direction (Christophe et al., 2003). Preference for the head direction’s prosodic patterns of the input language was argued to arise around 14 months, as German infants at that age were found to prefer NV phrases, following German’s canonical word order (SOV), to VN phrases (Hofmann et al., 2003). Second, head direction correlates with statistical regularities in the distribution of grammatical morphemes in the sentence. Whereas head-complement languages typically have their grammatical morphemes in initial position in the sentence, complement-head languages have them in final position (Gervain et al., 2008). Gervain and colleagues found that 8 month old Italian infants prefer an artificial language with frequent words (i.e., typically grammatical morphemes) situated in utterance-initial position in the sentences while Japanese infants prefer one with frequent words in final position. Hence, as early as 8 month old infants appear to develop knowledge of at least some of the surface properties that correlate with the word order of their mother tongue. All these observations were argued to support the view that acquisition of word order relies on a mechanism of ‘parameter setting’ (Gib森 and Wexler, 1994; Pinker, 1994) by which the infant, on the basis of prosodic and statistical cues present in the input, rapidly develops an abstract representation of head direction (e.g., Christophe et al., 2003; Gervain et al., 2008). In this view, children develop abstract knowledge of surface properties of the word order of their language before developing a lexicon.

But when do children start to know that syntactic positions map onto thematic roles at the semantic level? The experimental work by Hirsh-Pasek and Golinkoff using the preferential looking paradigm showed that as early as 14 months, English children interpret the noun phrase following the verb as its object (e.g., Hirsh-Pasek and Golinkoff, 1996). Although this report clearly attests that 14 month olds already use word order to comprehend sentences, it is insufficient to tease apart two theoretical interpretations as to the format in which word order is represented. According to what we will refer to as the grammatical hypothesis, children develop abstract knowledge of the mapping between syntactic positions and thematic roles from very early on. This hypothesis is grounded in the account of language acquisition in terms of ‘parameter setting’ according to which language acquisition amounts to fix the value of a number of universal properties that are subject to parametric variation across languages on the basis of limited input (e.g., Radford, 1990; Clahsen et al., 1996). In this view, child language may only diverge from adult language within the possibilities offered by Universal Grammar. Critical evidence was recently provided by Gertner and colleagues who tested children’s comprehension of sentences involving pseudo-verbs, i.e., verbs that children had never encountered before the experiment (Gertner et al., 2006). In line with the results obtained by Hirsh-Pasek and Golinkoff with known verbs, English children aged 21 months interpreted the argument preceding the pseudo-verb as the agent of the action (e.g., the duck in ‘The duck is gorping the bunny’) while they interpreted the one following the pseudo-verb as the patient (the bunny). Similar results were obtained with French infants aged 19 months (Franck et al., in press).

According to the lexical hypothesis, children learn the appropriate use of grammatical relations on a verb-by-verb basis. In this view, grounded in the constructivist, usage-based account of grammatical development, word order involves a slow and
The paper is addresses the recurrent problem of data analysis in these papers with regard to the since it reports data (and interpretation) goes from data back to theory, raising critical issues in data interpretation. The punch line is Under the assumption of the lexical hypothesis, the 'young' group

3. From theory to data: critical issues with empirical predictions

Different variables are manipulated in WWO studies. The first variable is age: performance of children from two (sometimes three) age groups is typically compared. Under the assumption of the lexical hypothesis, the 'young' group (between 2.2 and 3.6 years old, depending on the study) lacks grammatical knowledge of word order (i.e., has lexical knowledge that gradually turns into grammatical knowledge) while the 'older' group (from 4 years onward) has developed abstract word order representations. Hence, the young group constitutes the critical test case opposing predictions from the lexical hypothesis to predictions from the grammatical hypothesis. The second variable is the lexical status of the verb. Two conditions are usually contrasted: one involving high frequency (familiar) verbs, the other one involving low frequency or novel verbs. A third variable is also manipulated in some papers, namely the grammaticality of the modelled sentences (grammatical vs. ungrammatical model, i.e., SOV or VSO order). A summary of the experimental conditions illustrated with examples across the various papers is provided in Table 1.

Children's performance in the task is scored according to three categories: (1) matches, i.e., reproduction of the word order modelled, (2) mismatches, i.e., modifications of the word order modelled, or corrections constituting a subset of these responses, and (3) miscellaneous responses involving mostly non responses or responses including lexical units that were not in the model.

Three major predictions differentiating the younger and the older group are examined across papers. The first prediction bears on the reproduction effect which is based on the comparison between responses that 'Match' the input word order and responses that 'Mismatch' it. The lexical hypothesis predicts that when confronted with novel verbs, "the youngest children will be willing to spontaneously use the non-SVO orders. They are not expected to switch these orders to SVO. Older children, on the other hand, will either ignore these orders or will actively switch them to SVO, as they will have already formed the generalization that English is an SVO language" (Akhtar, 1999:343).

The second prediction bears on the frequency effect which is based on the comparison between high frequency verbs (familiar) and low frequency verbs (unfamiliar or pseudo verbs). The lexical hypothesis predicts that young children should match ungrammatical orders less if the verb is frequent than if it is infrequent, since they have lexical knowledge of the word order for the former. For the same reason, they will tend to correct ungrammatical orders with high frequency verbs more than with low frequency verbs. In contrast, no frequency effect, or at least a weaker effect is expected in the older group which is assumed to perform verb-generally, independently of lexical items. As Matthews and colleagues remark: "We anticipated that the number of occasions on which the younger children matched the SVO word order would be roughly

---

1 The paper by Abbot-Smith et al. (2008), also based on the WWO methodology, is addressed in section 6 since it reports data (and interpretation) radically different from those reported in the WWO papers reviewed here.

2 Although age is treated as a binary variable in these studies, one should keep in mind that the underlying assumption of the lexical hypothesis is that of a gradual process of language development.

3 As we will discuss it in the next section, virtually all mismatch responses consisted in corrections of ungrammatical orders since children nearly never changed an ungrammatical word order into another ungrammatical order. We nevertheless decided to keep the generic category of mismatch responses rather than 'corrections' since this category was used in the papers discussed.
inversely proportional to the frequency of the verbs they heard and that this effect of verb frequency would gradually diminish with age. (Matthews et al., 2005:124).

The third prediction bears on the grammaticality effect which is based on the comparison between grammatical and ungrammatical input models. Akhtar (1999) and Abbot-Smith et al. (2001) introduced grammaticality as part of their experimental designs and predicted that in the presence of a novel verb, young children would behave similarly whether the model sentence was grammatical or ungrammatical, and that they would tend to match the word order presented in both cases. In contrast, older children were expected to match grammatical orders more than ungrammatical orders and to mismatch-to-correct ungrammatical orders more than the grammatical one. A summary of the predictions is presented in Table 2.

Table 2 provides a preliminary view over the empirical evidence reported in the papers. While some of the predictions are indeed validated by the data, as indicated by a U in the table, others are not, as indicated by a × (see sections 4 and 5 for an in-depth analysis of the data).

More problematically, some of the predictions are clearly incorrect, either because they fail to take into account relevant factors that may play a role in performance, or because they ignore important findings from adults’ psycholinguistics. When confronted with grammatical sentences, the parser of languages like English or French assigns the agent role to the preverbal noun phrase in a declarative sentence, and assigns the patient role to the postverbal noun phrase. But if the noun phrases do not occupy their canonical grammatical positions, the parser cannot assign thematic roles: it has to guess. One source of information that may guide this inferential process is the visual context: the video illustrates who is the agent and who is the patient. If the verb is familiar, knowledge of its argument structure may also influence the inferential process by telling the child whether the two NPs constitute the two arguments of the verb (for transitive verbs) or a single argument in some sort of conjunct phrase (for intransitive verbs, see, e.g., Naigles, 1996, for evidence that argument structure is already finely processed by 2.6 year olds).

But the child’s performance to the task is also influenced by another type of guessing which concerns what the experimenter expects from her: correct the ungrammatical word order or adjust to the experimenter’s weird way of
speaking? The situation is particularly puzzling since the child hears the experimenter speak correctly during all the familiarization phase before the test sentences are presented. The role of the socio-pragmatic context in which the experiment takes place was demonstrated by Chang and colleagues who found that Japanese children matched WWO more often when speaking to a robot than when speaking to the experimenter (Chang et al., 2009). In a recent experiment replicating Matthews et al.’s (2007), we obtained significantly more WWO corrections with a slightly different procedure in which test sentences were presented via a computer voice (rather than the experimenter) and children were asked to describe the scene to a zebra whose eyes were hidden (Franck et al., 2011).

Thus, it is incorrect to predict that if children have grammatical representations, they will correct ungrammatical word orders with novel verbs. The reason is that the presence of abstract grammatical representations of word order is not a sufficient condition to trigger WWO’s correction with novel verbs: in order to correct a WWO, the child has to be able (a) to infer the argument structure of the novel verbs involved in the WWO input sentences and (b) to infer that the experimenter expects her to correct the incorrect structure from the input. Rather, the implication goes the other way around: if children do correct ungrammatical orders (above chance level) with novel verbs, they must have grammatical representations. It is also incorrect to predict that if children do not have grammatical representations, they will reproduce (i.e., match) the ungrammatical word order. Indeed, in order to be able to reproduce it, they must be able to infer (a) and (b) as well. Rather, the prediction is that if children lack grammatical representations, they should not correct ungrammatical word orders.

As for predictions related to the role of lexical frequency in children’s performance, they fail to take into account the literature in adults’ psycholinguistics showing massive influence of lexical frequency in language processing, including syntactic processing (e.g., MacDonald et al., 1994). We argued that when confronted with sentences with ungrammatical word orders, the parser is insufficient to build a semantic representation for these sentences. Lexical information on the verb provides important cues to this process since argument structure allows determining whether and which NPs may be part of the argument structure of the verb. Hence, lexical knowledge is expected to facilitate the processing of WWO sentences. However, this does not tell us what children will produce.

Thus, it is incorrect to predict that children will reproduce WWO less with high frequency (known) verbs: frequent verbs are expected to boost the general response rate, but the actual responses that children will produce (reproduction or correction of the WWO) depend on the inferences they will make, as discussed previously. It is also incorrect to predict that younger children will be more sensitive to frequency than older children, given that the adult’s mature parser, assumed to involve abstract grammatical knowledge, is significantly influenced by frequency (see Fisher, 2002, for an extended argumentation on the interpretation of lexical frequency effects in the literature on language development).

In sum, erroneous predictions have been formulated in the papers. Predictions related to the reproduction effect fail to take into account cognitive factors recruited in performing the task and which are independent of grammar. Predictions related to the frequency effect fail to take into account evidence about language processes in adults showing across-the-board sensitivity to frequency. In section 6, we discuss the consequences of these errors in the interpretation of the data.

4. The problem of missing data: critical issues with data analysis

One systematic aspect of the data sets reported in these studies is the very high rate of missing data of two types: (1) missing children: children who were excluded from the analyses because they never produced a single test verb throughout the whole experiment, and (2) non responses: responses from children who were included in the analyses but which failed to contain the target verb (either because nothing was produced or because another verb was used). We examine these two sorts of missing data in turn.

With respect to missing children, Abbot-Smith et al. (2001) report that 10 out of 42 children tested (i.e., 24%) never produced any of the novel verbs that were presented to them. Matthews et al. (2007) report that out of the 112 children tested, 23 failed “to use a test verb at any point and instead preferred to use alternative, often higher frequency, verbs in their place” (an additional set of 28 children was also discarded because of experimenter error or because they failed to use a multiword utterance using any verb). Similarly, in Matthews et al. (2005) 43 children out of the 139 children tested were not included in the study “due to experimenter error or because they failed to complete the testing session or to produce any multi-word utterance using any test verb”4 (Matthews et al., 2005:124). Hence, it seems that between 20% and 30% of the children in those studies were excluded from the analyses.

The high rate of Missing children would not be a problem if it was not asymmetrically distributed across experimental conditions. Table 1 in Matthews et al. (2007) shows that whereas only 2 children out of 56 (3.6%) failed to produce a test verb in the high frequency condition, 26 failed in the low frequency condition (53.6%). Similarly, in Matthews et al. (2005) only 1 child out of 32 failed to respond in the high frequency condition (3.1%) while 12 failed in the low frequency condition (37.5%). It is likely that this strong imbalance also characterizes the other studies, although they do not provide the distribution of Missing children. Hence, Missing children appear to be distributed according to frequency, a variable manipulated in the experiment.

Out of the children who were included in the analyses, Non responses constitute the vast majority of children’s performance across all experiments, ranging from 46% to 98.8%. Like Missing children, Non responses are not distributed

4 The authors did not break down experimenter errors from problems arising from the children (failing to complete the session or to produce a test verb); nevertheless, this is not an issue for the main point developed here which is that missing children are not missing at random.
equally across conditions. Akhtar (1999) reports 97% and 97.5% Non responses in the two ungrammatical conditions against 93.2% in the grammatical condition at age 2, a difference which is also found at ages 3 and 4. Abbot-Smith et al. (2001) report that some children did not use the target verb at all in one or two of the three experimental conditions. On average, 98.8% Non responses are reported in the ungrammatical VS condition against 95.4% in the corresponding SV grammatical condition. Response rates calculated over the numbers provided in Table 2 of Matthews et al. (2005) show that children aged 2.9 provided 92.8% Non responses in the low frequency condition against 77.5% in the high frequency condition, while a difference of 83.7% vs. 46.9% was found in 3.10 year olds. Matthews et al. (2007) report that two-year-old children failed to respond 82% of the time in the low frequency condition and 63% of the time in the high frequency condition, while the three-year-olds responded 55% and 52% of the time, respectively. Again, Non responses appear to be systematically distributed according to the variables manipulated, i.e., grammaticality, frequency, and age.

In sum, both Missing children and Non responses are what is called in the statistics literature Missing Not At Random (MNAR): they are not evenly distributed across experimental conditions. Therefore, the mechanism underlying missing children and missing responses cannot be ignored. To obtain valid inferences, a model of the data that includes the mechanism responsible for missing children and missing responses is required. Unfortunately, in the MNAR setting it is very rare to know the appropriate model for the missingness mechanism (e.g., Allison, 2001).

How did the authors deal with missing data? All studies except Akhtar (1999) discarded Missing children, that is, calculated mean frequencies on the basis of the subset of children who produced at least one test verb during the experimental session. Although Non responses were taken into account in Akhtar (1999) and Abbot-Smith et al. (2001), they were discarded in Matthews et al. (2005, 2007). In the latter two studies, statistical analyses were conducted over 'proportions', that is, mean frequencies of Match responses calculated over the subset of responses including a test verb (Match + Mismatch). By doing so, Matches and Mismatches are in complementary distribution. As a result, if young and old children produced a similar raw frequency of Match responses (as is the case in most of the studies) but older children produced more Mismatch than Match responses, the transformation into the mean frequency calculated on the subset of responses containing a test verb misleadingly suggests that young children matched more than older children. As an illustration, Figs. 1 and 2 contrasts the distribution of raw frequencies (Fig. 1, based on Table 2 in Matthews et al., 2005) to
that based on mean frequencies calculated on the subset of responses containing a test verb, that is, Match and Mismatch responses (Fig. 2, based on Fig. 1 in Matthews et al., 2005). Excluding missing data also impacts on the effects of age and lexical frequency. Table 3 shows the distribution of Match responses in Matthews et al. (2007) averaged over the two ungrammatical word orders (SOV and VSO). The first column reports raw frequencies as reported in Table 1 in the paper. The second column reports mean frequencies calculated on the subset of responses involving a test verb, i.e., excluding missing responses and missing children (data from Fig. 1 in Matthews et al., 2007). These mean frequencies were those considered for the statistical analyses. The third column reports mean frequencies calculated on the total number of responses expected from the subset of children included in the study, i.e., excluding missing children but including missing responses. The denominator used to obtain these mean frequencies is the number of children included in each condition (N in Table 1 of their paper) × 20 (the number of expected responses for each child). The fourth column reports mean frequencies calculated on the total number of responses expected from all children tested, including the children who failed to produce a test verb at any point in the experiment, i.e., including missing responses and missing children. The denominator is 28 (the number of children tested in each condition) × 20 (number of expected responses for each child) = 560.

What appears from Table 3 is that both the effect of age and the effect of lexical frequency on Match responses are dramatically different depending on whether or not missing data are included. Whereas the occurrence of Matches is similar for the two age groups in the low frequency condition when missing data are excluded (left column), it appears higher in the older group when all missing data are included (right column) as well as when only missing children are excluded (middle column). In contrast, the effect of lexical frequency is similar across the two age groups when all missing data are excluded (left column), but appears stronger in the older than in the younger group when missing children are excluded (middle column) and it disappears in the older group and is even inverted in the younger group when all missing data are included (right column).

In sum, studies using the WWO paradigm report a considerable amount of missing data that are critically Missing Not At Random. Missing data are distributed as a function of the experimental variables that were tested; hence, the probability of missingness appears to depend on the varying difficulty of the experimental conditions. This fact strongly suggests that they constitute an integrative part of children’s cognitive processing and that they can therefore not be discarded from the study.

5 The mean frequencies in the present figure are slightly different from those reported in Fig. 1 in Matthews et al. (2005). The reason is that, for the sake of clarity, we calculated these proportions on the basis of the raw frequencies reported in our Fig. 1. In contrast, Matthews et al. (2005) calculated proportions for each child individually and then averaged them. Note that this does not change anything with respect to data distribution.

6 The mean frequencies reported here correspond to those reported by the authors in their Fig. 1. They were calculated as the average of individual mean frequencies calculated proportions for each child (Matthews, personal communication). This explains why the numbers reported here (and in Fig. 1) do not exactly fit those obtained by dividing the total number of Match responses reported in Table 1 of the paper by the total number of responses (Match + Total reversion).

7 Missing responses consist in the failure to use a test verb, which means that this category of responses also involves cases where children replaced the verb presented with a familiar verb, which actually amounts to correcting the ungrammaticality of the model sentence.

5. From data to theory: critical issues with data interpretation

The influence of the three variables on children’s behaviour is now examined in turn: age, lexical frequency and grammaticality. The analysis shows that the effects argued to provide empirical evidence for the lexical hypothesis are inexistent or at best compatible with it or, more problematically, they are incompatible with it and even provide clear evidence for the grammatical hypothesis.

Table 3
Distribution of Match responses in the ungrammatical conditions (based on data reported in Matthews et al., 2007). The first column reports raw frequencies; the second column reports mean frequencies calculated on the subset of responses involving a test verb (data from Fig. 1 in Matthews et al., 2007); the third column reports mean frequencies calculated on the total number of responses expected from the subset of children included in the study; the fourth column reports mean frequencies calculated on the total number of responses expected from all children tested. The two ungrammatical orders manipulated in the study (SOV and VSO) were collapsed.

<table>
<thead>
<tr>
<th></th>
<th>Raw frequencies</th>
<th>Mean frequencies calculated on responses containing a test verb</th>
<th>Mean frequencies calculated on children included</th>
<th>Mean frequencies calculated on all children tested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low freq</td>
<td>High freq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.9</td>
<td>12</td>
<td>21</td>
<td>2.8</td>
<td>2.1</td>
</tr>
<tr>
<td>3.10</td>
<td>49</td>
<td>48</td>
<td>10.7</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>40.0</td>
<td>17.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.5</td>
<td>4.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.6</td>
<td>10.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.4</td>
<td>10.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.8</td>
<td>8.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6
5.1. Effect of age

The authors predicted that young children, who are assumed to lack grammar, would tend to re-use the ungrammatical word orders with novel verbs (Match) while, in contrast, older children would tend to correct them (Mismatch). Since no effect of age is expected on performance with familiar verbs (both age groups are expected to correct ungrammatical orders), we will concentrate on the condition involving novel verbs. We discuss in turn the effect of age on Match and on Mismatch responses.

5.1.1. Match responses

The data invalidate the authors’ prediction: young children do not produce more Match responses than older children, as Table 4 shows. Both groups of children produce extremely low rates of ungrammatical word orders. Raw numbers are reported for Matthews et al. (2005, 2007) because percentages depend on whether missing children are included or not. In Matthews et al. (2005), the young and the old group produced respectively 7.5% and 6.5% of Matches (4.7% and 4.1% if missing children are included) while in Matthews et al. (2007) these percents are 5.5% and 16.3%, respectively (2.1% and 8.8% if missing children are included). Hence, Matches represent between 0.9% and 7.5% of young children’s responses and between 0.3% and 16.3% of older children’s responses.

Abt (1999) did not report the statistics for the effect of age. Although the younger 2 year old group seem to match WWO (1.3%) more often than the older 4 year olds (0.3%), they seem to do so less often than the intermediate, 3 year old group (2.5%). Abbot-Smith et al. (2001), Matthews et al. (2005) and Matthews et al. (2007) all reported no effect of age on Match responses. An inverse effect of age, with more Match responses in the older group than in the younger group is even reported with familiar verbs in Matthews et al. (2005). This inverse effect seems also present in Matthews et al. (2007) but hidden by the transformation of the data into proportions. Indeed, the older group produced more Match responses than the younger group (97 vs. 33), an effect which would probably have shown up in the statistics if missing data (either missing responses only or both missing responses and missing children) had been taken into account (see Table 4).

Hence, Match responses do not appear differently distributed as a function of age, and under some analyses, Matches are even more frequent in the older group. Nevertheless, as summarized in Matthews et al. (2007), “in previous studies (Abbot-Smith et al., 2001; Abt, 1999; Matthews et al., 2005), the fact that older children were less likely to adopt a weird word order with novel items (...) was taken as evidence that older children have more abstract syntactic representations that hinder their adoption of a non-canonical order.” (Matthews et al., 2007:401).

Moreover, in most studies younger children did not match more than they mismatched incorrect sentences. Abt (1999) reported that “the two younger groups were equally likely to match the modelled structure as to make corrections” (Abt, 1999:347). Matthews et al. (2005, 2007) did not report statistics but Matches represent between 35% and 55% of children’s actual responses, that is, rather less than Mismatches (which are in complementary distribution, i.e., between 65% and 45%). Abbot-Smith et al. (2001) is the only one to report a significantly higher rate of Match (0.9%) than Mismatch (0.3%) responses in the younger group. Note however that the student t reported (t (15) = 1.9, p < .05) is calculated on a tiny subset of children’s responses (0.9%) and is critically not significant under a two-tailed test which should have been adopted given that the lexical hypothesis stands against another hypothesis which has been validated by other experimental work (that is, no a priori assumption should be adopted in the test).

The authors suggest that “this may reflect a general tendency to copy the experimenter’s word order at a rate just below 20% of the time. Any such general tendency to copy the experimenter’s word order would not explain the considerable effect of frequency on the 2 years old matches and reversions” (Matthews et al., 2005:133). It is hard to see why such an explanation would account for why older children match ungrammatical orders and not for why younger ones do so.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Match</th>
<th>Mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akhtar (1999)</td>
<td>Young: 1.3</td>
<td>Old: 1.5</td>
</tr>
<tr>
<td></td>
<td>Intermediate: 2.5</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>Old: 0.3</td>
<td>9.4</td>
</tr>
<tr>
<td>Abbot-Smith et al. (2001)</td>
<td>Young: 0.94</td>
<td>Old: 0.25</td>
</tr>
<tr>
<td></td>
<td>Old: 1.56</td>
<td>3.06</td>
</tr>
<tr>
<td>Matthews et al. (2005)</td>
<td>Young: 15</td>
<td>Old: 8</td>
</tr>
<tr>
<td></td>
<td>Old: 13</td>
<td>39</td>
</tr>
<tr>
<td>Matthews et al. (2007)</td>
<td>Young: 12</td>
<td>Old: 35</td>
</tr>
<tr>
<td></td>
<td>Old: 49</td>
<td>40</td>
</tr>
</tbody>
</table>
In sum, an incorrect picture of the distribution of Match responses arises when reading the papers under consideration here, according to which “the younger children were clearly willing to USE the ungrammatical structures” (Akhtar, 1999:347). This is not an appropriate description of the data showing that (1) young children’s production of ungrammatical orders with novel or low frequency verbs is a marginal phenomenon9 (between 0.9% and 7.5% depending on the study), (2) young children do not appear to match ungrammatical word orders more than older children,10 (3) young children do not appear to match ungrammatical orders more than they mismatch them.

5.1.2. Mismatch responses

As illustrated in Table 4, younger children nevertheless differ from older children in their tendency to mismatch the ungrammatical sentences presented by the experimenter. Apart from the study by Matthews et al. (2007) which reported similar rates of mismatch in both age groups, the other studies report less mismatches in the younger group than in the older group. Nevertheless, a closer look at the data shows that even though young children mismatch ungrammatical word orders less frequently than older children, they still do so, and sometimes even more often than they match (in studies by Akhtar, 1999; Matthews et al., 2007). Importantly, in studies involving transitive verbs (Akhtar, 1999; Matthews et al., 2005, 2007), there are 5 ways in which children could potentially mismatch the word order presented (6 possible orders, i.e., SVO, SOV, VSO, VOS, OSV, OVS, minus the one presented). Hence, the probability to mismatch by correcting the word order by chance, assuming that all orders have an equal probability, is 1/5. Yet, virtually all children’s productions were corrections. So if we consider the 35 mismatch responses reported in Matthews et al. (2007), the probability that they all consisted in corrections if children lacked grammar and responded at chance is 1/5^35, that is, virtually zero. Hence, what is crucial is actually not the small number of Mismatch responses obtained per se, but the consistency in these responses.

Akhtar (1999) proposed two interpretations for why young children produce corrections. Under the first account, “two year olds have more verb-general knowledge than they have previously been credited with” (Akhtar, 1999:352). But what does this knowledge consists of: is it abstract word order knowledge? Under the second, which the author favours, children would have formed schemas like ‘he-verb-him’ that allowed them to insert the novel verbs into the appropriate (sentence medial) position. However, in order to be able to use these schemas in their production of sentences involving full noun phrases, as in the experiments described here, young children must have extended the representation of these schemas to a more abstract representation like ‘NP agent-verb-NP patient’. This implies that they have built a grammatical representation of word order, something which is not acknowledged by any of the authors.

In sum, the distribution of Mismatch responses in young children shows that even though they mismatch ungrammatical orders less than older children, (1) they nevertheless occasionally do so and sometimes even more often than they match ungrammatical orders, and (2) Mismatches always amount to correct the ungrammatical order presented, a result which could not have arisen by chance. In sum, the Mismatches reported in young children, being all corrections, actually provide strong evidence that their performance is guided by knowledge of the correct word order.

5.2. Effect of lexical frequency

The authors predicted that the number of occasions on which the children would match the ungrammatical word order would be inversely proportional to the frequency of the verbs they heard. That is, the occurrence of Matches should increase as lexical frequency decreases, and conversely for Mismatches. Moreover, the effect of lexical frequency is expected to gradually diminish with age, that is, as grammar develops. It is important to note that the effect of age predicted by the authors is a logical corollary of the initial prediction. Indeed, if the same effect of lexical frequency is found in older children, who have grammar, then the effect observed in younger children cannot be taken as meaning that they lack grammar. We discuss in turn Match and Mismatch responses.

5.2.1. Match responses

In contrast to the prediction, Abbot-Smith et al. (2001) found no effect of lexical frequency in the younger group, but a significant effect in the older group, with more Matches with novel verbs. Matthews et al. (2005) reported an effect of frequency in the younger group in the expected direction but no effect for the older group, as predicted. However, when looking separately at each of the frequency conditions (low, medium, high), the authors found no effect of age. This finding questions the validity of the differential effect of lexical frequency observed for the two age groups. Matthews et al. (2007) reported a main effect of lexical frequency in both age groups, with no interaction with age. However, as Table 3 shows, young children actually produced more Matches in the high than in the low frequency condition (21 vs. 12), that is, in the opposite direction to the frequency effect reported in the statistical analysis conducted over proportions. The transformation of the data into Match proportions is entirely responsible for this reversion: the higher rate of Mismatches in the high

Note that the data do not allow determining whether the low rate of Matches is due to the fact that children do not want to produce ungrammatical orders or whether it is a consequence of the overall low response rate. Nevertheless, and independently of the explanation, it is wrong to claim that children are willing to use ungrammatical orders on the basis of the data.

The only paper that reports more Matches in young children than in older children is that of Akhtar (1999) who also reports more Matches in the intermediate age group than in the younger group. Hence, depending on how one looks at the data, one can either claim that young 2.8 year-olds produced more Matches than older children (than 4.4 year-olds) or less Matches than older children (than 3.6 year-olds). It seems therefore reasonable, given the reports from the other papers, to claim that there is no clear age effect.
frequency condition contributed to decrease the proportion of Matches in that condition. Unfortunately, Akhtar (1999) did not report the rate of Matches in her control experiment involving familiar verbs in ungrammatical sentences (Fig. 2 reports proportions with respect to Mismatch responses and only proportions are analysed in the statistics).

In sum, out of the 3 studies reporting statistics for the lexical frequency effect on Matches, one found an effect in the older group rather than in the younger group (Abbot-Smith et al., 2001), one found inconsistent statistical data (Matthews et al., 2005) and one failed to find the expected interaction with age (Matthews et al., 2007). Given this high variability in the results, it seems impossible to determine the role of frequency on Matches.

5.2. Mismatch responses

In the control experiment run by Akhtar (1999) as well as in the study by Matthews et al. (2007), lexical frequency was found to affect Mismatch responses similarly across all age groups. Abbot-Smith et al. (2001) found an effect in the younger group, but unfortunately failed to report the statistics for the older group. The data however suggest that older children were also sensitive to lexical frequency since more than twice more Mismatches were found in the condition with familiar verbs (6.5%) as compared to the condition with novel verbs (3.1%). Mismatches are not discussed in Matthews et al. (2005) since they are assumed to be the mirror image of Matches and therefore to show an effect of lexical frequency restricted to the younger group. But a closer look at the data of the older group suggests that, again, the transformation into Match proportions modified the results. Like in Matthews et al. (2007), children in this group produced more Matches in the high than in the low frequency condition (32 vs. 13). They also produced more Mismatch responses in the high than in the low frequency condition (138 vs. 39).

Since both Matches and Mismatches occurred more often in the high frequency condition, the transformation contributed to hide not only a probable effect of frequency on Mismatch responses in the older group, but possibly also an effect on Match responses in the opposite direction to that expected (i.e., more Matches in the high frequency condition).

In sum, in contrast to the expected predictions, two studies reported a frequency effect on Mismatches in all age groups (Akhtar, 1999; Matthews et al., 2007). One of these studies did not report statistics but the data go in the same direction (Abbot-Smith et al., 2001). The other study reported an effect restricted to the younger group however the lack of an effect in the older group seems entirely due to the transformation of the data into proportions (Matthews et al., 2005). It seems therefore reasonable to conclude that the production of Mismatches is sensitive to frequency in both groups of children. The finding that frequency also affects the performance of older children, who are assumed to have grammar demonstrates that this effect cannot be taken as an index of the absence of grammar.

5.3. Effect of grammaticality

Akhtar (1999) and Abbot-Smith et al. (2001) introduced grammaticality in their experimental designs. The hypothesis that young children do not have an abstract representation of word order predicts that they will behave similarly with grammatical and ungrammatical models involving novel verbs. In contrast, older children who are assumed to have developed abstract representations of word order are expected to match grammatical orders more than ungrammatical ones, and mismatch ungrammatical orders more than grammatical ones.

In both studies, young children, like older children, matched grammatical sentences four to five times more than ungrammatical sentences (respectively 5.8% vs. 1.3% on average in Akhtar, 1999 and 4.5% vs. 0.9% in Abbot-Smith et al., 2001). Moreover, young children virtually never mismatched grammatical orders (0% in Akhtar, 1999 and 0.1% in Abbot-Smith et al., 2001), which contrasts with their tendency, although smaller than older children, to mismatch by correcting ungrammatical orders (see discussion of Mismatch responses in section 5.1).

According to Abbot-Smith et al. (2001), “this indicates the influence of at least implicit verb-general knowledge on their performance in this task” (Abbot-Smith et al., 2001:698). The authors nevertheless conclude that “it is unlikely that all 2.4-year-olds have general word order ‘competence’ but are prevented from actively demonstrating this by some kind of ‘performance’ problem. If this were the case, one would need to account for the willingness of roughly 50% of these children to productively use ungrammatical word orders” (Abbot-Smith et al., 2001:690). But what the authors argue would need to be accounted for, i.e., the willingness to match ungrammatical word orders and their productive use, is not what they actually reported. First, it is hard to claim that young children are willing to use ungrammatical word orders given that their production of these orders represents less than 1% of their responses in that study (0.88% and 0.94% in the familiar and novel verb conditions, respectively).

Second, besides the fact that children seldom produce ungrammatical word orders, when they do so, they show absolutely no markers of productivity, which critically contrasts with what is found in their grammatical sentences. Two markers were examined by the authors: pronominalizations and verbal morphology. Abbot-Smith et al. (2001) reported that “when children used the non canonical order, they virtually never used a pronoun (a total of 3 out of 76 utterances). (. . .) Conversely, when they corrected the familiar VS verb and/or the novel VS verb to canonical order, they used pronouns much more often than nouns.” (Abbot-Smith et al., 2001:688). That is, whereas pronouns are virtually never produced in ungrammatical sentences (3% and 0% with novel verbs and familiar verbs, respectively), they replaced noun phrases in 52% of their grammatical sentences involving familiar verbs and 20% of those involving novel verbs (numbers provided in Table 2, Abbot-Smith et al., 2001:688).

A similar asymmetry between grammatical and ungrammatical sentences is reported in the other studies. Young children in Akhtar (1999) produced 0 pronominal arguments in their Matches of ungrammatical orders (62 lexical arguments),

---

11 The author ran statistics across all age groups which suggests that lexical frequency similarly affected all of them.
against 33 pronominal arguments in their Mismatches (37 lexical arguments). Matthews et al. (2005) found “a stark difference in pronoun use between SOV Matches and SVO reversions: in both age groups not a single child used a pronominal argument when matching SOV word order” (Matthews et al., 2005:131). Hence, although young children are willing to go beyond the input in their grammatical sentences, using pronouns that were not used by the experimenter, this productivity is absent in their production of ungrammatical sentences.

A similar lack of productivity in ungrammatical productions was observed with respect to verbal morphology. Abbot-Smith et al. (2001) reported that when using VS order, “none of the children ever added an auxiliary, no matter the form of the verb (progressive, past or bare). On the other hand, when using SV word order with the progressive, the 2-year-olds added the auxiliary 70% of the time with pronoun subjects and 26% of the time with nouns.” (Abbot-Smith et al., 2001:689).

In other words, whereas young children show productivity in the verbal morphology of their grammatical sentences, ungrammatical sentences appear to be generated by way of some rigid, non grammar-based process, calling on the application of some general-purpose, problem solving procedure.

In sum, these two studies show that young children, like old children, do not behave similarly with grammatical and ungrammatical orders: (1) they match grammatical orders more than ungrammatical orders, (2) they never mismatch grammatical orders, in contrast to ungrammatical orders, and (3) markers of productivity are present in their grammatical sentences (pronominalizations, verbal morphology) but absent in their sentences matching ungrammatical orders.

6. General discussion

The different results reported in the four papers under consideration are now summarized and discussed with respect to the question of the representation of word order in young children. We then briefly address more general issues with the rationale of the usage-based framework for language acquisition.

6.1. The representation of word order in young children

Eight major empirical results consistently emerged in all studies, independently of how missing data was being dealt with. They can be summarized as follows:

(1) Both groups of children seldom re-use ungrammatical word orders;
(2) Young children re-use ungrammatical orders at a rate similar to older children;
(3) Both groups of children correct ungrammatical orders more often if they know the verb;
(4) Young children correct ungrammatical orders less than older children do;
(5) Both groups of children tend to re-use the order presented more if the model is grammatical than if it is ungrammatical;
(6) Both groups of children mismatch ungrammatical orders only, not grammatical orders;
(7) Both groups of children correct ungrammatical sentences when they mismatch them.
(8) Both groups of children show the hallmarks of productivity in their grammatical sentences (pronominalisations, verbal morphology) but not in their ungrammatical sentences.

The papers reviewed capitalized on three major claims put forth as major empirical evidence for a lexically encoded word order in young children. The first claim is that young children correct ungrammatical orders less than older children. The second claim is that young children tend to produce sentences matching the ungrammatical word order modelled. The third claim is that young children correct ungrammatical sentences more when they know the verb than when they do not know it. We discuss them in turn.

The first claim is in line with the data (observation 4). However, as we pointed out in section 3, the logical implication between grammar and corrections of ungrammatical orders is one in which corrections are only a sufficient condition for the presence of grammar (Correction → Grammar). The reason is that in order to correct, the child not only needs a grammar but also other cognitive abilities: the ability to infer the argument structure of the verb from a WWO sentence and the ability to infer the experimenter’s expectations. Abbot-Smith and colleagues proposed yet another factor involved in the ability to correct which has to do with processing limitations (Abbot-Smith et al., 2008). They found that when presenting 2-year-old English-speaking children with ungrammatical OVS sentences with novel verbs (rather than SOV or OSV as in previous studies), children corrected them more often than in previous studies. The authors interpret this higher rate of correction as resulting from syntactic priming: model sentences being of the NVN form would have played a role in ‘freeing up processing capacity’ allowing children to focus on linking semantic roles to positions of the noun phrases in the sentence. Critically, a different conclusion from that reached in their previous work is put forth in this more recent paper. Indeed, the authors now propose that children, by age 2, already have abstract grammatical representations, although these are assumed to be ‘weak’ and gradually become stronger (see section 6.2 for a discussion of ‘graded’ representations).

In contrast to the first claim, the second and the third claims are not in line with the data reported, as observations (1)–(3) above show. First, young children seldom produce ungrammatical sentences which represent between 0.9% and 10% of their responses, depending on the experiment. Moreover, finding (8) strongly suggests that these ungrammatical sentences, when they are produced, are not at all productive. Second, older children, who are assumed to have grammar, re-use ungrammatical orders at a similar rate to young children; hence, this cannot be taken as an index of a lack of grammar in the
younger group. Finally, older children’s performance is also affected by lexical frequency. Hence, the effect of lexical frequency cannot be taken as an index of the absence of grammar either.

In light of these critical observations, the conclusion that young children have a lexicalized representation of word order has to be revisited. The first possibility is to conclude that the effects that were assumed to be an index of an absence of grammar, i.e., the reproduction effect and the frequency effect, are actually not evidence for the absence of grammar. This amounts to invalidating the predictions of the lexical hypothesis. While this may not be sufficient to disconfirm the lexical hypothesis entirely, a search for new valid predictions and empirical support is needed. In light of observations (5)–(7), another possibility is to abandon the lexical hypothesis altogether, in favour of the grammatical hypothesis. Indeed, in contrast to the interpretive indeterminacy of observations (1)–(4), observations (5)–(7) provide strong evidence that young children actually do have a grammar. If the young child did not have an abstract representation of word order, she would not re-use grammatical word orders more than ungrammatical orders (5), she would not mismatch the order presented more when it is ungrammatical than when it is grammatical (6), and she would not revert ungrammatical orders to the correct order more than to another incorrect one (7). In other words, it seems hard to account for this set of results without assuming that the young child already has an abstract representation of the word order of her language.

Further evidence comes from a study which we conducted on French-speaking children, replicating Matthews et al. (2007). Two changes were introduced to the initial study. First, rather than involving exclusively ungrammatical sentences, the materials contained both ungrammatical and grammatical models. Second, model sentences were preregistered and presented via loudspeakers rather than by the experimenter. These two changes appear to have played an important role in children’s performance since the younger group (aged 2.11, as in Matthews et al., 2007) was found to correct ungrammatical sentences at a similar rate to the older group (Franck et al., 2011). We believe that the effect of these changes operates at the level of the inferences children made as to the experimenter’s intentions. Introducing grammatical models helped children realize that ungrammatical models were weird (which is only possible if the child has a grammar) and the computer presentation of ungrammatical sentences reduced their tendency to imitate the experimenter. Finally, we ran the same experiment on four 8–9 years old children. All of them were puzzled with the task and asked questions as to what they were expected to do. Half of them systematically re-used (with effort!) the ungrammatical structure presented by the experimenter. When asked whether they believed their sentences to be correct, children said they did not, but did what the experimenter asked for. The moral is that children, like adults, can imitate weird behaviours, but the functional meaning of such imitations has to be taken with great caution, since in the absence of additional elements, it is impossible to determine whether it reflects some internal abnormality or simply the very basic ability to imitate.

In sum, we have shown that the results presented in the four papers reviewed provide two sorts of empirical evidence with respect to the debate opposing the lexical hypothesis to the grammatical hypothesis. The first set, observations (1)–(4), contains evidence that is compatible with the lexical hypothesis, although observations (1)–(3) would require that the initial predictions attributed to the lexical hypothesis be revised. Hence, one may at best conclude that the lexical hypothesis is a sufficient condition for the data at stake; however, the data are also compatible with the alternative, grammatical hypothesis. The second set of results, observations (5)–(8), contains evidence that is not only incompatible with the lexical hypothesis, but suggests that the grammatical hypothesis is a necessary condition without which it is hard to account for children’s performance. One therefore is brought to conclude that the grammatical hypothesis has more explanatory power than the lexical hypothesis.

6.2. Critical issues with the usage-based framework

The major claim of the usage-based framework is that a system that tends to re-use a particular structure is not productive, whether the structure is correct (as in diary data, e.g., Tomasello, 1992), or incorrect (as in the WWO task). This view was developed in the foundational paper of Tomasello (2000) presenting the usage-based theory in which the critical argument against the continuity hypothesis (i.e., the grammatical hypothesis) is that “young children’s creativity – productivity – with language has been grossly overestimated; beginning language learners produce novel utterances in only some fairly limited ways.” (Tomasello, 2000:210).

However, as argued by Yang (2011), the theory critically fails to provide a statistical test for assessing productivity. Computational modelling work by Yang (2011) demonstrated that the low overlap in determiner-noun combinations found in children’s productions (nouns are typically produced either with ‘a’ or with ‘the’, Pine and Lieven, 1997), considered as evidence for children’s lack of productivity, is actually a direct consequence of the statistical properties of linguistic units (Zipf’s power law) and their combinations in the language. Moreover, Yang showed that children’s productions are inconsistent with the hypothesis of memorized determiner-noun pairs of the usage-based account. Rather, a model relying on the productive rule “DP → DN” appears to provide a quasi perfect fit to corpus data from children’s very early productions. Hence, Yang’s modelling work points to the very early availability of an abstract and productive grammar.

In addition to the lack of a statistical approach to productivity, the usage-based framework relies on at least two non obvious underlying assumptions. We briefly review them in turn, and then address a third issue with respect to the recent hypothesis of ‘graded representations’ put forth as a key property of linguistic representations within recent constructivist frameworks.

The first assumption is that children’s productions provide a transparent window over their linguistic representations. Research on the acquisition of word order is confronted with a major contradiction between studies showing that children below age two already represent word order and use it in sentence comprehension (e.g., Hirsh-Pasek and Golinkoff, 1996;
Gertner et al., 2006), and studies suggesting that word order is encoded as a lexical property of the verbs (e.g., the papers reviewed here). At the methodological level, when facing two experiments, one showing a significant effect of a variable, the other one showing no effect, it is generally concluded that the latter involves a task that is not sensitive enough for the effect to show up. As an illustration in the field of language acquisition, children were found to process pronouns in virtue of binding constraints from age 2.6 in truth value judgement tasks (McKee, 1992) but not in yes/no tasks in which performance was only successful at the age of 5 (Chien and Wesler, 1990). No one would want to argue that children’s grammar varies with the task; rather, it is concluded that truth value judgement tasks are more adequate to tap into the child’s grammar than yes/no tasks, which appear to put an additional cognitive burden on the child’s response (see Crain and Thornton, 1998, for a critical assessment of experimental methods in the study of language development).

In the present case, what distinguishes experiments showing that young children have a grammatical representation of word order from those showing that they lack it is that while the former involve measures of children’s abilities in comprehension, the latter involve measures of their abilities in production. Research on the acquisition of syntax has reported many cases of such dissociation between comprehension and production, with a systematic advantage for the former (see Naigles, 2002, for an extended argumentation of this point). For example, studies of children’s processing of function words by Gerken et al. (1990) have shown that young children, who typically omit function words in their productions, have problems processing sentences that lack those words in comprehension tasks. Work on object clitic omission in French speaking children with Specific Language Impairment (SLI) also shows that even though they omit the object clitic, these children interpret their presence or absence correctly (Grüter, 2005). Indeed, the same verb is interpreted as intransitive in the absence of the object clitic, but as transitive in its presence. Work on agreement by Roulet and Jakubowicz (2006) showed that French SLI, who typically fail to produce gender agreement (they produce masculine forms as default), are nevertheless slowed down when processing noun phrases containing a gender agreement error.

Why do production studies systematically show poorer knowledge of grammar than comprehension studies? Whereas the experimental study of sentence comprehension only requires that the child looks or points at a picture, production experiments involve many processes, amongst which the comprehension of the linguistic materials presented (rendered even more difficult when the verb is novel or when the sentence is ungrammatical), the processes involved in grammatical encoding, phonological encoding, articulatory planning (e.g., Bock and Levelt, 1994), and even sometimes the complex inference about the experimenter’s intentions. Hence, it is most plausible that the poor performance exhibited by young children in production tasks, particularly so in weird word order experiments as attested by the extremely high rate of missing data, leaves only little space for grammatical knowledge to show up.

The second assumption of usage-based accounts is that the tendency to re-use grammatical structures attests to a lack of productivity. Such a position ignores the literature in adults’ psycholinguistics showing that priming is pervasive at all levels of language production: semantic, syntactic, lexical, phonological and even motor. With respect to our concerns with syntax here, numerous studies ever since Bock (1986) have shown that hearing, reading, speaking and writing a sentence with a given structure increases the probability of subsequently producing the same structure, regardless of whether the prime and target share lexical content (see Branigan, 2006, for a synthesis). Syntactic priming was also reported in corpus studies of spontaneous productions (e.g., Gries, 2005) and was found to be persisting for long intervals (Bock and Griffin, 2000). Word order, in particular, was also found to be affected by priming in adult sentence production (e.g., Bock and Loebell, 1990; Hartsuiker et al., 1999). In addition, young children were found to be sensitive to syntactic priming for a variety of structures (like passives, double objects, prepositional objects, see for example Huttenlocher et al., 2004; Tothathiri and Snedeker, 2008).

The finding that adults tend to repeat syntactic structure across otherwise unrelated utterances can obviously not be taken as evidence that they lack grammar, as is suggested by the usage-based interpretation of this effect in children. Rather, syntactic priming has been argued to provide evidence (1) for the role of abstract structural representations that in many ways resemble those proposed by theoretical linguists (Branigan, 2006), and (2) for syntactic representation being subject to activation levels (e.g., Hartsuiker et al., 2004), possibly by way of a similar mechanism of spreading activation proposed in the literature on semantic, lexical and phonological processing (e.g., McClelland and Rumelhart, 1981). Syntactic priming in young children may be of a different sort, but one would need empirical and theoretical arguments for interpreting it differently from that found in adults.

Finally, a third assumption which gained popularity within recent usage-based frameworks concerns the gradedness of linguistic representations in children (e.g., Abbot-Smith and Tomasello, 2006; Abbot-Smith et al., 2008; Arnon and Clark, 2011). This account proposes that lexical, morphological and syntactic representations are graded in strength: representations become stronger as the number of exemplars which the child is exposed to increases. Weak representations are accessible, but noisy, which explains why they do not always show up in performance. Abbot-Smith et al. (2008) proposed to interpret the results from their WWO study within this framework. The authors reported cross-linguistic differences between English and German 2 year old children in a WWO experiment involving ungrammatical Object–Verb–Subject sentences. Results show that German children correct ungrammatical orders more often than English children when confronted with novel verbs. The authors argue that this difference is due to the presence of stronger cues for thematic roles in German (i.e., case marking), leading to “more robust verb-general representations of agent and patient marking”. According to the authors, transitive representations of children are graded in strength depending on the amount of input they received and the salience of perceptual cues.

However, there is an intrinsic confusion underlying the ‘graded representations account’. At the theoretical level, this view clearly assumes that what is graded are linguistic representations. Gradedness may take a different meaning whether
one is concerned with grammatical representations (e.g., Abbot-Smith et al., 2008) or lexical representations (e.g., Arnon and Clark, 2011). But with respect to grammatical representations, the key assumption of constructivist models with respect to early grammatical knowledge lies in the lack of abstractness (in initial proposals) or, as more recently formulated, in the weakness of abstract representations. But how can a property such as word order (or more specifically here the representation of the fact that subjects precede verbs while objects follow them) be ‘more or less abstract’? One could consider the possibility that knowledge is partial, as for example the fact that the object follows the verb, while the child does not yet know that the subject precedes it. Still, although this representation is partial with respect to the overall grammar, it is abstract in that it is independent of specific lexical units. In other words, it is hard to conceive abstractness as a gradable property of representations. Hence, grammatical representations do not seem to be the locus of the graded performance effects reported.

In contrast, the processes by which these representations are accessed/retrieved from memory appear as a natural locus for gradedness effects. Indeed, psycholinguistics abounds in observations attesting of graded levels of activation of linguistic units of all types (phonemes, words, phrases, sentence structures). In fact, careful reading of the authors of the graded representations account suggests that they may actually mean ‘graded retrieval process’. This transpires in many instances of their writing which call on the role of various processing factors. For example, “with constructions in which the agent is marked with 2 or more cues, the young 2 year olds may find it easier to overcome the processing difficulties involved in producing a transitive with a novel verb, than with constructions in which the agent is marked with only 1 cue” (Abbot-Smith et al., 2008). That is, the number of cues plays a role in reducing processing load, not in the representation of the transitive construction itself. “If a child’s representation is still ‘weak’, it is accessible but because its signal is noisy, it may only show up with dependent measures which do not tax executive functions mechanisms and memory systems” (Abbot-Smith et al., 2008). Again, this suggests that what is graded is the ease with which an existing representation is retrieved from memory, a process which critically relies on non linguistic factors like memory and executive functions.

If this interpretation of the new formulations of the constructivist framework is correct, then this position is entirely aligned with the grammatical hypothesis, and the debate may be closed. If it is not, then understanding what opposes the constructivist view to the grammar-based account will require that the notion of graded grammatical representations be conceptually clarified.

7. Conclusion

A fine analysis of a series of four papers using the weird word order paradigm led us to conclude that this methodology, and in particular the way it is used in these studies to tap into grammatical representations, does not support the theoretical conclusions adopted in these papers. Three lines of arguments were developed. First, we argued that the task in weird word order experiments mobilizes inferential components independent of grammar that contribute substantially to performance. Second, we showed that most of the observations reported in these studies consist in missing data that are distributed as a function of the variables manipulated and can therefore not be discarded as has been done in the papers reviewed. Third, a closer look at the data themselves showed that even though some are compatible with the hypothesis that word order is lexicalized, they are also compatible with the hypothesis that word order is represented abstractly. More critically, assuming that the child has an abstract representation of word order appears as a necessary condition to account for other aspects of the data that are truly incompatible with the lexical hypothesis.

The reanalysis of the empirical evidence reported in these papers therefore wipes out the apparent contradiction between the initial conclusions of these papers and the literature showing that children, by age two, already have encoded word order in an abstract format. This is not to say that the data invalidate the claim that language acquisition is data-driven or that it is a gradual process; the data discussed do not address the issue of the acquisition process. However, what the data strongly suggest is that as early as researchers have been able to test it, young children appear to dispose of abstract knowledge of this grammatical property of their native language.

Acknowledgements

The work reported here was supported by grant number 100014_126924 of the Fonds National de la Recherche Suisse to Julie Franck. Many thanks to Jeff Bowers, Anne Christophe, Stephen Crain, Uli Frauenfelder, Sven Matthys, Alissa Melinger, Severine Millotte, Julien Musolino, Akira Omaki, Luigi Rizzi, and Rushen Shi for the exciting discussions that the writing of this paper generated. We take full responsibility for the paper.

References


Chien, Y.-C., Wexler, K., 1990. Children’s knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. Language Acquisition 1, 225–295.


