Linguistic variation and micro-cues in first language acquisition

Abstract: Children are often exposed to considerable variation in the input. Nevertheless, there is very little overgeneralization in child language data and children are typically found to make errors of omission, not errors of commission, a fact which is often referred to as conservative learning. In this paper, these findings are accounted for by a model of micro-cues, a generative approach to language acquisition arguing that children are sensitive to fine syntactic distinctions from early on. The micro-cues are small pieces of abstract syntactic structure resulting from parsing the input. This means that UG provides children with principles, features, and the ability to parse, but not the micro-cues themselves, which are considered to be part of the knowledge of a specific language. The model also considers children's errors to generally be due to economy and the language acquisition process to be development in small steps, from specific to more general knowledge.

1. Introduction

While traditional generative theory has mainly been concerned with variation across languages, variation within languages is not uncommon. Thus, children are typically exposed to considerable variation in the input. Based on findings from the acquisition of such variation, this short paper discusses and further develops the model of micro-cues in language acquisition (Westergaard 2009a, 2013). The main focus of the paper is on

principles of economy, aspects of conservative learning, and arguments that the acquisition process is a stepwise development from specific to general knowledge.

The paper is organized in the following way: In the next section, I give a brief overview of some child data on word order patterns that involve variation in the input, showing that children generally make fine distinctions in syntax and information structure from early on (Westergaard 2009a, c, 2011, Anderssen & Westergaard 2010, 2012, Westergaard & Anderssen forthcoming). In section 3, I discuss the occasional errors that are typically found in child language data and argue that most of these are due to a principle of economy in the acquisition process, and I relate this to the idea of conservative learning (Snyder 2007). In the section 4, I discuss the idea of grammar competition (Yang 2002, 2010), arguing that it generally predicts massive overgeneralization in early child data and that competition must therefore be restricted to low-level variation, i.e. affecting very small parts of the grammar. I then outline the model of micro-cues and the idea of "learning by parsing" (Fodor 1998, Lightfoot 1999, 2006, Westergaard 2009a). Finally, in section 6 I briefly compare traditional generative accounts to constructivist accounts of language acquisition and argue that learning is from specific to general (Ambridge & Lieven 2011, Westergaard 2013) and not the other way around (Biberauer & Roberts 2012). Section 7 is a brief conclusion.

2. Variation in the Input

In traditional generative theory, variation across languages is accounted for by the existence of parameters provided by Universal Grammar (UG); see e.g. Chomsky (1981, 1986), Snyder & Lillo-Martin (2011). These are typically considered to be mental switches for aspects of grammar where languages differ, e.g. whether heads precede or follow their complements (head parameter), whether subjects may be null or not (pro-drop parameter), or whether verbs have to appear in second position (V2 parameter). The switches will be turned to the correct value as a result of exposure to a particular language early in the language acquisition process (Wexler 1999). On this view, children's errors are often considered to be due to parameter mis-setting; see e.g. Hyams (1986) for one of the most cited examples of this, where she argues that subject omissions in the early production of children learning a non-prodrop language is due to a mis-

setting of the prodrop parameter. The idea of parameter mis-setting has turned out to have a number of problems; see e.g. Valian (1990, 1991) for a thorough discussion of Hyams' account.

There is also considerable variation within languages. For several years, the acquisition research group at the University of Tromsø has been concerned with how children deal with variation in the input, especially in the domain of word order. The most important research questions that have been addressed are whether children have an early preference for one of the two options, possibly indicating parameter setting, and how early they master the often fine distinctions between the two word orders in the adult language, with respect to both syntax and information structure. Here I will briefly overview some work on variable verb second (V2), different subject positions (subject shift), and word order variation inside the DP (possessives), but the group has also produced relevant work on object shift (Anderssen, Bentzen & Rodina 2012, Anderssen & Bentzen 2012, Bentzen, Anderssen & Waldmann 2013), object scrambling (Mykhaylyk 2011, 2012), embedded clause word order (Westergaard & Bentzen 2007, Westergaard 2009a) and double object constructions (Mykhaylyk, Rodina & Anderssen 2013, Anderssen, Rodina, Mykhaylyk & Fikkert 2014).

It is well known that V2 word order is not obligatory in Norwegian (e.g. Vangsnes 2005, Westergaard 2009b). The variation is dependent on factors such as clause type (e.g. declarative vs. wh-question), the initial constituent (e.g. phrasal vs. monosyllabic wh-elements), and information structure (whether the subject conveys given or new and/or focused information). Investigating the spontaneous utterances of three children in an acquisition corpus (Anderssen 2006), Westergaard (2009a) finds that all three children produce target-consistent V2 as well as non-V2 word order in appropriate contexts from early on, without any overgeneralization. For example, if the initial element in a declarative is the adverb *kanskje* 'maybe', the target language allows both V2 and non-V2 word orders, but speakers prefer non-V2 as often as approximately 95% (cf. Westergaard 2008b), while all other non-subject-initial elements (adverbs, objects, etc.) require V2. Examples (1) and (2) illustrate that the children produce non-V2 with initial *kanskje* 'maybe', while V2 is produced with other non-subject-initial material (here the

adverb *no* 'now'), and according to Westergaard (2008b, 2009a), the proportions of each word order in the child data are more or less the same as in adult child-directed speech.

(1)   *kanskje **det var*** *en anna dag.* (Ina.9, age 2;2.12)
        maybe it be.PAST an other day
        'Maybe it was another day.'

(2)   *no er   det* *borte.* (Ina.06, age 2;1.0)
        now be.PRES it   gone
        'Now it is gone.'

In wh-questions, there is even more word order variation: Although the standard language requires V2, non-V2 is widespread in most parts of the country (cf. e.g. Westergaard, Vangsnes & Lohndal 2012). There is considerable micro-variation across dialects, but a common distinction is that V2 is required if the wh-element is long (more than one syllable), as in (3), while both word orders are grammatical if the wh-element is monosyllabic; see (4a-b). Westergaard (2009b) uses the Head Principle of van Gelderen (2004) to argue that this distinction is due to monosyllabic wh-words being heads and longer wh-elements being phrases. This means that the monosyllabic wh-words may move into the head position that the verb normally moves to (the head of the Interrogative Phrase), thus blocking (generalized) V2. In contexts where both word orders are allowed, the choice between the two is dependent on information structure: More specifically, non-V2 is used if the subject has been mentioned in previous discourse and thus conveys given information (typically a pronoun, as in 4b) and V2 if the subject expresses new information (often a full DP, as in 4a). V2 may also be used with given subjects if they are focused or contrasted. Westergaard (2009b) thus argues that a lower functional head is involved (the head of the Topic Phrase), attracting the verb only when the subject conveys new and/or focused information. Although there is quite a bit of variation across different speakers with respect to the frequency of V2 in this context (between 2.5% and 68.4% in the data investigated in Westergaard 2009b), children are typically exposed to considerable proportions of both word orders. The three children in

the acquisition corpus seem to have no problem producing V2 and non-V2 in appropriate contexts, as shown by the following examples.

(3)   *koffer **har**   **han** fått den?* (Ina.22, age 2;10.2)

        why   have.PRES he   got that

        'Why did he get that?'

(4)   a.   *kor   er   **Ann sin   dukke** hen?* (Ann.04, age 1;11.0)

                where be.PRES Ann POSS doll   LOC

                'Where is Ann's doll?'

        b.   *ka   **du   gjør**?* (Ann.10, age 2;3.9)

                what you do.PRES

                'What are you doing?'

   Norwegian also displays word order variation with respect to the position of the subject in all non-subject-initial sentences with V2, i.e. wh-questions and non-subject-initial declaratives. If negation or another adverb is present, the subject may either precede or follow this element, i.e. the word order is either XP-V-Neg-S or XP-V-S-Neg. The choice of word order is again dependent on information structure, informationally new and/or focused subjects (typically DPs) following negation and given subjects (typically pronouns) preceding negation. This is sometimes referred to as subject shift (e.g. Westergaard 2008a, 2011), involving movement of an informationally (and often prosodically) light subject to the higher position. In spontaneous discourse, the high subject position is considerably more frequent than the lower position, since subjects tend to be given information. In the Tromsø acquisition corpus (Anderssen 2006), the high subject position is attested 81% (1351/1667) in relevant utterances in the adult data. The three children in the corpus are sensitive to this distinction from early on (Westergaard 2008a, Anderssen & Westergaard 2010, Westergaard 2011), typically producing DP subjects in the low position (following negation) and pronominal subjects high (preceding negation), as illustrated in (5) and (6).

(5)   *korfor kommer   **ikke mummien** sæ   laus?* (Ole.17, age 2;8.24)

why   come.PRES not mummi.DEF REFL loose

'Why is the Mummi troll stuck?'

(6)   *og   no   kan **æ ikke** drikke det.* (Ole.19, age 2;10.0)

and now can I not   drink it

'And now I can't drink it.'

Word order variation is also found inside the DP, in that the possessor may either precede or follow the head noun, depending on whether the possessor is topical or focal, yielding N-POSS and POSS-N word orders respectively; cf. Lødrup (2011), Anderssen & Westergaard (2010). Investigations of corpora of spontaneous speech reveal that the postnominal possessor construction (N-POSS) is far more frequent than the prenominal one in children's input, being attested approximately 75% (Anderssen & Westergaard 2012, Westergaard & Anderssen forthcoming). Nevertheless, the corpus data investigated reveal that the children produce both word orders from early on, as shown in (7a-b). The interpretation of (7a) would be that the possessor is focused and contrastive (i.e. *my* dress, not somebody else's), while the possessive relationship in (7b) is neutral/non-contrastive.

(7)   a.   *det er **min kjole**.* (Ina.07, age 2;1.23)

               it   is my   dress

               'It is my dress.'

        b.   *nei no   dætt ned   mannen   på **foten**   **min**.* (Ina.08, age 2;1.29)

               no now fall down mann.DEF on foot. DEF my

               'Oh no - now the man is falling down on my foot.'

To conclude this section, the data from children's spontaneous production of word order variation, both at the clausal and the phrasal level, show that they produce both options from early on. Furthermore, they generally produce the two word orders in appropriate contexts. In these domains, therefore, there does not seem to be any evidence that children are setting (or mis-setting) parameters, which would have resulted in massive and indiscriminate overgeneralization of one of the word orders in early production.

3. Economy and conservative learning

Considering the data reviewed in the previous section, it is an important question whether children *ever* make mistakes in spontaneous production. In fact, they do. But their errors are of a particular type. In the three domains discussed in the previous section, verb placement, subject placement and the position of the possessor in relation to the noun, young children have been found to occasionally produce the element in question in a lower position than what the target language requires. Example (8) shows that the verb has failed to move across the subject to verb-second position in a declarative, (9) shows a pronominal subject in a position following negation, and (10) shows that the child produces POSS-N word order (without N-movement across the possessor) in a context where the adult investigator produces N-POSS.

(8)   *nå æ skal (s)t(r)ikke litt   til.* (Ole.10, age 2;4.6)

         now I shall knit   little more

         'Now I will knit a little more.' Target: Nå skal æ strikke litt til.

(9)   *det får   ikke æ lov   til.* (Ole.12, age 2;5.18)

         that get.PRES not   I   allowed to

         'That I am not allowed to do.' Target: Det får æ ikke lov til.

(10)   a.   Ina: *i min munn*. (Ina.20; age 2;8.27)

                          in my mouth

                          'Into my mouth.' Target: I munnen min.

         b.   Inv: *ja og opp i munnen   din.*

                          yes and up   in mouth.DEF your

                          'And into your mouth, yes.'

The examples in (8)-(10) all illustrate lack of syntactic movement. Similar findings have been attested in children's production of object positions (e.g. Anderssen, Bentzen, Rodina & Westergaard 2010). Thus, I have claimed (e.g. in Westergaard 2009a) that this production is not due to a defect in the children's I-language grammar, such as a mis-set parameter. Instead this is argued to be due to a third factor, commonly seen in the process

of language acquisition: Economy. That is, children are economical in their production and will not produce an element, perform a movement operation or build syntactic structure, unless there is clear evidence for it in the input. This means that there is little or no overgeneralization in child language data; in fact, we often find the opposite. Roeper (1999:175) also claims that there is widespread evidence for "undergeneralization" in child language.

Similarly, Snyder (2007) provides an overview of a number of language acquisition studies, focusing on very different areas of grammar than we have done here, e.g. verb-particle constructions in English or preposition stranding vs. pied-piping in English and Spanish. Snyder (2007) convincingly shows that children's errors are generally restricted to errors of omission, while the number of errors of commission is negligible in child language data. He refers to this as 'grammatical conservatism' and argues that traditional approaches to learnability, such as the Trigger Learning Algorithm (Gibson & Wexler 1994), are not sufficient to explain the acquisition process of a conservative learner, since such approaches necessarily predict massive errors of commission as the child moves from one grammar to another (i.e. sets and re-sets parameters). Instead, Snyder suggests that the ideas proposed in Fodor (1998) may be compatible with conservative learning: Fodor argues that children's initial grammars are endowed with small pieces of syntactic structure, so-called 'treelets', and that children use these to identify possible parses for the input that they are exposed to. If the parse is unambiguous, the grammar will use this to set a parameter. This is referred to as "learning by parsing" and will be returned to in section 5 below, as it is in principle very similar to the idea behind the micro-cue model.

Finally in this section, I would like to point out that children's economic lack-of-movement errors are not always random. Occasionally it is possible to find that children make certain distinctions in their non-target-consistent production that are not reflected in the input. For example, when V2 fails in Norwegian or Swedish child language, this typically happens when the subject is a pronoun and/or the verb is another verb than *be*, as in (8) above (Westergaard 2004, Waldmann 2008, 2013). This means that V2 word order is initially preferred with *be* and DP subjects, just like the target-consistent V2 in wh-questions. This preference is relatively short-lived but found to be statistically significant (Westergaard 2009a). It has also been argued that some English-speaking

children's lack of subject-auxiliary inversion is systematically related to certain wh-items, typically distinguishing between *what* and *where* on the one hand, which trigger inversion early and almost consistently, and *why* on the other, which triggers inversion only at a much later stage (e.g. de Villiers 1991, Thornton 2008, Westergaard 2009c). A significant distinction has also been found between *be* and auxiliaries in English children's wh-questions (e.g. Westergaard 2009c, forthcoming). This means that children are *systematically* "undergeneralizing", i.e. producing less movement than what is required in the target language.

4. Grammar competition and the "size" of rules

Roeper's (1999) seminal article on "universal bilingualism" introduced the idea that monolingual children who are exposed to variation in the input may entertain two different grammars for an extended period of time. For example, English children are exposed to a grammar where the verb *be* inverts with the subject, but other lexical verbs do not, as illustrated in (11)-(12).

(11)   Where is she?

(12)   *What drinks she? / What does she drink?

By comparison, German children are exposed to a grammar where all lexical verbs invert, and they are thus assumed to set the V2 parameter to its positive value at an early stage. English-speaking children, on the other hand, will have to have a lexically restricted V2 grammar (affecting *be* and a few other verbs), while at the same time entertaining a productive non-V2 grammar applying in all other cases. Roeper (1999: 184) also shows that there are many other "pockets of bilingualism … within Standard English", and this means that all monolingual speakers must have a grammar that has certain bilingual properties.

In the spirit of this idea, Yang (2002) has developed an approach to language acquisition called the Variational Model, combining UG and statistical learning (see also Yang 2010). On this view, children are endowed with a highly specified UG where all possible human grammars are represented; e.g. for pro-drop, children may choose

between an Italian-type pro-drop language (with rich agreement), a Chinese-type pro-drop language (which also allows object drop), and a non-prodrop grammar such as English. Like Snyder (2007), Yang (2010: 134) argues against a triggering approach to parameter setting, as this would predict "sudden qualitative and quantitative changes in children's production", which are generally not attested in child language data. Instead, children keep track of the input that favors one or the other grammar and use statistical evidence in the input to strengthen or demote them: For example, a child learning English will relatively quickly discard an Italian-type pro-drop grammar, as English does not have rich agreement and this is evident in almost every sentence. A Chinese-type pro-drop grammar will take somewhat longer to rule out, as the necessary evidence is only found in sentences with expletive subjects, and children encounter such sentences in the input much more rarely (1.2%, according to Yang 2010: 135).

According to Snyder (2007), the Variational Model is successful in analyzing areas of the grammar where children typically omit material that would be required in the adult language (such as subjects and objects). However, the model predicts "rampant errors of commission in other parts of the grammar" (Snyder 2007: 185). In my view, it is also problematic that all possible grammars have to be provided by UG. This would entail an extraordinarily high number of different grammars in UG, as e.g. not all null-subject grammars are like Italian or Chinese: For example, Russian, Hebrew and Inuktitut all allow null subjects under conditions that are somewhat different from those found in Italian and Chinese (Gordishevsky & Avrutin 2004, Allen & Schröder 2003). It has also recently been argued that there is systematic micro-variation between Spanish and Italian with respect to the interpretation of pronominal subjects (Filiaci, Sorace & Carreiras 2013). Nevertheless, children learning all these languages very early zoom in on the target grammar. Furthermore, to my knowledge, there is no evidence in the literature that (monolingual) children ever produce more overt subjects that the target language, which might have been expected if children could (more or less randomly) select any parameter setting provided by UG; in all cases they typically drop subjects slightly more than adults. In my view, this could simply be considered to be a result of children's general tendency for economy.

When there is variation in the input, some kind of grammar competition seems to be inevitable. But given findings from acquisition data such as those reported in Snyder (2007) and in the work of the Tromsø research group mentioned in section 2, children very early make the crucial distinctions that exist in the target grammar and produce the two (or more) options *in appropriate contexts*. To return to the example of V2 in Norwegian discussed in section 2, the children do not seem to be computing the overall percentages of V2 vs. non-V2 in the input and indiscriminately "weighing" the two grammars (in terms of the setting or re-setting of a macro-parameter), but are sensitive to the linguistic contexts that the different word orders appear in. Thus, given that children master these fine distinctions from early on, there is very little evidence for grammar competition in child language data. As I have argued in Westergaard (2014), this shows that children do not initially expect there to be competing forms in the input, but instead assume that they are exposed to principled variation and therefore try to figure out what this is based on. This means that grammar competition should not be the initial hypothesis of a child on exposure to variation, but rather a *last resort*, to be entertained only when children fail to find a distinguishing property between the options.

Nevertheless, there must obviously be some grammar competition in language, for example in cases where there is free variation in the target grammar, such as the optionality between V2 and non-V2 after the adverb *kanskje* 'maybe' (see section 2). In these cases, children seem to be quite good at statistical learning, producing the two options with similar frequencies as in the adult data from early on (96.4% (27/28) non-V2 compared to approximately 95% in the adult language; cf. Westergaard 2008b, 2009a). Thus, in my view, grammar competition (and the corresponding statistical learning) does not apply at the level of macroparameters, but should be restricted to cases of low-level variation, where the differences are quite small, affecting a subcategory, a feature or a lexical element rather than major categories.

This means that the "size" of rules is crucial. In Biberauer & Roberts (2012) and related work, the concept of parameter is broken down into a hierarchy of four distinct types, dependent on the size of the context in which they apply. That is, parameters may be macro, meso, micro or nano, depending on the class of elements that undergo the relevant process: 1) all elements of a given type, e.g. all heads in the language

(macroparameter), 2) a featurally specifiable subset of the elements of this category, e.g. all verbs or all nouns (mesoparameter), 3) the smallest definable sub-class of elements of this category, e.g. auxiliaries or pronouns (microparameter), and 4) one or more individual lexical items (nanoparameter).

Biberauer & Roberts (2012) mainly discuss verb movement, and identify the V-to-I movement operation found in earlier stages of English as well as many present-day languages, e.g. French, as a mesoparameter. This rule moves all heads of the verbal category to a higher functional position, resulting in a word order where the finite verb precedes negation and other adverbs, as shown in (13). In present-day English, this operation has been reduced to a microparameter, affecting auxiliaries only, i.e. a sub-class of verbal elements. An example of subject-auxiliary inversion is provided in (14).

(13)   if I **gave not** this accompt to you (Early Modern English, 1557)

        if I gave not this account to you

        'if I didn't give this account to you'

(14)   John **has not** kissed Mary.

        (Examples from Biberauer & Roberts 2012: 271-2)

The historical development from (13) to (14) also displayed a stage where certain lexical verbs still underwent the movement operation, e.g. *know* and *doubt*. This corresponds to a nanoparameter, affecting only specific lexical items in the language. Finally, at the other end of the hierarchy, Biberauer & Roberts (2012: 276) provide generalized head movement as an example of a macroparameter: The positive value of this parameter would entail syntactic movement to all heads, which is found in some polysynthetic languages. The other value of this macroparameter is found in languages that have no head movement at all, e.g. Mandarin and other Chinese varieties.

According to Biberauer & Roberts (2012), parameters are not given by UG, but considered to be emergent properties, resulting from the interaction of a minimal UG, the primary linguistic data and certain 3rd factors (Chomsky 2005), e.g. what Biberauer & Roberts refer to as acquisition strategies. In my view, breaking down major parameters

into processes that affect smaller parts of the grammar is a promising development within generative theory, given the micro-variation that has been found to exist between different languages, especially as a result of many dialect studies in recent years, and also given findings from language acquisition research showing that children cope quite well with such variation. Thus, I believe that in order to gain further understanding of language variation and language acquisition, our studies should focus on the micro-level; i.e. at the level where variation is dependent on fine linguistic distinctions between sub-classes of categories.

5. *A model of micro-cues*

In recent work, I have developed a model of language acquisition based on children formulating micro-cues in their I-language grammars (e.g. Westergaard 2009a, 2013). The model is inspired by Lightfoot's (1999, 2006) cue-based theory of acquisition and change. In this theory, a cue is a piece of abstract syntactic structure, formulated as in (15) for the word order Object-Verb or Verb-Object (corresponding to the head parameter) and (16) for V2 word order.

(15)   Cue for OV word order: $_{VP}$[DP V]

(16)   Cue for V2 word order: $_{CP}$[XP $_C$V...]

According to Lightfoot, the cues are provided by UG; thus, children know what to look for in the input that they are exposed to. This means that the cues do not correspond to input strings. The input serves as *triggers* for cues that are already present in UG, some of which will be activated in the language acquisition process and others that will not. Lightfoot (2006: 78) formulates it in the following way: "a sentence EXPRESSES a cue if the cue is unambiguously required for the analysis of the sentence." This means that the child's primary linguistic data are the triggering experience, while the cues are mental representations in the child's I-language.

Lightfoot's cue-based theory is similar to Fodor's (1998) idea of treelets as unambiguous triggers, mentioned in section 3 above. A treelet is similar to a cue in that it is defined as "a small piece of tree structure (a few nodes, perhaps only partially specified

in features; in the limiting case a single feature) that is made available by UG and is adopted into a learner's grammar if it proves essential for parsing input sentences." (Fodor 1998: 6). This means that both cues and treelets are assumed to be innate structural templates. In Fodor's system, the treelets represent parametric options provided by UG, and each treelet will thus trigger a specific parameter setting. In Lightfoot's theory, on the other hand, "cues … are the points of variation between grammars and there is no need for an independent notion of a parameter." (Lightfoot 2006: 78).

Lightfoot's cues are formulated in terms of major categories such as V or DP, and for this reason they make the same predictions as traditional macro-parameters. But the child language data discussed in section 2 show that children are sensitive to much finer distinctions than that. In my view, therefore, if a theory is to account for variation, it is necessary to formulate a number of much smaller cues, i.e. micro-cues. In the model of micro-cues, the context for a particular word order (e.g. V2 or non-V2) needs to be specified as part of the cue. This captures the fact that children do not only need to acquire a specific word order, but also the contexts in which this word order is relevant. Examples of such micro-cues are provided in (17)-(21), accounting for the micro-variation discussed in section 2.

(17) Micro-cue for V2 in wh-questions with monosyllabic wh-elements:

$$_{IntP}[\ _{Int'}[wh]\ _{TopP}[\ _{Top^o}[V\ldots XP_{[+FOC]}\ldots\ ]]]$$

(18) Micro-cue for V2 in declaratives: $_{DeclP}[XP\ _{Decl'}V\ \ldots]$

(19) Micro-cue for word order in declaratives with clause-initial *kanskje* 'maybe':
$$_{TopP}[\ kanskje\ XP\ \ldots\ _{VP}[\ V]]$$

(20) Micro-cue for subject shift: $_{InTopP}[\ DP\ _{[-FOC]}\ \ldots]$

(21) Micro-cue for N-POSS word order: $_{DP}[N\text{-}DET\ POSS\ _{[-FOC]}\ \ldots]$

Both Fodor (1998) and Lightfoot (1999, 2006) state that for cues or treelets to be able to trigger a particular structure (or parameter), they must be unambiguous. Fodor (1998: 6) formulates this as a principle of acquisition in the following way: "one absolute rule for language learners is *Do not learn from ambiguous input.*" She refers to her model as a "wait-and-see device", which, unlike the Triggering Learning Algorithm of Gibson & Wexler (1994), makes no changes in the child's grammar when the input is ambiguous. In the micro-cue model, this requirement for unambiguous cues is captured by adding relevant context into the formulation of the cue itself. This is also discussed in Westergaard (2008, 2009a), where I have argued that children search for micro-cues only in constructions or clause types where there is clear evidence. For example, in order to formulate a micro-cue for V2 word order, children only consider non-subject-initial sentences, as only these will contain relevant information, while SVO sentences will simply be disregarded in this respect. This means that children must have the ability to parse the input and distinguish between ambiguous and non-ambiguous cues. This ability will also reduce the amount of grammar competition in children's grammars. In the micro-cues formulated above, only (18) and (19) are in conflict with each other, specifying the word order in non-subject-initial declaratives (in general) and the particular word order found in declaratives introduced by the adverb *kanskje* 'maybe'. For the other micro-cues, the specification of the context ensures that there is no competition.

An important difference between Lightfoot's (1999, 2006) cue-based theory and Fodor's (1998) treelets on the one hand and the micro-cue model on the other is that the former models assume that the cues or treelets themselves are provided by the innate language faculty. In the micro-cue approach, on the other hand, the micro-cues are considered to be language-specific. In fact, they must be, given that they in some cases refer to particular lexical items, e.g. the micro-cue in (17), providing information about the word order in declaratives introduced by the adverb *kanskje* 'maybe'. Children are clearly sensitive to information at this level of detail from early on, and this must therefore be part of their linguistic knowledge, i.e. their I-language. This means that the micro-cues represent a speaker's knowledge of a *specific* language. However, the micro-cues are made up of syntactic primitives and built according to principles provided by

UG. This makes the micro-cue model different also from constructivist accounts, which typically claim that children are not endowed with any innate knowledge of categories or structure (e.g. Ambridge & Lieven 2011).

Thus, the micro-cue model is a generative approach which assumes the existence of a UG consisting of syntactic primitives (categories, features) and general principles of structure building.[1] This innate endowment is in some sense restricted compared to what is assumed in more traditional generative accounts, in that UG does not contain any parameters, nor does it provide the learner with any pre-built cues or treelets. But UG is still quite rich, in that it enables children to *parse* the primary linguistic data that they are exposed to. In parsing the input, children select the relevant primitives from the universal set and build syntactic structure based on the principles provided by UG. The micro-cues are then small pieces of syntactic structure that result from this parsing. In turn, they trigger the syntactic operations necessary to produce the relevant target structures, e.g. verb movement or subject shift. The language acquisition process is also affected by so-called 3rd factors (Chomsky 2005), for example general cognitive limitations (such as memory) or principles of economy, as discussed in section 3. Under this approach, language acquisition is considered to be what Snyder (2007) refers to as 'learning by parsing', and it typically results from an interaction between UG, input, and economy.

6. From specific to general and general to specific

One important issue within the field of first language acquisition is the question whether the child's development is from knowledge of general principles to knowledge about more specific details of the ambient language or the other way around. A traditional generative approach to acquisition assuming parameter setting will generally consider development to be from general to specific; that is, parameter setting is early and automatic and based on very little input (e.g. Wexler 1999), while any language–specific exceptions will have to be learned from more extensive exposure to the input and therefore take longer. Constructivist accounts, on the other hand, which assume no linguistically specific genetic endowment, argue that development is from specific to

---

[1] An anonymous reviewer asks how specific these features are in the micro-cue model. In my view, that is an empirical question, which we may find answers to by studying children's fine-grained distinctions that are not in the input (i.e. their 'undergeneralizations').

more general; that is, early multi-word utterances are initially item-specific, then develop into frames or schemas with slots for different word types, and only gradually become more abstract and general (e.g. Tomasello 2003, 2006).

The micro-cue model recognizes a relatively rich UG and argues that children's early grammars have categories/features, structure and rules. Nevertheless, the model is similar to constructivist accounts in that it assumes that development is from specific to general. The main reason for this is the conservative nature of children's production, indicating that they do not generalize a pattern or a rule until they have encountered positive evidence for this in the input.

In section 4, I discussed the new syntactic model proposed by Biberauer & Roberts (2012), where parameters are split up into four kinds of rules depending on size (macro, meso, micro and nano). The top of the parameter hierarchy involves less specific knowledge than the lower levels; that is, lower positions in the hierarchy have longer and more detailed descriptions and are therefore more complex. This means that setting a macroparameter is simpler than setting a parameter at the meso-level, which is again simpler than parameters at the micro- or nano-levels. There is especially one factor that is of importance in this respect, viz. the Input Generalization, formulated in the following way (Biberauer & Roberts 2012: 269, originally from Roberts 2007).

(22) Input Generalisation (IG): If a functional head F sets parameter P_j to value v_i then there is a preference for similar functional heads to set P_j to value v_i.

The Input Generalization ensures that there is a strong tendency for all functional heads to point in the same direction. Biberauer & Roberts (2012) relate the simplicity of this to the conservativeness of the child in the acquisition process. The higher levels represent the least amount of linguistic knowledge on the part of the learner and are thus assumed to "represent the acquirers' initial hypotheses" (Biberaurer & Roberts 2012: 270). Thus, these will be "automatically 'chosen' by the acquirer based on early 'ignorance'" (Biberauer & Roberts 2012: 270-271). The process of language acquisition then involves the learner moving down the hierarchy, making more and more fine-grained distinctions.

But children's conservative learning is usually used to refer to the opposite process, i.e. the lack of (over-)generalization. This was seen in the work referred to above (e.g. Snyder 2007, Westergaard 2013, Roeper 1999). And in the constructivist literature, there are numerous reports of especially experimental studies in which children are found not to generalize across the item-specific knowledge that they possess at a specific stage (see e.g. Ambridge & Lieven 2011). I would therefore argue that the Input Generalization cannot be a general property of language acquisition and that it must be severely modified.

Nevertheless, we also know that the adult language is not simply accumulated knowledge of a high number of specific constructions; it also consists of a productive grammatical system. This means that there must be a certain generalization taking place at some point in the acquisition process. In the micro-cue approach, this kind of generalization is considered to be development in a stepwise fashion. The crucial point is that these steps are small, involving only the addition of a new sub-category, a new lexical item or an extra feature. For example, given the cases of "undergeneralization" that we saw in section 3, an initial formulation of the micro-cue for word order in Norwegian or Swedish declaratives could be as in (23), specifying that V2 appears in sentences with the verb *be* and DP subjects. Since V2 is a more general process in the target version of the two languages, affecting all verbs and all subjects, this micro-cue must be extended to the formulation that we saw in (18), repeated here as (25), possibly with an intermediate stage where V2 is generalized to all subject types but still only applying to the verb *be*, as in (24).

(23)   Micro-cue for V2 in declaratives (initial version): $_{DeclP}$[XP $_{Decl'}$*be* …[DP …]

(24)   Micro-cue for V2 in declaratives (intermediate version): $_{DeclP}$[XP $_{Decl'}$*be* …]

(25)   Micro-cue for V2 in declaratives (adult version): $_{DeclP}$[XP $_{Decl'}$V …]

Similarly, the micro-cues for subject-auxiliary inversion in English wh-questions could be formulated as the developmental process illustrated in (26)-(28): This shows

development from an item-based process, affecting the wh-words *what* and *where* first as well as the lexical verb *be*, via a stage where the verbal element is generalized to also include auxiliaries (i.e. all elements that appear in the I position in English), and finally to a stage where the initial element is generalized to include all wh-items.

(26)   Micro-cue for inversion in wh-questions (initial version): $_{IntP}$[*what/where be …*]

(27)   Micro-cue for inversion in wh-questions (intermediate version):
        $_{IntP}$[*what/where* I …]

(28)   Micro-cue for inversion in wh-questions (adult version): $_{IntP}$[WH I …]

There may of course be many more steps in the process than what is indicated here, and the order may also be different. The duration of the various stages will vary from child to child and also clearly be dependent on the frequency with which a child is exposed to positive evidence in the input that the current formulation of a micro-cue should be generalized. However, given the general speed of language acquisition, the small steps in the development should typically be short-lived. For this reason, the evidence for this kind of development should be sought in very dense corpora of spontaneous child speech, which are unfortunately not abundant among the existing resources to date.

Finally, an important question is whether children never *over*generalize? Given their sensitivity to the input combined with conservative learning, examples of overgeneralization are also quite difficult to find in child language data. One that has been attested concerns verb movement in English, which, as I argued above, normally does not generalize beyond what there is positive evidence for in the input; i.e. it only extends from *be* to auxiliaries. Nevertheless, Roeper (1999) has attested occasional examples of inversion with the verbs *mean* and *call* in English child data, lasting for a very limited time (about a week), cf. examples (29) and (30).

(29)   What **means that**? (Roeper 1999: 175)

(30)   What **calls that**?

    Assuming that this child has learned inversion with *be* (given that this is an early acquisition, cf. above), the interesting issue here is that overgeneralization does not affect a major category, i.e. all verbs, but only takes place across a class or subcategory. That is, both *mean* and *call* are semantically similar to *be*, belonging to a class of equative verbs, and are therefore affected by this short-lived overgeneralization. Inversion with *mean* has also been found in data from other English-speaking children (Westergaard & Bentzen 2010, Westergaard forthcoming). An obvious advantage of this minor type of overgeneralization is that is reduces the need for "unlearning" in the language acquisition process.

7. Conclusion

In this paper, I have reviewed some early acquisition data where children are exposed to variation in the input, showing that young children are conservative learners, typically not (over-)generalizing across major categories. Such findings are difficult to explain in traditional parametric accounts of language acquisition, including theories of grammar competition. In the model of micro-cues, the acquisition data are accounted for in the following way: Children are endowed with a UG consisting of syntactic primitives and principles of structure building, which enables them to parse the input that they are exposed to. In the acquisition process, they build small pieces of abstract syntactic structure, the micro-cues, which become part of their knowledge of a specific language. The acquisition process is a development from specific to more general knowledge, and this development takes place in small steps based on positive evidence in the input, where each step represents the addition of a feature, a sub-category or one or more lexical items. This ensures that any overgeneralization will also be minor, reducing the need for "unlearning". The acquisition process is also affected by a principle of economy, accounting for the general lack of errors of commission in child language data.

References

Allen, S.E.M. & Schröder, H. 2003. Preferred argument structure in early Inuktitut spontaneous speech data. In J. W. Du Bois, L. E. Kumpf & W. J. Ashby (eds.), *Preferred argument structure: grammar as architecture for function*, 301-338. Amsterdam: John Benjamins.

Anderssen, M. 2006. *The acquisition of compositional definiteness in Norwegian.* PhD dissertation. University of Tromsø.

Anderssen, M., K. Bentzen, Y Rodina & M. Westergaard. 2010. The acquisition of apparent optionality: Word order in subject and object shift constructions in Norwegian. In M. Anderssen, K. Bentzen & M. Westergaard (eds.), *Variation in the Input: Studies in the Acquisition of Word Order.* [Studies in Theoretical Psycholinguistics 39], 241-270, Springer Verlag.

Anderssen, M. & K. Bentzen. 2012. Norwegian object shift as IP-internal topicalization. In K. Bentzen & A. Fábregas (eds.) *Nordlyd 39.1: The Grammar of Objects*, 1-23.

Anderssen, M., K. Bentzen & Y. Rodina**.** 2012. Topicality and complexity in the acquisition of Norwegian object shift. *Language Acquisition* 19.1: 39-72.

Anderssen, M., Y. Rodina, R. Mykhaylyk & P. Fikkert. 2014. The acquisition of the Dative Alternation in Norwegian. *Language Acquisition: A Journal of Developmental Linguistics* 21.1: 72-102.

Anderssen, M. & M. Westergaard. 2010. Frequency and economy in the acquisition of variable word order. *Lingua* 120.11: 2569-2588.

Anderssen, M. & M. Westergaard. 2012. Tospråklighet og ordstilling i norske possessivkonstruksjoner. [Bilingualism and word order in Norwegian possessive constructions]. *Norsk Lingvistisk Tidsskrift* 30: 170-197.

Ambridge, B. & E. Lieven. 2011. *Child language acquisition: contrasting theoretical approaches.* Cambridge: Cambridge University Press.

Bentzen, K., M. Anderssen & C. Waldmann. 2013. Object Shift in Mainland Scandinavian: A corpus study of Danish, Norwegian and Swedish. *Nordic Journal of Linguistics* 36(2): 115-151.

Biberauer, T. & I. Roberts. 2012. Towards a parameter hierarchy for auxiliaries: diachronic considerations. *Cambridge Occasional Papers in Linguistics* 6: 267–294.

Chomsky, Noam. 1981. *Lectures on Government and Binding.* Dordrecht: Foris.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use.* New York: Praeger.

Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry* 36: 1–22.

de Villiers, J. 1991. Why questions? In T.L. Maxfield & B. Plunkett (eds.), *Papers in the acquisition of* wh: *Proceedings of the Umass Roundtable, May 1990.* Amherst, MA: University of Massachsetts Occasional Papers.

Filiaci, F., A. Sorace & M. Carreiras. 2013. Anaphoric biases of null and overt subjects in Italian and Spanish: a cross-linguistic comparison. *Language and Cognitive Processes, Language, Cognition and Neuroscience*. DOI:10.1080/01690965.2013.801502.

Fodor, J. D. 1998. Unambiguous triggers. *Linguistic Inquiry* 29.1: 1-36.

Gelderen, E. van. 2004. *Grammaticalization as Economy. [Linguistik Aktuell 71]*. Amsterdam and Philadelphia: John Benjamins.

Gibson, P. & K. Wexler. 1994. Triggers. *Linguistic Inquiry* 25: 355-407.

Hyams. N. 1986. *Language acquisition and the theory of parameters*. Dordrecht: Reidel.

Lightfoot, D. 1999. *The Development of Language: Acquisition, Change and Evolution.* Malden, MA and Oxford: Blackwell.

Lightfoot, D. 2006. *How New Languages Emerge.* Cambridge University Press.

Lødrup, H. 2011. Norwegian possessive pronouns: Phrases, words or suffixes? In M. Butt & T. H. King (eds.), *Proceedings of the LFG11 Conference*, 383-403. Stanford: CSLI Publications.

Mykhaylyk, R. 2011. Middle Object Scrambling. *Journal of Slavic Linguistics* 19:2: 231-72.

Mykhaylyk, R. 2012. Factors contributing to child scrambling: Evidence from Ukrainian. *Journal of Child Language* 39:03: 553-579.

Mykhaylyk, R., Y. Rodina & M. Anderssen. 2013. Ditransitive constructions in Russian and Ukrainian: Effect of givenness on word order. *Lingua* 137: 271-289.

Roberts, I. 2007. *Diachronic Syntax.* Oxford: Oxford University Press.

Roeper, T. 1999. Universal bilingualism. *Bilingualism: Language and Cognition* 2 (3): 169-186.

Roeper, T. 2007. What frequency can do and what it can't. In I. Gülzow & N. Gagarina (eds.), *Frequency effects in language acquisition: Defining the limits of frequency as*

an explanatory concept [Studies on Language Acquisition], 23-48. Berlin: Mouton de Gruyter.

Snyder, W. 2007. *Child Language: The Parametric Approach.* Oxford: Oxford University Press.

Snyder, W. & D. Lillo-Martin. 2011. Principles and Parameters Theory and langugae acquisition. In P. Hogan (ed.) *The Cambridge Encyclopedia of the Language Sciences*, 670-673. Cambridge, UK: Cambridge University Press.

Sorace, A., L. Serratrice, F. Filiaci & M. Baldo. 2009. Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua* 119: 460-477.

Thornton, R. 2008. Why continuity? *Natural Language and Linguistic Theory* 26: 107-146.

Tomasello, M. 2003. *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. 2006. Acquiring linguistic constructions. In D. Kuhn & R. Siegler (eds), *Handbook of Child Psychology*, 255-298. Hoboken, NJ: Wiley.

Valian, V. 1990. Null subjects: A problem for parameter-setting models of language acquisition. *Cognition* 35.2: 105-122.

Valian, V. 1991. Syntactic subjects in the early speech of American and Italian children. *Cognition* 40: 21-81.

Vangsnes, Ø. A. 2006. Microparameters for Norwegian *wh*-grammars. *Linguistic Variation Yearbook* 5: 187-226. Amsterdam/Philadelphia: John Benjamins.

Waldmann, C. 2008. *Input och output: Ordföljd i svenska barns huvudsatser och bisatser*. PhD dissertation, Lundastudier A 65, University of Lund.

Waldmann, C. 2012. Moving in small steps towards verb second: A case study. *Nordic Journal of Linguistics* 34(3): 331–359.

Westergaard. M. 2004. The interaction of input and UG in the acquisition of verb movement in a dialect of Norwegian. *Nordlyd: Tromsø Working Papers in Language Acquisition* 32.1: 110-134.

Westergaard, M. 2008a. Verb movement and subject placement in the acquisition of word order: Pragmatics or structural economy? In P. Guijarro-Fuentes, P. Larranaga &

J. Clibbens (eds.), *First Language Acquisition of Morphology and Syntax: Perspectives across languages and learners*. [Language Acquisition and Language Disorders 45], 61-86, Amsterdam: John Benjamins.

Westergaard, M. 2008b. Acquisition and change: On the robustness of the triggering experience for word order cues. *Lingua* 118.12: 1841-1863.

Westergaard, M. 2009a. *The Acquisition of Word Order: Micro-cues, Information Structure and Economy.* [Linguistik Aktuell/Linguistics Today 145], Amsterdam: John Benjamins.

Westergaard, M. 2009b. Microvariation as diachrony: A view from acquisition. *Journal of Comparative Germanic Linguistics* 12.1: 49-79.

Westergaard, M. 2009c. Usage-based vs. rule-based learning: The acquisition of word order in *wh*-questions in English and Norwegian. *Journal of Child Language* 36.5: 1023-1051.

Westergaard, M. 2011. Subject positions and information structure: The effect of frequency on acquisition and change. *Studia Linguistica* 3: 299-332.

Westergaard, M. 2013. The acquisition of linguistic variation: Parameters vs. micro-cues. In T. Lohndal (ed.), *In Search of Universal Grammar: From Old Norse to Zoque*, 275-298. Amsterdam: John Benjamins.

Westergaard, M. 2014. Complexity and conflicting grammars in language acquisition. *Second Language Research* 30.1: 85-89.

Westergaard, M. Forthcoming. Word order and finiteness in acquisition: A study of English and Norwegian wh-questions. In K. M. Eide (ed.), *Finiteness matters. On finiteness-related phenomena in natural language.* Amsterdam: John Benjamins.

Westergaard, M. & K. Bentzen. 2007. The (Non-)effect of input frequency on the acquisition of word order in Norwegian embedded clauses. In I. Gülzow & N. Gagarina (eds.), *Frequency Effects in Language Acquisition: Defining the Limits of Frequency as an Explanatory Concept*, [Studies on Language Acquisition], 271-306. Berlin/New York: Mouton de Gruyter.

Westergaard, M. & K. Bentzen. 2010. Word order and finiteness in the acquisition of English and Norwegian wh-questions. In K. Franich, K. M. Iserman & L. L. Keil (eds.), *Proceedings of the 34th annual Boston University Conference on Language*

Development, Vol 2: 457-467. Somerville, MA: Cascadilla Press.

Westergaard, M., Ø. A. Vangsnes & T. Lohndal. 2012. Norwegian *som*: The complementizer that climbed to the matrix Left Periphery and caused Verb Second violations. In V. Bianchi & C. Chesi (eds.), *Enjoy Linguistics! Papers offered to Luigi Rizzi on the occasion of his 60th birthday,* 329-343. Siena: CISCL Press.

Westergaard, M. & M. Anderssen. Forthcoming. Word order variation in Norwegian possessive constructions: Bilingual acquisition and attrition. In J. B. Johannessen & J. Salmons (eds.), *Germanic heritage languages in North America: Acquisition, attrition and change* [Studies in Language Variation]. Amsterdam: John Benjamins.

Wexler, K. 1999. Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. In A. Sorace, C. Heycock & R. Shillock (eds.), *Language Acquisition: Knowledge Representation and Processing,* special issue of *Lingua,* 23-79. Amsterdam: Elsevier.

Yang, C. 2002. *Knowledge and Learning in Natural Language.* Oxford: Oxford University Press.

Yang, C. 2010. Universal Grammar, statistics or both? In C. Yang (ed.), *Language Acquisition: Critical Concepts in Linguistics* II, 128-140. London and New York: Routledge.