Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Lingua 118 (2008) 522–553
Lingua
www.elsevier.com/locate/lingua

Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling

John N. Williams *, Chieko Kuribara

Research Centre for English and Applied Linguistics, University of Cambridge, English Faculty Building, 9 West Road, Cambridge CB3 9DP, United Kingdom

Received 15 December 2006; received in revised form 5 March 2007; accepted 10 March 2007
Available online 19 July 2007

Abstract

The aim of the present study was to evaluate how successfully the UG (nativist) and connectionist (emergentist) frameworks can account for early L2 development, focusing on the acquisition of Japanese word order by adult native English speakers. We conducted a laboratory-based language learning study in which participants were exposed to a semi-artificial language based on Japanese, and we measured incidental learning of scrambling (and head-direction). Although there was some evidence of learning a generalised notion of ''free word order'', there was no evidence for accessing the relevant UG parameterised properties. The lack of clustering effects expected as a result of acquiring scrambling led us to conclude that adult SLA is not guided by UG. On the other hand, a connectionist simple recurrent network that was trained and tested on the same structures provided a close approximation to the participants' data, suggesting that they had acquired a good sense of the statistical structure of the input. Nevertheless, we argue that such a model cannot provide a complete account of learning the word order phenomena that we investigated without being supplemented by symbolic rule-learning mechanisms.

Keywords: Second language acquisition; Incidental learning; Universal grammar; Connectionism; Scrambling; Japanese

1. Introduction

Word order is more flexible in some languages than in others. Languages like English are relatively rigid; whereas, languages such as Japanese are considerably more flexible: the verb must come at the end of the clause, but the order of other phrases is free.

* Corresponding author.
E-mail addresses: jnw12@cam.ac.uk (J.N. Williams), ckuribara@hotmail.com (C. Kuribara).

0024-3841/$ – see front matter
doi:10.1016/j.lingua.2007.03.003

(1) a. Jon-ga ano resutoran-de piza-o tabe-ta.
John-nom that restaurant-in pizza-acc eat-past
'John ate pizza in that restaurant.'

b. Jon-ga piza$_i$-o $[_{VP}$ ano resutoran-de t$_i$ tabe-ta].
John-nom pizza-o that restaurant-in eat-past

c. Ano resutoran$_i$-de $[_{IP}$ Jon-ga t$_i$ piza-o tabe-ta].
that restaurant-in John-nom pizza-acc eat-past

d. Piza$_i$-o $[_{IP}$ Jon-ga ano resutoran-de t$_i$ tabe-ta].
pizza-acc John-nom that restaurant-in eat-past

However, this does not mean that there is no base order in this language. All the sentences (1b–d) are considered to be derived from the canonical word order, (1a), through the application of a syntactic operation called ''scrambling'' (see Nemoto, 1999 for a detailed summary on the nature of scrambling).

Although scrambled orders are possible, scrambling sentences are extremely rare in the input (see Iwasaki (2003) for a summary of previous corpus studies of Japanese; Chung (1994) reports the frequency of scrambling sentences produced by a Korean mother, which accounts for less than 1% of all of her utterances). However, in spite of impoverished positive evidence, children successfully acquire scrambling. A recent study by Murasugi and Kawamura (2005) suggests that scrambling may be acquired at around the age of two, as early as the acquisition of canonical word order. Thus, scrambling can be seen as an aspect of grammar that reveals that frequency is not a crucial factor in child language acquisition. The child's early successful acquisition despite the limited availability of relevant input calls for the presence of innate linguistic knowledge, which we assume is in the form of Universal Grammar (UG).

In contrast, frequency seems to play a crucial role in adult L2 acquisition. Iwasaki (2003), having conducted a study on acquisition of Japanese SOV and OSV order by English speakers, reports that the accuracy of the L2 learners' performance does not increase with their proficiency level.[1] She also reports substantial individual differences in their knowledge of scrambling. Iwasaki suspects that some of her participants think that only canonical SOV is a possible word order in Japanese and that one reason for this may be that scrambled sentences are infrequent in the input.

The above observation leads us to a question: what is it that causes this difference? One possible answer is provided by the Fundamental Difference Hypothesis or the no parameter resetting position, whereby adult SLA is carried out by L1 transfer and a non-domain-specific learning system (Bley-Vroman, 1989, 1990; Clahsen and Muysken, 1989; Tsimpli and Smith, 1991; Tsimpli and Roussou, 1991; Neeleman and Weerman, 1997). Another candidate is the Full Transfer/Full Access Model put forward by Schwartz and Sprouse (1994, 1996). Adult learners make use of all the properties available from their L1, and acquire their target language by gradually re-setting parameters. The re-setting is motivated by the failure to assign a representation to the input with the learners' syntactic knowledge at hand; therefore, it may not always be correctly carried out if relevant input is obscure or rare. This implies that the interlanguage grammar may differ from the target grammar during the course of development (and/or at ultimate attainment), but the interlanguage, consisting of a set of parameterised UG properties, will always be UG-constrained.

[1] Iwasaki conducted a series of tasks: a picture description task, a fill-in-the-blank task, a grammaticality judgement task, and an interview. These were designed to elicit her subjects' knowledge of case particles (i.e. subject and object case markers) in SOV and OSV sentences.

Both models claim that SLA is (at least initially) carried out by L1 transfer where possible. However, the models crucially differ as to how the grammar develops from there. In other words, how do L2 learners get to learn structures that are not (mis-) analysable by their L1 grammar? According to Schwartz and Sprouse, if input provides clear and sufficient evidence, learners would be able to reset a parameter to a correct value, leading to the acquisition of all underlyingly related structures (Meisel, 1995). This is referred to as a ''clustering effect'' in the relevant literature. A lack of the clustering effect would imply that adult SLA is not achieved by parameter resetting, and this would serve as evidence for the Fundamental Difference Hypothesis.

According to this hypothesis, the reason why frequency plays a crucial role in adult SLA is to do with the involvement of a domain-general learning system. As far as adult SLA is concerned, therefore, the Fundamental Difference Hypothesis shares a common view with what Gregg (2003) has called ''empiricist emergentism'', as exemplified by connectionism, an approach that claims that everything that is learned is induced from input alone. In this article we shall use a common data set to explore a frequency-based account of acquisition of Japanese word order within a connectionist model, whilst at the same time addressing the issue of UG availability.

Testing the validity of the connectionist approach requires us to compare human learning with that of a connectionist network. An example of this kind of work within SLA is a study by Ellis and Schmidt (1998). Using a miniature linguistic system, they demonstrated the well-known regularity by frequency interaction in morphological processing and learning, whereby frequency affects processing of irregular items but not regular ones. They showed that a connectionist model, trained and tested on the same data as their human participants, reproduces the same effect in learning. We follow the same strategy here, but in the domain of syntax.

The aim of the present study was to evaluate how successfully the UG (nativist) and connectionist (emergentist) frameworks can account for early L2 development, focusing on the acquisition of Japanese word order by adult native English speakers. We conducted a laboratory-based language study in which participants were exposed to a semi-artificial language based on Japanese in the context of a meaning-focussed task, and measured how well they incidentally learned word order properties using a subsequent grammaticality judgement task. A laboratory-based language learning situation was used so as to have total control over the input structures, and their frequencies. A semi-artificial language consisting of English lexis and Japanese syntax was used so as to eliminate the burden of learning and processing novel vocabulary, and to allow us to examine syntax learning even on initial contact with the language.

This article is organised as follows. First we describe the relevant aspects of Japanese and English, and lay out the theoretical assumptions behind the design of the materials. We then describe the predictions of the UG framework before moving onto a description of the study. The results are first reported and discussed from the perspective of the UG framework, and then evaluated against a connectionist network. The adequacy of the connectionist approach as an explanation of SLA is considered in section 7.

2. Theoretical assumptions

As has been observed in example (1), scrambling can take place within the VP, e.g. the movement of direct object over indirect object or VP-adjuncts, as shown in (1b).[2] A phrase may move to the clause-initial position, as exemplified in (1c) and (1d). A phrase can scramble

[2] Miyagawa (1997) argues that both orders, i.e. Subject–Indirect Object–Direct Object–Verb, and Subject–Direct Object–Indirect Object–Verb, are base-generated.

long-distance, moving across a clausal boundary to the sentence initial position of a matrix clause, as shown in (2b).

(2) a. Mearii-ga [Jon-ga ano resutoran-de piza-o tabe-ta to] omot-ta.
Mary-nom John-nom that restaurant-in pizza-acc eat-past that think-past
'Mary thought that John ate pizza in that restaurant.'
b. Piza$_i$-o Mearii-ga [Jon-ga ano resutoran-de t_i tabe-ta to] omot-ta.
pizza-acc Mary-nom John-nom that restaurant-in eat-past that think-past

Scrambling can also be applied to wh-phrases. Japanese is a wh in situ language. Thus wh-phrases remain in their base-generated positions, as shown in (3a); however, they can be moved to the three different positions described above, just like their non-wh-counterparts.

(3) a. Jon-ga ano resutoran-de nani-o tabe-ta no.
John-nom that restaurant-in what-acc eat-past Q
'What did John eat in that restaurant?'
b. Jon-ga nani$_i$-o [$_{VP}$ ano resutoran-de t_i tabe-ta] no.
John-nom what-o that restaurant-in eat-past Q
c. Nani$_i$-o [$_{IP}$ Jon-ga ano resutoran-de t_i tabe-ta] no.
what-acc John-nom that restaurant-in eat-past Q
d. Nani$_i$-o Mearii-ga [$_{IP}$ Jon-ga ano resutoran-de t_i tabe-ta to] omot-ta no.
what-acc Mary-nom John-nom that restaurant-in eat-past that think-past Q
'What did Mary think John ate in that restaurant?'

It has been of much dispute as to whether the syntactic movements described above essentially involve the same operation,[3] whether these movements are feature-driven (hence obligatory) or optional,[4] and where exactly the scrambled phrases land.[5] However, we will base our study on Saito and Fukui's (1998) parametric account of English and Japanese, where the head parameter determines the distribution of scrambling. That is, setting the head parameter has a consequence for the formation of phrase structure, which is beyond simply determining the head position. The theory is desirable from the child language acquisition point of view: the lower the number of parameters, the fewer the parameters there are for the child to deal with (Neeleman and Weerman, 1997).

Saito & Fukui (S&F) argue that scrambling is an optional operation, by showing that it exhibits different characteristics from those of operator-variable movements like topicalisation and wh-movement in English. These movements are superficially similar to scrambling in that they move a phrase to the left as in (4b) and (5b), and they do so also across a clause boundary, as in (6b) and (7b):

(4) a. Ano hon$_i$-o Jon-ga t_i yon-da.
that book-acc John-nom read-past
b. That book$_i$, John read t_i.

[3] See Miyagawa (1997, 2001, 2003) and Saito (2004) for a relevant discussion.

[4] See Fukui (1995), Kuroda (1988), Saito (1985, 2002, 2004), Saito and Fukui (1998), Tada (1993) among others for arguments for the optional nature of scrambling. Opponents of this view are Miyagawa (1997, 2001, 2003), Grewendorf and Sabel (1999), and Kawamura (2001).

[5] For example, see Saito (2002), Grewendorf and Sabel (1999) and Kawamura (2001) for varied opinions.

(5) a. Nani$_i$-o Jon-ga t_i tabe-ta no.
what-acc John-nom eat-past Q
b. What$_i$ did John eat t_i ?

(6) a. Ano hon$_i$-o Mearii-ga [John-ga t_i yon-da to] omot-tei-ru.
that book-acc Mary-nom John-nom read-past that think-asp-pres
b. That book$_i$, Mary thinks that John read t_i.

(7) a. Nani$_i$-o Tim-ga [otooto-ga t_i kowashi-ta to] it-ta no.
what-acc Tim-nom young brother-nom break-past that say-past Q
b. What$_i$ did Tim say that his brother broke t_i ?

However, topicalisation and wh-movement are different from scrambling. They are feature-driven movements and subject to the Minimal Link Condition, where a syntactic requirement must be met by the closest constituent if there is more than one constituent of the relevant kind (Chomsky and Lasnik, 1993). This means that the movement of a constituent is blocked if there is a constituent of the relevant kind which is closer to the landing site. This constraint does not apply to scrambling, which makes sense if we assume that the movement is not triggered by formal features.

Therefore, no more than one constituent can be moved via topicalisation and wh-movement, whilst there is no such limit for scrambling. This is demonstrated in (8) and (9).

(8) a. John-ni$_i$ sono hon$_j$-o Mearii-ga t_i t_j watashita.
b. ??To John$_j$, that book$_i$, Mary handed t_i t_j.
(Saito and Fukui, 1998, p. 444)

(9) a. Dare$_i$-ni nani$_j$-o Mearii-ga t_i t_j watashita no.
b. ??To whom$_j$, what$_i$ did Mary hand t_i t_j?

The fact that scrambling does not yield superiority effects whilst wh-movement does also follows from the same principle.

(10) a. Dare-ga nani-o kat-ta no.
who-nom what-acc buy-past Q
b. Who bought what?

(11) a. Nani-o dare-ga katta no.
what-acc who-nom buy-past Q
b. *What did who buy?

S&F point out that English has a syntactic operation equivalent to Japanese scrambling. Heavy NP shift can move wh-phrases as in (12), and also allows multiple application as in (13). Only the difference is that heavy NP shift moves a phrase to the right, instead of the left.

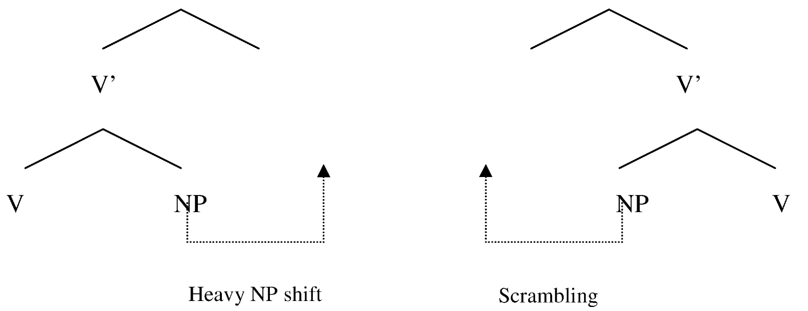
(12) Who$_i$ t_i borrowed t_j from the library [which book that David assigned in class]$_j$?

(13) John told t_i t_j yesterday [a most incredible story]$_i$ [to practically everyone who was willing to listen]$_j$.
(Saito and Fukui, 1998, p. 225)

According to S&F, the reason why Japanese scrambling and its English counterpart move a phrase in the opposite direction to each other is to do with the setting of the head parameter in each language. English selects [head-initial], thus forming a right-branching phrase structure; whereas Japanese selects [head-final], therefore forming a left-branching structure:

(14) a. John [$_{VP}$ ate pizza].
b. Jon-ga [$_{VP}$ piza-o tabe-ta].
John-nom pizza-acc eat-past
"John ate Pizza."

(15) a. English: head-initial b. Japanese: head-final



Scrambling and heavy NP shift preserve the headedness values, and thus they are considered to be special cases of Merge (the term/notion introduced by Chomsky (1994)). Merge is costless, and therefore does not need any driving force (Chomsky, 1995); hence the operation is optional. Only the movements that do not conform to the head parameter values need to be triggered by feature checking. Thus, wh-movement and topicalisation in English necessarily involve some features in relevant functional heads. Although S&F are not specific about the identities of the features and functional heads, we assume that wh-movement is motivated by the strong [+wh] feature of C, and that topicalisation is triggered by some feature associated with Topic Phrase (TopP), which is located between CP and IP (Muller and Sternefeld, 1993).

To summarise, an optional movement is available in English as well as in Japanese. The operation is basically a special case of Merge. The direction of the movement/Merge is determined by the value of the head parameter. Japanese is right-headed thus allows phrases to be scrambled to the left; whilst English is left-headed thus allows phrases to be scrambled to the right. Wh-movement and topicalisation in English are superficially similar to Japanese scrambling, in the sense that they move a phrase to the sentence/clause-initial position and even across a clause boundary. However, they are different from scrambling, since they involve a movement to the Spec of a functional projection. Similarities and differences between scrambling and wh-movement/topicalisation are summarised in Table 1.

3. Research question and hypotheses

With the assumption that child language acquisition is a process of setting UG parameters into the values appropriate for a given input, the aim of the current research is to find out whether the early development of adult SLA is also guided by the same mechanism. If not, what do L2 learners acquire? Is it that their performance is a reflection of probability-based

Table 1
A summary of syntactic characteristics associated with leftward movement available in Japanese and English

	Scrambling in Japanese	Wh-movement/topicalisation in English
Type of phrase	Wh and non-Wh	Wh-phrase/non-Wh
Landing site	IP and VP	CP/TopP
Long distance?	Yes	Yes
No of application	Single and multiple	Single only
Superiority effect?	No	Yes

morphophonological patterns that they encounter while they are exposed to a target language? Or, do the learners generalise the observed patterns and formulate some rules?

We tackle these issues by investigating the acquisition of word order, i.e. head-direction and scrambling, in Japanese by native speakers of English. We base our study on Saito and Fukui (1998), who claim that the value for the head parameter determines in which direction free recursion of the X'-level or Merge can take place, so that the representation can be built to accommodate an optional movement. In the case of Japanese, scrambling, an optional movement, moves a phrase to the left, since the language, being head-final, can expand a phrase structure only to that direction. If this theory is on the right track, the correlation should be manifested in the child's language as follows. Suppose that the head-direction is already set to head-final and therefore that free recursion is possible to the left of the head. Suppose also that scrambling is already available in the child grammar. Then, as soon as the representation up to the IP level is built, the landing site for the sentence-initial scrambling becomes available. Scrambling can now move a phrase over the subject to the left hand side of the Infl. This prediction seems to be empirically supported by child language acquisition data.

A recent study by Murasugi and Kawamura (2005) reports that children who interpret active sentences (SOV) correctly, also perform accurately on scrambled sentences (OSV). The results of the act-out experiment have shown that a child at the age of two scored as high as 83% of the time for both sentence types. Murasugi and Kawamura suggest that the acquisition of scrambling can be as early as the acquisition of basic sentences. That is, once the canonical word order is acquired, the scrambled order is also possible in the child's grammar.

It has also been argued in the child language acquisition literature, that the head-direction parameter is set at a very early stage, i.e. as soon as the child starts producing the Verb and its complement (about 18 months old according to Wexler, 1998), and that the word order does not fluctuate (for the OV order, De Haan, 1987 for Dutch; Poeppel and Wexler, 1993 for German; Sugisaki, 2005 for Japanese; for the VO order, Bloom, 1970; Brown, 1973; Lebeaux, 1989, for English). Thus, it can be concluded that the child requires an extremely short period of time (less than several months) for setting head-direction and acquiring scrambling, and that s/he does so at an extremely young age (before or at around the age of two).

If the same mechanism works for adult SLA, it is predicted that leftward scrambling and the head-final setting of the head-direction parameter should be available, and hence manifest in learner language, as soon as the canonical word order of Japanese is acquired. It is then expected that in the data of L2 learners who have already learned the SOV order, the following patterns should be observed:

(i) Movement has the characteristics of scrambling, i.e. properties of a non-operator-variable operation,

(ii) verb-complement order is clearly (re-)set to [head-final], and
(iii) movement takes place only to the left.

It has been widely acknowledged that L2 learners utilise their L1 properties in order to (mis-) analyse L2. In particular, Schwartz and Sprouse specifically state that learners transfer all the syntactic properties available in their native language, i.e. projection up to the CP level in the case of native English speakers. As we have seen in the previous section, English has syntactic operations that are superficially similar to scrambling: wh-movement and topicalisation, both of which move a phrase to the left within and across a clause. Therefore, it is possible that having observed this surface similarity, the learners may apply the syntactic options, which leads to the apparent acquisition of scrambling (cf. Hawkins and Hattori, 2006; Kuribara, 2000, 2004; Miyamoto and Iijima, 2003). Wh-movement and topicalisation, being operator-variable movement, show different characteristics from those of scrambling. As described above, neither wh-movement nor topicalisation allows multiple application, and wh-movement is subject to superiority; in contrast, such constraints do not apply to the scrambling operation. These properties can be used to test the true nature of the learners' knowledge of scrambling. If it is the case that the learners use L1 syntactic options in order to analyse scrambled structures, they would show a dispreference for the sentences which violate superiority and those which involve multiple movements. If this scenario is the correct one, it implies that the learners have somehow failed to identify scrambling despite the fact that there is an equivalent operation in English, and/or to (re-)set the head parameter into [head-final] (as a result of which, no landing site is available on the left for the scrambling operation).

A question still arises for the case where no superiority and multiple application effects are present in the data. Would this mean that the learners have accessed the knowledge of the optional movement and managed to (re-)set the head parameter? Caution is again called for; their apparent success may be to do with the use of some non-linguistic cognitive ability. If scrambling is truly available for adult L2 learners, their data should also meet the rest of the conditions listed above (i.e. the expected patterns (ii) and (iii), as well as (i)), as observed in the child L1 acquisition literature. On the other hand, if there is no such clustering, we take this as evidence that the acquisition of word order in the early stage of adult SLA is not guided by UG.

4. Method

As outlined in section 1, we used stimuli in which Japanese word order and case marking were combined with English lexis. We shall refer to this language as ‘‘Japlish’’. In the exposure phase participants performed semantic plausibility judgements on Japlish sentences. Thus, any learning of syntax was incidental to the meaning-focussed task in which they were engaged. In the test phase they were then given a (surprise) grammaticality judgement task. Grammaticality judgement performance was compared to that of a second group of participants who did not receive any prior exposure to Japlish and just performed the test phase. The performance of this *no exposure group* informs us of their initial state, and provides a baseline for measuring learning effects in the *exposure group*.

4.1. Participants

A total of 41 undergraduate and postgraduate students from the University of Cambridge participated in the experiment. They were drawn from a variety of disciplines and were all native

Table 2
The structures used in the exposure phase

Structure type	Structures (and number of items)	Examples
Simple canonical	SV (20), SOV (28), SIOV (22)	Horse-ga when fell? Pilot-ga that runway-o saw. Student-ga dog-ni what-o offered?
Simple canonical multiple wh-questions	S when what-o V? (10), S who-ni what-o V? (10)	Bill-ga when what-o sang? That doctor-ga who-ni what-o showed?
Complex canonical	S [SOV]V (20), S[SIOV]V (20)	John-ga angrily Mary-ga that ring-o lost that said.
Short IP-scrambling	**Adj**$_i$ S t_i V (8), **O**$_i$ S t_i V (16), **O**$_i$ S I t_i V (16)	When Bill-ga danced? That sandwich-o John-ga ate.
Short VP-scrambling	S **O**$_i$ I t_i V (8)	Vet-ga injection-o elephant-ni gave.
Long IP-scrambling	**O**$_i$ S[S t_i V]V (8) **O**$_i$ S[S I t_i V]V (8)	That disease-o vet-ga cow-ga have that declared. What-o Mary-ga professor-ga students-ni taught that said?

Note: S = subject, V = verb, I = indirect object, O = direct object, Adj = adjunct, t = trace (the position which a scrambled phrase originates from). The number of sentences corresponding to each type of structure is indicated in brackets. Segments in bold are those that have undergone movement.

speakers of English who had no knowledge of Japanese. Sixteen of them formed the no-exposure group, mean age 21, and 25 formed the exposure group, mean age 22.

4.2. Materials

4.2.1. Exposure structures

The structures used in the exposure phase, and the number of items per structure, are listed in Table 2. We shall use *S* for subject, *O* for direct object, *I* for indirect object, *V* for verb, *Adj* for adjunct, and *t* for trace (the position that a scrambled phrase originates from). For convenience, we use the terms ‘short’ and ‘long’ instead of ‘within-clause’ and ‘long-distance’ in referring to different types of scrambling.

Adjunct phrases were inserted after S for some of the sentences, indicating to the learners that they are optional constituents. Apart from the multiple wh-question structure, half of the sentences contained a wh-phrase. The fact that half of the instances of movement involve wh-phrases and half non-wh-phrases indicates to the learners that the movement can be applied to both kinds of phrases. Plenty of canonical structures where both wh- and non-wh-phrases stay in situ indicate to the participants that movement is non-obligatory. All the embedded clauses involve a ‘that’ complement clause in the clause final position.[6] Note that every sentence and

[6] A reviewer suggested that the use of *that* for the complementizer in Japlish may not be appropriate, because ‘that’ is a free morpheme in English, whereas an important aspect of consistently head-final languages, like Japanese and most of the Asian languages is that the complementizer occurs as a suffix at the end of the clause. The reviewer also points out that those languages have complex inflectional suffixation on the verb, another property which is missing from Japlish. Since these morphological characteristics are missing, the participants might have taken Japlish to be a mixed language in terms of headedness (such as German or Chinese). However, our understanding is that consistently head-final languages “tend” to be agglutinative (Lehmann, 1973), which means that such languages “do not need” to be agglutinative, i.e. agglutination is not a necessary condition for a language to be consistently head-final. Besides, all the sentences/clauses we presented to our participants during the exposure phase were V-final, and the complementizer appeared in the clause-final position. Nevertheless we accept that it would be interesting to see if different results would be obtained if a different morpheme were used instead of ‘that’ (although not Japanese *-to* for obvious reasons).

clause ends with a verb. Because the participants' task was to perform a plausibility judgement on each sentence, half of the sentences were semantically plausible, and half were implausible. Examples of implausible sentences are: "Simon-ga which bowl-o ate?", "Applicant-ga company-ni which job-o offered?". Note that in the latter type of example, the participants had to rely on the Case system in order to make the correct plausibility judgement. Finally, note that one third of the exposure sentences had scrambled word order.

4.2.2. Test constructions and predicted outcomes

The present study focuses on the acquisition of IP-scrambling, a type of scrambling that moves a phrase to a sentence-/clause-initial position. Test constructions and some examples are presented in Table 3.[7] The test items that repeat an exposure pattern are categorised as 'Trained'; the items that only appear in the test phase are categorised as 'New'. Each construction was tested with four sentences. With the exception of the items relevant to superiority effects, two sentences contained a non-wh-phrase and the other two contained a wh-phrase. All of the test sentences were semantically plausible. Segments in bold are those that have undergone movement. Note that a construction with an indirect object is counted as a separate construction from the one without. For example, the total number of test sentences for the Complex canonical structures is eight, rather than four. The items relevant to superiority effects were an exception. There were four items of each type (Canonical, Reverse), two with an indirect object and two without. There were 88 test sentences altogether.

The scrambled structures that the learners come across during the exposure phase only involve movement of the direct object to the sentence initial position. If the learners manage to internalise this movement operation in some way, instead of accepting the relevant constructions by simply relying on their memory, it would be expected that they also accept scrambling constructions that they have never come across before (i.e. I_iSt_iOV, $S[O_iS(I)t_iV]V$, and $I_iS[St_iOV]V$).

However, we need to determine the nature of the movement. If it is some sort of operator-variable movement which is available in the learners' L1 such as topicalisation and wh-movement, then the learners would not allow multiple applications. In addition, they would show superiority effects by rejecting the Wh-O Wh-S order whilst accepting the Wh-S Wh-O order.

Nevertheless, even if those effects are absent, it does not yet guarantee that the movement is scrambling. Leftward scrambling should correlate with the [head-final] setting of the head parameter. The setting of the head parameter into the [head-final] value should lead the learners to accept the canonical structures, S[S(I)OV]V. However, since these constructions appear in the exposure phase, it is possible that the learners accept them by simply relying on their memory. Therefore, it is important to see if they also successfully reject the English V-initial order (*SV(I)O). Nevertheless, the English canonical structures can be analysed as having the combination of the [head-final] setting and the right movement of object(s), i.e. $*S\ t_i\ t_j\ V(I_i)O_j$. But if the head parameter is truly (re-)set to the correct value, then scrambling is licensed to the

[7] A reviewer suggested that the use of a demonstrative, *that*, in addition to a complementizer, *that* (see examples in Table 3e) might have caused a problem for the participants when they try to analyse test sentences: because of the morphological similarity and the non-English word order, the participants might have misanalysed the demonstrative as the complementizer, or vice versa, in some of our test sentences. However, we argue that this kind of misanalysis is unlikely to have taken place, since a picture provided for each sentence included the demonstrative with the rest of the noun phrase (e.g. *that ring-o*), but not the complementizer.

Table 3
The structures used in the test phase

a. Grammatical items to test whether any learning of canonical structures has taken place

‘Trained’
Complex canonical: S [S (I) O V] V

e.g. John-ga nurse-ga surgeon-ni knife-o handed that claimed.

b. Grammatical items to test whether any learning of direct object scrambling has taken place, and whether the knowledge extends to the case of indirect object scrambling

‘Trained’	‘New’
Simple short scrambling: **O**$_i$ S (I) t$_i$ V	Simple short scrambling: **I**$_i$ St$_i$ O V Complex short scrambling: S[**O**$_i$ S (I) t$_i$ V]V
Long scrambling: **O**$_i$ S [S (I) t$_i$ V] V	Long scrambling: **I**$_i$ S [S t$_i$ O V] V

e.g. Horse-ni farmer-ga hay-o gave., Who-ni waiter-ga customer-ga tip-o handed that said?

c. Grammatical items to test the presence of multiple application effects

Single application (New)	Multiple application (New)
I$_i$ S t$_i$ O V, **Adj**$_i$ S t$_i$ O V	**I**$_i$ **O**$_j$ S t$_i$ t$_j$ V, **Adj**$_i$**O**$_j$S t$_i$ t$_j$ V

e.g. That girl-ni scarf-o John-ga gave., When what-o Mary-ga read?

d. Grammatical items to test the presence of superiority effects

Canonical (New)	Reverse (New)
Wh-S Wh-O (I) V	Wh-O$_i$ Wh-S (I) t$_i$ V

e.g. What-o who-ga John-ni showed?

e. Ungrammatical items to test the directionality of movement and headedness

Right movement of S	Right movement of O	English canonical structures
*t$_i$ I O V **S**$_i$	*S I t$_i$ V **O**$_i$	*S V (I) O
*S [t$_i$ O V **S**$_i$] V	*S [S t$_i$ V **O**$_i$] V	

e.g. Artist-ga friend-ni showed that picture-o., Bill-ga chef-ga cooked what-o that said?

Note: Each construction was tested with four sentences. Apart from the items relevant to superiority effects, two sentences contained a non-wh-phrase and the other two contained a wh-phrase. Segments in bold are those that have undergone movement.

opposite direction from the head (i.e. left), which should automatically deter scrambling to the right. This would lead the learners to reject all the ungrammatical structures listed in Table 3e.[8]

It then follows that if the learners successfully build syntactic representations that conform to the parametric options selected by Japlish, they should be able to correctly accept all the grammatical constructions and reject ungrammatical ones. We will interpret such a data pattern to imply that the adult learners have acquired Japlish word order by accessing the mechanism endowed by UG. If, on the other hand, the learners fail to show such a pattern, there are grounds to argue that that is not the case.

The sentences were to be presented in both spoken and written form and were recorded by the first author (a native speaker of British English) using the natural stress patterns of Japanese (as described in Ishihara, 2001).[9]

4.3. Procedure

4.3.1. Exposure phase

This phase of the experiment was divided into four blocks of trials, moving from simple canonical, to complex canonical, and finally to a mixture of simple and complex canonical and scrambled sentences. Each block was preceded by written instructions which were on paper, and available for the participants to consult only during that block. The first set of instructions explained the function of the case markers -ga (subject), -o (direct object), and -ni (indirect object). They were given one example SIOV sentence, and instructed on how to perform the plausibility judgement task. They then performed the plausibility judgement task on the first block of 52 simple canonical sentences. For the second block the instructions explained that in the example sentence ‘‘Fred-ga John-ga apple-o ate that said’’ Fred did the saying and John did the eating. Block 2 contained 22 complex canonical sentences. The instructions for Block 3 simply said that simple and complex sentences would be mixed up. Block 3 contained 9 complex and 18 simple sentences. Block 4 was preceded by the following instructions: ‘‘In this section your task is the same as before, and the sentences are no more complicated than those you have already had. But the words will start to come up in different orders. Still, by paying attention to the word endings it should be possible to work out the meaning of the sentences whatever the word order’’. No example sentences were provided. Block 4 contained 20 simple canonical, 9 complex canonical, 48 simple scrambled sentences, and 16 complex scrambled sentences. Note that the first 101 out of a total of 194 sentences had canonical word order. We hoped that this would help the learners to establish the canonical order of Japlish prior to receiving scrambled structures. The sentences within each block were presented in an individually randomised order for each participant.

Each sentence was presented both auditorily and visually. The visual presentation remained on the screen until the participant made their plausibility judgement (by pressing either of two designated keys on the keyboard). If their plausibility judgement was incorrect the sentence was repeated until a correct response was entered.

[8] Japanese actually allows structures such as (I)OVS and S(I)VO. Nevertheless, these word orders are not possible inside the embedded clause, and moreover, the dislocated phrases, S and O respectively, cannot be wh-phrases (Sugisaki, 2005).

[9] For declarative sentences, the main stress was placed on the phrase immediately preceding the verb, following the ‘‘nuclear stress rule’’. For wh-questions, we followed the ‘‘additional stress assignment and deaccenting’’ rules, where the wh-word received an additional stress and the rest of the words were deaccented. Where there were two wh-words, both were stressed.

4.3.2. Test phase

The test phase procedure was identical for the exposure and no-exposure groups apart from slight differences in the instructions. Participants made grammaticality judgements on sentences presented auditorily and visually. Unlike in the exposure phase, the sentences disappeared at the offset of the auditory sentence. This was to encourage the participants to make immediate decisions. After responding (using one key for 'likely to be grammatical' and another for 'unlikely to be grammatical') they also had the option of repeating the trial (in theory they could do this as many times as they wished, although in practice they rarely did so more than once). This was felt to be necessary because of the processing and memory demands imposed by the longer sentences. In order to facilitate comprehension (especially for the no-exposure group) each sentence was preceded by a diagrammatic representation of its meaning. An example is shown in Fig. 1. The left–right orientation of the figures was varied randomly so that there was no correlation with the word order of the sentence. The participants viewed the figure for as long as they wished, and when they pressed a designated key the figure disappeared and the sentence was presented.

The instructions for the exposure and no-exposure groups were broadly similar. Both groups first read the following: ''Languages vary with respect to the word order patterns that they allow. Languages like English have a fairly consistent word order pattern, but other languages (exposure group: ''like Japlish'') have a greater variety of word order possibilities. But in all languages there are rules that determine which word orders are possible, i.e. *grammatical*, and which are not''. The no-exposure group were then told that ''we are interested in your intuitions about which word order patterns seem to you to be more likely to be grammatical in a language that you do not actually know''. The use of the case markers was then explained to them. The exposure group were told ''the grammar of Japlish allows a variety of word orders. All of the sentences you have had up until now were grammatical in Japlish. We are now interested in your intuitions about which sentences you think are likely to be grammatical, and which sentences you think are not likely to be grammatical, in Japlish''. The diagrammatic representation of sentence meaning was then introduced in the context of three practice trials.

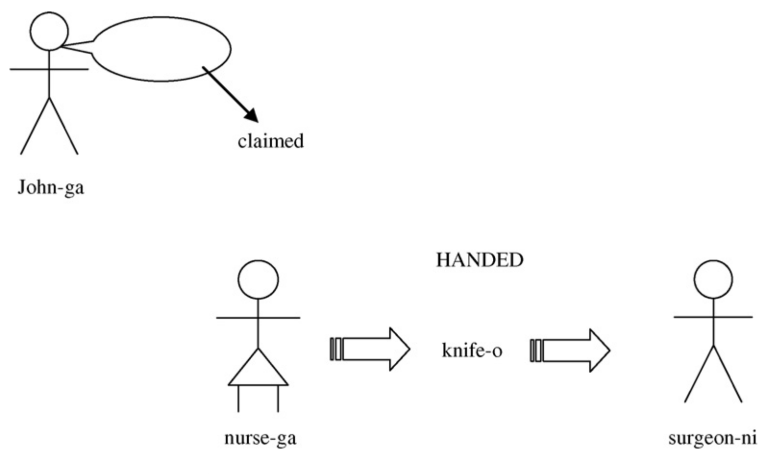


Fig. 1. The diagram used to convey the meaning of the sentence ''John-ga nurse-ga surgeon-ni knife-o handed that claimed''.

5. Results

Acceptance rates for each construction are sorted by group, and are analysed using generalised linear modelling with the statistical package SAS. The rates are also converted into mean percentage scores. For convenience trace markers and indices are not included in structure labels.

5.1. No-exposure group

As we described above, the data of the no-exposure group serve two functions. One is simply to serve as a baseline for measuring the effect of exposure, and the other is to determine the starting point, or initial state of L2 development. The latter is the purpose of this section. Our investigation will be focused on L1 transfer, which has been shown by a number of previous studies to play a critical role in the early stage of SLA.

Table 4 shows the no-exposure group's acceptance rates on short and long scrambling. The data show that they clearly accept scrambled structures as long as they involve within-clause scrambling. The acceptance rates on OS(I)V, ISOV, and S[OS(I)V]V are all significantly above chance.

Table 5 shows the results of the tests for multiple application/scrambling and superiority effects. An analysis of the mean of ISOV/AdjSOV and IOSV/AdjOSV showed that single scrambling is accepted significantly more than multiple scrambling ($\chi^2 = 15.31$, d.f. = 1, $p < 0.001$). The overall rate for single scrambling is significantly higher than chance ($\chi^2 = 24.29$, d.f. = 1, $p < 0.001$), whilst the acceptance rate on multiple scrambling is not different from

Table 4
No-exposure group acceptance rates (%) for short and long scrambling

Type of scrambling	Structure	% Acceptance	Statistical results
Short scrambling	OS(I)V	80	$\chi^2 = 38.71$, d.f. = 1, $p < 0.001$
	ISOV	75	$\chi^2 = 14.48$, d.f. = 1, $p < 0.001$
	S[OS(I)V]V	59	$\chi^2 = 4.45$, d.f. = 1, $p < 0.05$
Long scrambling	OS[S(I)V]V	35	$\chi^2 = 10.94$, d.f. = 1, $p < 0.001$
	IS[SOV]V	38	$\chi^2 = 3.50$, d.f. = 1, $p = 0.061$

Table 5
No-exposure group acceptance rates (%) for single and multiple application, superiority, and canonical Japlish and English

Single application		Multiple application	
ISOV	75	IOSV	53
AdjSOV	70	AdjOSV	44
Mean	72	Mean	48
Superiority: canonical		Superiority: reversed	
Wh-S Wh-O V	66	Wh-O Wh-S V	44
Canonical Japlish		Canonical English	
S[S(I)OV]V	71	SV(I)O	85

chance ($\chi^2 = 0.12$, d.f. = 1, $p = 0.72$). The difference in acceptance rates between Wh-S Wh-O and its reverse order is also significant ($\chi^2 = 5.66$, d.f. = 1, $p < 0.05$). The acceptance rate on the canonical order is significantly higher than the chance-level ($\chi^2 = 6.04$, d.f. = 1, $p < 0.05$), whereas the rate on the reversed order is not different from chance ($\chi^2 = 0.77$, d.f. = 1, $p = 0.3788$).

Table 5 also shows the acceptance rates on the canonical structures for Japlish and English. They reflect the head-final and head-initial settings for the head parameter, respectively. The no-exposure group has a slight preference for the head-initial structures over the head-final; however, this is not a fair comparison because the former consists of a single clause whilst the latter has a complex clausal structure. Nevertheless, one thing is clear: both types of structures are accepted at well-above chance ($\chi^2 = 22.52$, d.f. = 1, $p < 0.001$ for the head-initial structures; $\chi^2 = 20.72$, d.f. = 1, $p < 0.001$ for the head-final structures).

To summarise, (1) the no-exposure group accept scrambled structures (at least the ones that involve short scrambling), (2) they show multiple scrambling and superiority effects, and (3) no clear preference is observed for the setting of head direction. The second finding suggests that although the participants correctly accept scrambled structures, they do so by implementing operator movement, a syntactic operation available in their L1. This implies that the projection up to the CP level is available in native English speakers' L2 initial state, because operator-variable movements like topicalisation and wh-movement necessarily involve the specifier position of TopP and CP, respectively. This outcome in turn supports the Full Transfer model of SLA, and implies that despite the fact that the participants know that some leftward movement is involved in the formation of the scrambled sentences, they are not able to access the knowledge of the optional movement available in their L1. This may be reasonable, if we assume that they have not committed themselves to the [head-final] setting for the head parameter, which licenses free recursion to the direction opposite from the head. Without the free recursion, no landing site is available. Without a landing site, (leftward) scrambling is impossible; hence resorting to operator movement.

A question remains about the low acceptance rates for long scrambling. As we have seen in section 2, operator movement also allows movement across a clause boundary. We suspect that the reason for the below-chance acceptance for these structures is to do with processing difficulties. One might imagine that in a structure such as OS[SV]V (with an acceptance rate of 33%) the problem is the distance between the fronted object and the verb. However, distance cannot be the explanation because there is a high acceptance rate for OSIV (80%), even though the same number of words intervene the object and the verb in the two cases. Rather the problem would appear to lie in the fact that there is a clause boundary intervening the object and the verb. We assume that this introduces processing difficulties that are not present in complex sentences that do not involve extraction from the embedded clause.

With this characterisation of the initial state we have established the baseline against which to measure the effect of exposure.

5.2. *Exposure group*

5.2.1. *Exposure phase task*

Overall accuracy in the plausibility judgement task was 88%, indicating that the exposure group participants had good comprehension of the training sentences. An analysis of variance indicated that there was no effect of whether sentences were canonical or not, $F(1,24) = 2.05$, $p > 0.1$. However, there were effects of complexity, $F(1,24) = 18.3$, $p < 0.001$, and an interaction between canonicity and complexity, $F(1,24) = 8.55$, $p < 0.01$. The latter arose because the effect of complexity was greater for scrambled structures than canonical ones (canonical simple = 89%, canonical complex = 85%; scrambled simple = 95%, scrambled complex = 84%). Note, however, that even in the hardest condition, scrambled complex sentences, accuracy was still very high at 84%.

5.2.2. Grammaticality judgement task

In this section we focus on the comparison between the exposure and no-exposure groups assuming that any differences between them indicate the impact of exposure on the initial state. The exposure group only outperformed the no-exposure group on the trained complex canonical S[S(I)OV]V ($\chi^2 = 12.27$, d.f. = 1, $p < 0.001$), and the trained long scrambling structures, OS[S(I)V]V ($\chi^2 = 9.6$, d.f. = 1, $p < 0.01$). Somewhat surprisingly, the exposure group actually showed *lower* acceptance rates than the no-exposure group on two types of structures: trained short scrambling, OS(I)V, and new short scrambling, ISOV, with the difference actually being significant in the case of trained short scrambling ($\chi^2 = 5.53$, d.f. = 1, $p < 0.05$). Even though the exposure group had received examples of these structures in the training phase they actually accepted them at a significantly lower rate than participants who had had no prior exposure.

Closer examination of the data for the trained short scrambling structures, OS(I)V, revealed a high degree of variability between participants (S.D. = 0.2671) and a highly skewed distribution of scores with 17 out of 25 of participants accepting these structures at 7 or 8 out of 8, but 8 out of 25 participants actually accepting them at or below chance. The data for the no-exposure group were also skewed, although less so, with 3 out of 16 showing acceptance rates at or below chance.

We therefore have to acknowledge the fact that acquisition of scrambling (i.e. accessing the knowledge of the optional syntactic movement and selecting the [head-final] value for the head parameter), if it occurred at all, was limited to a subset of the exposure group participants. As we will report at the end of this section, it actually appears that some of the participants simply imposed a preference for canonical word order and responded to all other structures, including trained scrambling structures, at a similar chance level. We therefore selected participants who scored at or above 75% on the trained short scrambling structures, OS(I)V. Applying this criterion to *both* the exposure and no-exposure groups resulted in a group of 14 exposure group participants and 11 no-exposure group participants. We shall refer to these groups as the exposure/no-exposure ‘‘scramblers’’ (using quotation marks to indicate that we are using this term non-technically, and without any implication that they have acquired scrambling in the technical sense described earlier).

If the exposure group ‘‘scramblers’’ have managed to internalise the movement operation in some way or make some generalisation of the phenomenon, then one would expect them to accept scrambling constructions that they have never come across before, and that their acceptance rates on new scrambled structures would be above that of even the no-exposure group ‘‘scramblers’’. However, if these rates are not higher than the no-exposure group, then their behaviour could not be claimed to be a result of exposure, but simply a reflection of the initial state. Fig. 2 shows that there are indeed differences between the groups on new complex short scrambling ($\chi^2 = 10.16$, d.f. = 1, $p < 0.01$), and a marginally significant effect for new complex long scrambling ($\chi^2 = 3.47$, d.f. = 1, $p = 0.06$), as well as effects for trained long scrambling ($\chi^2 = 15.70$, d.f. = 1, $p < 0.001$). Although we have failed to obtain a difference for new short scrambling, in all other cases the exposure group showed higher acceptance rates than the no-exposure group.

We conclude from these data that exposure has impacted upon the initial state and increased acceptance of scrambled structures, at least for the ‘‘scramblers’’. In the case of new complex short scrambling, and to some extent new complex long scrambling, there is evidence of learning

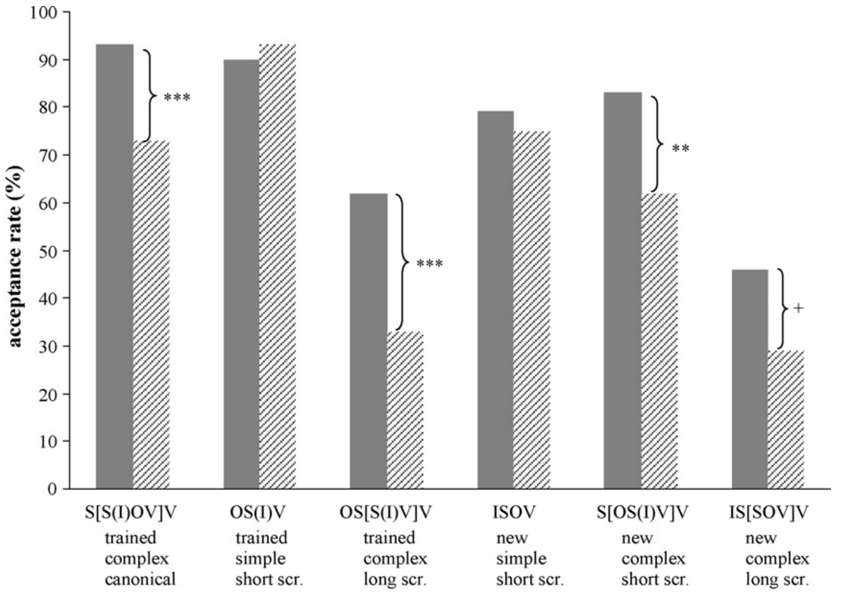


Fig. 2. Acceptance rates (%) for the exposure and no-exposure sub-groups who scored ≥75% on OS(I)V. *Note*: Dark bars, exposure group; striped bars, no-exposure group. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{+}p = 0.06$.

that generalises beyond specific patterns received in training. But is this a reflection of acquisition of scrambling? To answer this we need to see how the ‘‘scramblers’’ fare on other test structures.

Figs. 3 and 4 show the exposure and no-exposure group ‘‘scramblers’’ acceptance rates for multiple scrambling and superiority structures. As has already been reported, the effect of

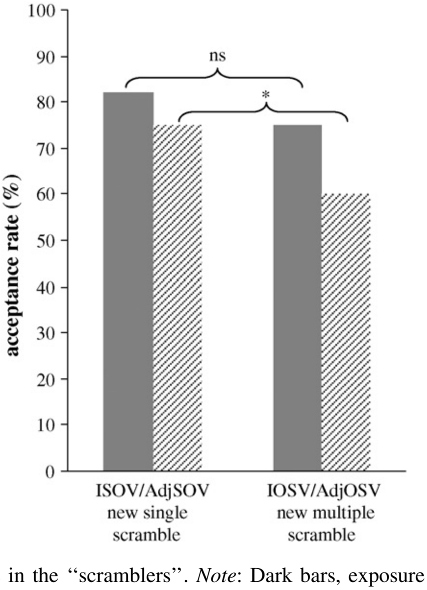


Fig. 3. Multiple application effects in the ‘‘scramblers’’. *Note*: Dark bars, exposure group; striped bars, no-exposure group. ns, not significant, $^{*}p < 0.05$.

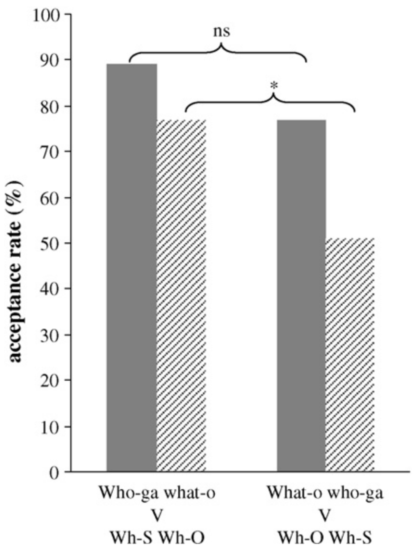


Fig. 4. Superiority effects in the scramblers. ns, not significant, $^{*}p < 0.05$. *Note*: Dark bars, exposure group; striped bars, no-exposure group. ns, not significant, $^{*}p < 0.05$.

multiple scrambling and superiority holds for the no-exposure group as a whole. Figs. 3 and 4 show that these effects are also significant for the no-exposure group ‘‘scramblers’’, for both multiple scrambling ($\chi^2 = 4.33$, d.f. = 1, $p < 0.05$) and superiority ($\chi^2 = 6.23$, d.f. = 1, $p < 0.05$). In contrast, for the exposure group ‘‘scramblers’’ there are no statistically significant effects of multiple scrambling ($\chi^2 = 1.60$, d.f. = 1, $p = 0.21$), nor for superiority ($\chi^2 = 2.98$, d.f. = 1, $p = 0.\ 08$), and their acceptance of all these structures is significantly above chance ($\chi^2 = 37.64$, d.f. = 1, $p < 0.001$ for single scrambling; $\chi^2 = 25.35$, d.f. = 1, $p < 0.001$ for multiple scrambling; $\chi^2 = 24.08$, d.f. = 1, $p < 0.001$ for Wh-S Wh-O; $\chi^2 = 14.28$, d.f. = 1, $p = 0.001$ for Wh-O Wh-S). These results are consistent with the hypothesis that this group is not analysing scrambled structures using L1 operator-variable movement.

Nevertheless, even if it is argued that for the exposure group ‘‘scramblers’’ as a whole multiple scrambling and superiority effects are absent, this does not guarantee that the acquired knowledge is scrambling. (Leftward) scrambling should correlate with the [head-final] setting of the head parameter. The setting of the head parameter into the [head-final] value should lead the exposure group ‘‘scramblers’’ to accept the trained canonical structures, S[S(I)OV]V, and more importantly, reject the English head-initial word order (*SV(I)O), which is ungrammatical in Japlish. In Fig. 2 we saw that the acceptance rate for the Japlish canonical order S[S(I)OV]V, is significantly greater in the exposure group than the no-exposure group ($\chi^2 = 12.27$, d.f. = 1, $p < 0.001$). In contrast, the two right-most structures in Fig. 5 show that the exposure group have significantly lower acceptance for English canonical structures, SV(I)O than the no-exposure group ($\chi^2 = 20.62$, d.f. = 1, $p < 0.001$). On the face of it this would seem to indicate that the exposure group ‘‘scramblers’’ have reset the head parameter, but it is important to note that the acceptance rates of the exposure ‘‘scramblers’’ on the English canonical structures are not different from the chance level ($\chi^2 = 2.53$, d.f. = 1, $p = 0.11$).

If the head parameter is truly reset to the correct value, and scrambling is available, then optional movement should be licensed to the opposite direction from the head, i.e. to the left. If this is indeed the case, then exposure group ‘‘scramblers’’ should correctly reject the structures involving

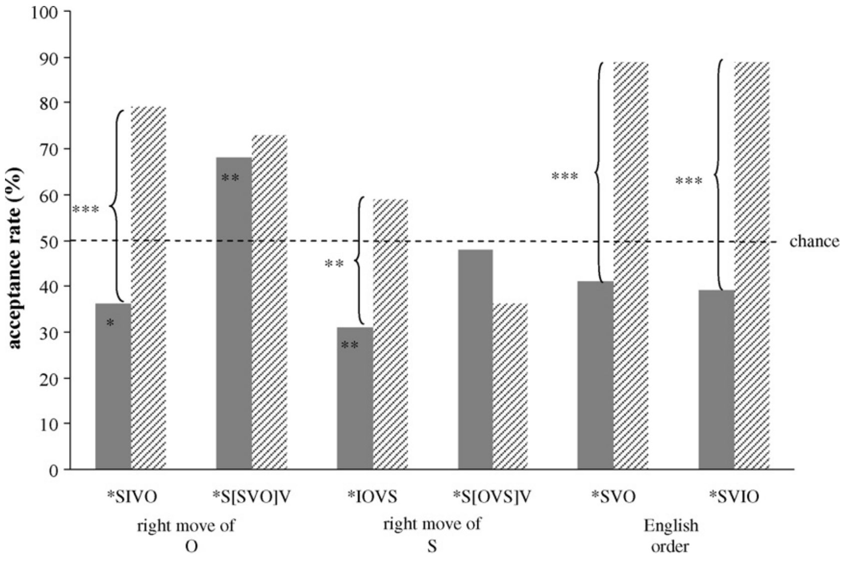


Fig. 5. Acceptance rates for ungrammatical structures involving right movement: exposure and no-exposure ''scramblers''. *Notes*: Dark bars, exposure group; striped bars, no-exposure group. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Asterisks on bars indicate the significance of the difference from chance.

movement to the right. However, this is not what our data show. Fig. 5 shows the acceptance rates on the ungrammatical structures involving right movement of the object or the subject.

Given that all of the sentences that the exposure group received in training had sentence/clause-final verbs, one would expect that it should be easy for them to reject the structures that do not end with a verb, and that their acceptance rates would be significantly lower than those of the no-exposure group. This expectation is met only by the results of the simple structures. The acceptance rates on SIVO and IOVS are 36% and 30%, respectively, which are significantly lower than the acceptance rates of the no-exposure group ($\chi^2 = 35.76$, d.f. = 1, $p < 0.001$; $\chi^2 = 8.04$, d.f. = 1, $p < 0.01$, respectively) and below the chance level ($\chi^2 = 5.92$, d.f. = 1, $p < 0.05$; $\chi^2 = 8.16$, d.f. = 1, $p < 0.01$, respectively). Nevertheless, if the exposure group ''scramblers'' had indeed accessed the knowledge of optional movement and reset the head parameter according to the given linguistic data, this should also be manifested in their behaviour on the complex structures. Our data show that this is not the case. The acceptance rate on *S[SVO]V is well above chance ($\chi^2 = 6.82$, d.f. = 1, $p < 0.01$), and is not statistically different from the no-exposure group ($\chi^2 = 0.28$, d.f. = 1, $p = 0.60$). Moreover, the exposure group's acceptance rate on *S[OVS]V is not significantly different from chance ($\chi^2 = 0.07$, d.f. = 1, $p = 0.79$), and, worse still, it is higher than that of the no-exposure group, although the difference is not statistically significant.

Finally, although the above analyses have focused on whether the ''scramblers'' acquired scrambling it should not be forgotten that many of the participants did not appear to like scrambling structures, even ones that they had been trained on. Whilst these ''non-scramblers'' accepted canonical structures at a rate of 94%, trained scrambling structures were accepted at just 41%, new scrambling structures at 45%, and ungrammatical structures at 36%. They clearly had a strong preference for canonical word orders. Non-canonical structures were responded to randomly regardless of whether they had occurred in training, or whether they were ungrammatical. This result is reminiscent of Iwasaki's (2003) finding that knowledge of Japanese scrambling is highly variable in L2 at all levels of proficiency.

5.2.3. Individual analysis

The degree of correlation between head direction and the direction of scrambling was put to a further test in which we examined the relationship between (a) acceptance of new single short scrambled structures, (b) acceptance of multiple scrambling and reversed superiority structures, and (c) rejection of ungrammatical structures resulting from the [head-initial] word order and right movement, on a participant by participant basis in the exposure group as a whole. The criteria were as follows. For (a), the mean acceptance rate had to be at least 75% over ISOV, AdjSOV, S[OS(I)]V; for (b) the mean acceptance rate had to be at least 75% over IOSV, AdjOSV, and What-O who-S V; and for (c) the mean acceptance of the [head-initial] word orders and right movement structures had to be 25% or less. There were three participants who managed to meet the three criteria. Of these three, one participant showed categorical acceptance of all short scrambling structures (hence no multiple scrambling or superiority effects), and categorical rejection of all [head-initial] and right movement structures, even the embedded ones. The remaining two showed relatively high acceptance rates for embedded right movement. The remaining 22 participants failed to show a correlation between the three conditions: six of them failed to meet any of the three criteria, whilst the rest only succeeded in one or two out of the three. This kind of data pattern would not be expected under the parameter resetting view. Therefore, although there was one participant who behaved as if she had acquired scrambling, it is possible that this was achieved by formulating an explicit rule to arrive at correct answers. In short, there was no evidence for the correlation between performance on the three types of item that would be predicted by UG.

5.2.4. Summary of the exposure-group results

To summarise the results of the exposure group: a sub-set of the participants did indeed appear to have higher acceptance of scrambling structures than the no-exposure group, even when the two groups were matched on acceptance rate for trained simple scrambling structures. Unlike the no-exposure group, their acceptance of scrambling generalised to multiple scrambling and reversed superiority structures. However, this group was still not sure whether head-initial word orders are allowed in Japlish. Furthermore, they did not show across-the-board rejection of all [head-initial] structures and right movement structures: only managing to reject reliably the single-clause right movement constructions. In general there was a lack of correlation between generalised acceptance of scrambling and rejection of the ungrammatical structures, and we take this as evidence that the exposure group's knowledge of "scrambling" is unlikely to be UG-licensed. So if the exposure group's data cannot be explained within a UG framework, can it be explained within an emergentist one?

6. An emergentist approach

Our goal in this part of the study was to see what outcome would be expected if learning were purely input-driven, with no contribution from L1 or UG. Our hypothesis is that participants base their responses on the statistical structure of the sentences they encountered in the exposure phase. Our working assumption is that connectionist networks compute input statistics in a psychologically plausible way, and so we shall examine the degree of correspondence between the participant test data and the outputs of a network trained and tested on the same sentences. We should stress at the outset that our purpose here is not to test a specific model of the human learning process. Rather we are concerned with evaluating *what* is learned, not how it is learned (this point will be elaborated in section 7).

Elman (1990) introduced the simple recurrent network (SRN) architecture as a means of modelling sequence learning (see Chang et al., 2006, for an elaboration of this work in the context of natural language syntax). Networks of this type learn sequences by being trained to predict the next element on the basis of the preceding elements. For example, for the sentence ''John loves Mary'', ''John'' is presented as a pattern of activation over the input layer. Activation spreads through a layer of hidden units to a layer of output units, and the output is compared to the correct prediction, in this case ''loves''. Since at the start of training all the weights in the network are set to random values, there will be a high degree of error in this prediction. Small adjustments to the weights are made in order to approach the weight strengths that would be necessary to make the correct prediction. ''Loves'' is then presented, and the network is taught to predict ''Mary''. Learning occurs through incremental adjustments to the weights on the connections between the units in different layers of the network over many sentences, and the goal of learning is to find a configuration of connection weights that minimises the prediction error over all training sentences. One might imagine that the SRN would just learn the strength of association between successive words in the sentences it is trained on. This would be true if it were not for the addition of ''context'' units that cumulatively record the hidden unit activation created by segments in the sequence. When ''loves'' is presented, activation from the input units is combined with the hidden unit activation that was created by ''John'' so that both of these sources of information can be used to predict ''Mary''. In this way the network can, in principle, learn to make context-dependent predictions. Elman (1993) showed that SRNs can learn long-distance dependencies. For example an SRN could learn that after the fragment ''The dog who the boys see ...'' a singular verb is predicted, even though a plural is compatible with the most recent noun in the sentence.

6.1. Simulation details

The first issue is to decide over what level of representation the SRN is to compute the statistical structure of the training sentences. The very fact that the participants were able to reliably accept canonical sentences even though they contained different words from training sentences shows that they were able to learn patterns over categories that are more abstract than specific word forms.[10] That is, they recognise that in the sentences ''Bill-ga pizza-o ate'' and ''Mary-ga book-o read'', *Bill-ga* and *Mary-ga* are in some sense equivalent, as are *pizza-o* and *book-o*, and *ate* and *read*. It seems appropriate, therefore, to code the inputs to the network using categories that reflect these equivalences.[11] For convenience we shall continue to refer to the relevant categories as Subject, Object, Indirect Object, and Verb. But we are not assuming that participants necessarily compute statistical structure (if they do so at all) over these grammatical categories. Computations could be performed within the domain of thematic roles (Agent, Theme,

[10] This is not to say that there could not have been learning at the lexical level, i.e., lexical collocations, but that we assume that this level of learning was irrelevant to the outcome of the experiment.

[11] Chang et al. (2006) also recognise that SRNs need to be trained on syntactic categories rather than word forms if they are to show generalisation to novel sentences. Although their model is presented with word forms, an initial ''compression'' network is used to force learning of the categories that are relevant to syntactic prediction in the sequencing (SRN) component. We assume that our participants are able to construct the categories that are relevant to learning word order on the basis of what they are told about Japlish case markers and their L1 knowledge of grammatical categories or thematic roles, and so representations corresponding to these categories were presented to the SRN directly.

Table 6a
The coding scheme used for the simulation

	Unit no.								
	1	2	3	4	5	6	7	8	9
Segment	Begin	Subject	Indirect object	Direct object	Verb	Wh-word	Adjunct	End	Blank
Abbreviation	Beg	S	I	O	V	Wh-	Adj	End	Blank

Table 6b
Example coding: OSIV

	Beg	S	I	O	V	Wh-	Adj	End	Blank
Beg	1	0	0	0	0	0	0	0	0
O	0	0	0	1	0	0	0	0	0
S	0	1	0	0	0	0	0	0	0
I	0	0	1	0	0	0	0	0	0
V	0	0	0	0	1	0	0	0	0
End	0	0	0	0	0	0	0	1	0
Blank	0	0	0	0	0	0	0	0	1
Blank	0	0	0	0	0	0	0	0	1

Recipient, Action) or for nouns, surface forms (-ga, -o, and -ni), or some combination of these. For the purposes of coding input to the network, however, all of these possibilities would result in the same coding scheme. The coding scheme and an example are given in Tables 6a and 6b.

Fig. 6 shows the architecture of the model. The arrows in the figure indicate how the layers of units are connected together. Each unit in one layer is connected to every unit in the other, and the connections have modifiable strengths (or weights) that determine how much activation passes along them. Learning consists of modifications of these connection weights.

The ‘Tlearn’ program was used to run the simulation (Plunkett and Elman, 1997). Each training sentence was coded as a sequence of grammatical categories ensuring that each type of structure was presented with the same frequency as in the actual training phase. Training sentences were presented to the network in a random order for a total of 50 cycles through the training set.[12] The different types of test item were coded using the same scheme and presented to the network after training was complete (learning was switched off during the testing phase). For each item type the activation over the output units in response to each segment was recorded. The simulation was repeated 10 times and the results averaged.[13]

[12] When the sentences were blocked as in the actual experiment the learning achieved in one block interfered with the learning achieved in the previous one (this is the well-known phenomenon of catastraophic interference, O’Reilly and Rudy, 2001). As a result the network vascillated wildly between adaptations to the types of item in each block. Also, the number of cycles required cannot be compared to that of humans because of the incremental nature of learning in connectionist networks. Fifty cycles represent an intermediate level of training at which the trained structures have not yet been learned perfectly. The fact that the training conditions of the model do not match those of the humans is irrelevant to our purpose because we do not intend to provide a model of the actual learning process. We are attempting to derive what might be computed, not model how it is computed. See section 7.

[13] Variation in performance is because the “initial state” of each run of the network is a unique set of random connection strengths. Time to reach a solution (to the problem of finding a set of weights that minimise prediction error over the entire set of training items) varies from run to run, as does the nature of the solution itself. Each run results in a unique set of connection weights.

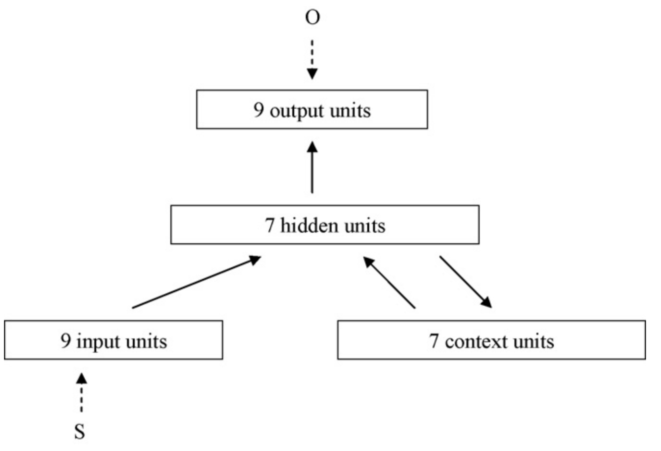


Fig. 6. The simple recurrent network used in the present simulations. The figure shows the processing of the first segment of an SOV structure.

The averaged output for one test structure is given in Table 6c. After the input 'Begin' the network's strongest predictions are for S, O, or wh. Output strengths presumably reflect the probability of each possible sentence beginning in training. When the network is given O as the next input it correctly predicts S. After S the network's strongest prediction is for S, which presumably derives from the trained OS[S(I)OV]V structure, as well as the frequent S[S pattern in complex structures. The prediction for Adj comes second (deriving from the S Adj sequence in training) followed closely by I. After I it correctly predicts V, but there is also a strong prediction for O (derived from the canonical SIOV pattern), and after V it correctly predicts the end of the sentence.

How can network output be related to grammaticality judgements? First, the strength of the activation on the target output unit (the unit corresponding to the following segment) was expressed as the proportion of the total output activation, i.e., the output of the correct unit divided by the total activation over all units; this is known as the Luce ratio (as used in a similar context by, for example, Timmermans and Cleeremans, 2001). We take this as a measure of the

Table 6c
Example output after training

Input	Target	Beg	S	I	O	V	Wh-	Adj	End	Blank	Target Luce ratio
Beg	O	0.05	0.67	0.02	0.25	0.01	0.14	0.06	0.03	0.00	0.2010
O	S	0.02	0.79	0.01	0.02	0.02	0.02	0.12	0.12	0.00	0.7060
S	I	0.02	0.27	0.18	0.05	0.03	0.06	0.19	0.05	0.01	0.2138
I	V	0.06	0.06	0.06	0.22	0.59	0.07	0.05	0.04	0.01	0.5070
V	End	0.01	0.02	0.01	0.02	0.22	0.01	0.10	0.85	0.01	0.6794

Note: Cells for target units are highlighted. The mean output strength for this item is 0.461.

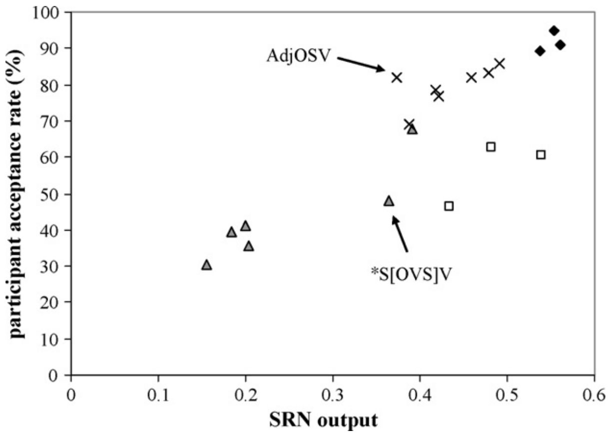


Fig. 7. Exposure group ‘‘scrambler’’ acceptance rates in relation to SRN output. (◆) Canonical, (□) long scrambling, (▲) ungrammatical, and (×) short scrambling.

extent to which the network is able to anticipate the following segment. The Luce ratios for each word were then averaged over the whole sentence to give a global measure of how well the network was able to predict the segments of the sentence.

6.2. Comparisons between participant and network performance

Fig. 7 shows network performance on each test structure in relation to the mean acceptance rates of the exposure group ‘‘scramblers’’. Each point corresponds to one test structure. The canonical structures are S[SOV]V, S[SIOV]V, Who-ga what-O V. The short scrambling structures are ISOV, IOSV, AdjSOV, AdjOSV, What-o who-ga V, S[OSV]V, and S[OSIV]V (OSV and OSIV have been removed from these figures and subsequent analyses because they are diagnostic of the ‘‘scrambler’’ and ‘‘non-scrambler’’ groupings).

Considering the ungrammatical, short scrambling, and canonical structures there is a linear relationship between participant acceptance rate and SRN output. Long scrambling structures do not pattern with the others, however. We have suggested that acceptance of long scrambling is generally suppressed because of processing difficulties associated with extraction out of a clause, and so it is not surprising that these structures are accepted at a lower rate than other structures with a similar SRN output strength. Over the remaining structures, however, there is a strong correlation between participant and network performance, Pearson $r = 0.955$. The linear regression model indicates that 91.3% of the variance in participant acceptance rates can be accounted for by the network.

There are however, two points that do not fit the regression, S[OVS]V (standardized residual = −2.57) and AdjunctOSV (standardized residual = 2.43). These are labelled in Fig. 7. In the former case, the network’s relatively strong output is because of a high prediction strength for V in the embedded clause. For the network this sentence is coded as SOVSV, and since SOV is a frequent word order in training, the network makes a strong prediction from SO to V, boosting the overall output strength for the sentence as a whole. The participants would presumably not have made the SO → V prediction because they knew from the diagrammatic representation of the sentence meaning that the first S was the matrix subject. But without the diagram it would indeed have been logical for the participants to take the SO sequence as highly predictive of V

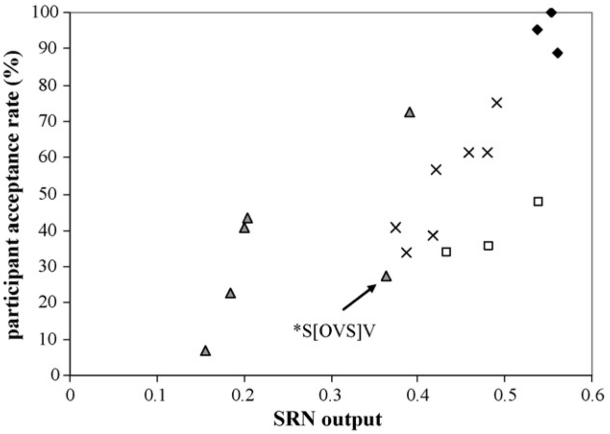


Fig. 8. Exposure group "non-scrambler" acceptance rates in relation to SRN output. (◆) Canonical, (□) long scrambling, (▲) ungrammatical, and (×) short scrambling.

(they would essentially have been "garden-pathed" by this sentence). Hence the misfit we observe for S[OSV]V is merely a consequence of the fact that the network had no way of knowing in advance whether the sentence was simple or complex.

Participant responses to AdjunctOSV are above the level predicted by the network. Here it is possible that the participants base their decisions on the OSV fragment and simply segment off the initial adjunct, which they might perceive as an optional element. Indeed, their response rate of 0.875 is close to the predicted value of 0.83 based on entering the network output strength for OSV into the regression equation. However, the network is forced to compute predictions from the initial Adjunct, thereby reducing its output for that structure.[14]

Fig. 8 shows that the "non-scrambler" acceptance rates also bear a relationship to the network's output. Excluding the long scrambling structures the linear regression model accounts for 69.7% of the variability in participant acceptance rates. Like the "scramblers" there is a linear trend over the ungrammatical and canonical structures (excluding S[OVS]V), but unlike the "scramblers" the short scrambling structures fall below this trend. Yet within the short scrambling structures there is still a correlation with network performance, $r = 0.897$.

To summarise, the network provided a good fit to the participant data. Divergences between network output and participant performance could be traced to additional processing effects (long scrambling) or to differences in the information available to the participants and the network (S[OVS]V). Note that the variability in participant acceptance rates for the ungrammatical items turns out to be consistent with the network. For example, the embedded right movement structure S[SVO]V has a particularly strong network output (0.39) and is also the one with the highest "scrambler" and "non-scrambler" acceptance rate.

7. Discussion

The purpose of the present study was to examine the nature of what is learned on initial exposure to a novel language. We first examined the nature of the "initial state" by looking at

[14] If we accept the above assumptions, and remove S[OSV]V from the regression, and give AdjunctOSV the network output strength for OSV, then network output accounts for 98.4% of the variance in participant acceptance rates.

grammaticality judgements to novel syntactic structures, and found evidence that participants utilised L1 knowledge in making their judgements. Scrambling structures were analysed as operator-variable movement, and head direction was transferred from L1. These observations support the view that ''the initial state of L2 acquisition is the final state of L1 acquisition'' (Schwartz and Sprouse, 1996, p. 40) including structure up to CP (contra Vainikka and Young-Scholten, 1994, 1996; Bhatt and Hancin-Bhatt, 2002).

We then asked, what is the impact of L2 input on the initial state? We found evidence for a significant impact even after just 30 min of exposure. It needs to be emphasised that all of the learning effects were in the context of an incidental learning situation in which the participants' only task was to make semantic plausibility judgements on Japlish sentences. One might have thought that they could accomplish this task simply by paying attention to the case markers and that, in simple sentences at least, they could effectively ignore word order. Yet, at least when the ''scramblers'' were considered, increased acceptance rates with respect to the no-exposure group were obtained for both trained and new structures (see Fig. 2). For example, OS[S(I)V]V was accepted at a significantly higher rate in the exposure group despite the fact that there were only 16 examples in the training. This shows an impressive sensitivity to word order patterns after limited exposure in a meaning-focussed task.

For present purposes the essential question is how this initial learning can be characterised. In particular, can the participant's learning be shown to be UG-guided? Our analyses of the grammaticality judgement data suggested not. On the one hand, ''scrambler'' participants, defined by their high acceptance of trained short scrambling, did appear to accept new scrambled structures at well above-chance levels, and even for the less-preferred structures (such as long distance scrambling), they out-performed the no-exposure group ''scramblers'' (Fig. 2). Further, they did not show multiple application and superiority effects, suggesting that they were not dealing with scrambled structures by applying operator-variable movement from the L1. This is rather surprising given that one might have expected the use of English lexis to increase the probability of L1 influence. Yet at the same time there was little evidence that they had reset the head direction parameter and accessed knowledge of optional movement: they failed to reject English head-initial word orders and the structures involving right movement inside an embedded clause. Thus, although there was evidence for incidental learning of word order patterns, this learning did not appear to be guided by UG.

But if learning was not guided by UG, then how is it to be characterised? We showed that there is a good correspondence between the ''scramblers'' acceptance rates and the output of a connectionist SRN that was trained on the same sentences. This result is sufficient to allow us to conclude that the participants' behaviour can be explained largely by the interaction between domain-general, frequency-sensitive, learning processes and the input received in training (we say ''largely'' because the difference between scramblers and non-scramblers, and the low acceptance of long-distance scrambling are not predicted by the SRN). But if we want to move beyond this general characterisation of what the participants had learned it will be helpful to consider more carefully the nature of learning in the network.

SRNs learn to make predictions by essentially computing contingencies between events. But contingency is not the same as frequency. Just because event B follows event A with a certain frequency does not mean that A is predictive of B. Predictability also depends upon how often A is not followed by B. Taking this into account gives the familiar measure of transition probability. But according to Shanks (1995) a more precise characterisation of human and animal learning is obtained by also taking into account the probability of B in the absence of A. This gives the statistic known as 'delta P', which is the probability of the outcome (B) given the cue (A) minus the probability of the outcome (B) in the absence of the cue (A). Ellis (2006a,b) proposes that much of SLA can be understood in terms of contingency learning. Shanks (1995, Chapter 4) also shows how a simple connectionist network essentially computes delta P when trained to asymptote (see also Perruchet and Peereman, 2004). From this perspective, connectionist models offer no more than a useful way of computing contingency statistics. Whether they are realistic models of how those computations are carried out in the brain is irrelevant.

However, the present SRN does not simply compute the contingency between one grammatical category and the next in the input sequences. The SRN's predictions were dependent upon the context in which a word occurred. For example, in complex sentences, after the first (embedded) V the network correctly predicted a second (matrix) V. In contrast in simple sentences, after the first, and only V, it correctly predicted the end of the sentence. The prediction of either a V–V or V–end sequence depended upon whether S occurred twice at the beginning of the sentence. This was true even of complex structures not encountered in training and showed that the network was sensitive to global context, not just the preceding word. The SRN was able to do this because of the inclusion of recurrent connections to context units. This sensitivity to context means that the SRN's predictions are in fact a blend of the effects of local and non-local constraints. The degree of fit between the SRN's output and participant acceptance rates implies that the participants were also computing contextually-dependent contingencies, although as noted earlier we have no way of knowing whether their computations were over forms of case markers, grammatical categories, or thematic roles.

If the participants were absorbing the statistical properties (as defined above) of the training sentences in the training phase, what were they judging in the test phase? The SRN's mean output strength for each test sentence is a function of its similarity to training sentences. Thus, it seems likely that even though the participants were instructed to judge the grammatical acceptability of the test sentences they were strongly influenced by the similarity between test and training sentences. It is not necessary to appeal to a notion of "grammaticality" in explaining their behaviour.

The present results are consistent with previous work showing rapid acquisition of the distributional structure of syllable sequences. In Saffran et al. (1996a) participants heard an unsegmented stream of syllables such as *ba-bu-pu-du-ta-ba-bu-pa-da-tu-ti-bu-pa-tu-bi-pi-da-bu* . . . . The participants were not told that the sequence was actually composed of just six trisyllabic words (*babupu*, *bupada*, *dutaba*, *patubi*, *pidabu*, *tutibu*) presented in random order. After 21 min of exposure participants showed a preference for syllable strings that corresponded to words over strings that did not. This demonstrated an ability to learn the transition probabilities between syllables, and to recognise points of low transition probability in the sequence as evidence of potential word boundaries. These effects have been shown in 8-month-old infants after only 2 min of exposure (Saffran et al., 1996b) and in Tamarin monkeys (Hauser et al., 2001). They can also be reproduced in an SRN (Elman, 1990). Clearly, humans are endowed with a natural talent for what Ellis (2006a) refers to as "tallying" – keeping track of the contingencies between events in the environment.

The crucial question in the current context, however, is what role contingency learning plays in second language acquisition of syntax. To evaluate this we need to look beyond the very initial stage of learning that we examined here. Would the same degree of fit between human and network learning still be obtained at later stages of development? Would both the participants and the SRN eventually acquire syntactic properties associated with scrambling in Japlish? We can easily examine one half of this issue by training the SRN for much longer, say for 5000 cycles through the training sentences instead of 50. The output strengths for the two simulations on different classes of test structure are shown in Fig. 9.

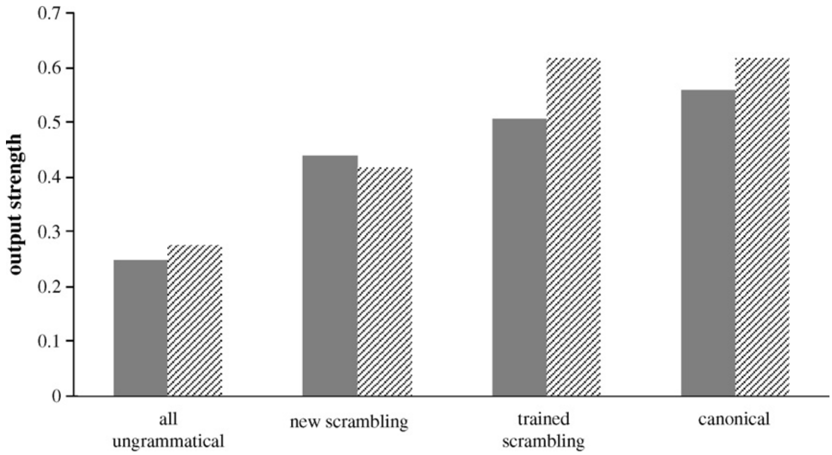


Fig. 9. SRN outputs at different levels of training. *Notes*: Dark bars, 50 cycles of training; striped bars, 5000 cycles of training.

Greater training increases the strength of the output to trained structures. Output strength for new scrambling goes down slightly, and for ungrammatical structures it actually goes up slightly. In other words, the network's ability to discriminate ungrammatical and new scrambling structures becomes if anything worse with increased training. In particular, the output for the embedded ungrammatical right movement structure S[SVO]V (0.35) is similar to grammatical scrambling structures such as ISOV (0.36) and IOSV (0.34). In short, the network would never be able to reliably discriminate grammatical from ungrammatical Japlish.

It is, of course, an empirical question whether human participants would learn Japlish from increased input. We know from studies of L2 acquisition of Japanese scrambling that a preference for canonical word orders can persist even at advanced levels of proficiency (Iwasaki, 2003). But we can safely assume that the head-final characteristic of Japanese is learned even at low levels of proficiency. Yet an SRN's output to non-head-final structures actually increases with exposure.

What the SRN lacks is the ability to move beyond a global assessment of the familiarity of a string to an output that is determined by certain criterial features, like head position. If it were given explicit feedback concerning grammaticality, then it would presumably learn that verb position is diagnostic of grammaticality. But without such feedback it cannot. Neither would it be sufficient to make its judgements sensitive to zero predictability of a segment, since this is a feature of new grammatical scrambling structures as well. The network has no way of knowing which zero probability events are diagnostic of grammaticality and which are not.

The problem is that the SRN does not represent the regularities that it discovers in symbolic form. As Marcus (2000) has pointed out, it is simply not possible to model some simple aspects of language and cognition without including symbolic representations. For example it is not possible to explain the ability to recognise that ABA is more similar to XYX than XXY without invoking an abstract representation of alternation as opposed to doubling. In the present case we would argue that symbolically represented rules would be needed to produce categorical rejection of non-verb-final clauses, or to allow all other constituents to appear in any surface position regardless of the familiarity of the sentence as a whole. Such rules could be formed by domain-general learning procedures, or they could be a reflection of the setting of parametric options provided by UG. Connectionist networks might be very good at capturing the statistical structure of the input, and by responding to new inputs by analogy with old, but they cannot produce the kind of categorical behaviour that we assume is possible in this domain.

Williams (2003) presents another example of the difficulties of learning natural language systems in both adult implicit learning and connectionist models. The learning target was an abstract grammatical gender system in which the set of articles that accompanied a noun depended upon its grammatical gender. The gender system was purely arbitrary, defined only by the distribution of the articles. Participants who performed a meaning-focused task showed good memory for trained items, but no generalisation to novel article-noun combinations. A connectionist model that was trained on the same items in an analogous task showed similar results. As here, we see a failure to learn underlying linguistic generalisations under incidental/implicit training conditions in both humans and networks.

This is not to say that sensitivity to statistical structure is not important. For example, Ellis (2002) documents the pervasiveness of frequency effects in language learning and processing. And given the facility with which people can pick up the statistical structure of the input it seems unlikely that this process is completely dissociated from symbolic, and possibly explicit, learning processes. The task then is to specify how statistical learning can feed into other forms of symbolic learning mechanism (as in hybrid connectionist models for example, Sun et al., 2001), and to specify if, and why, conscious awareness is critical in making this transition.

But, in the case of language learning, all of these arguments are specific to the SLA context. It is striking that learning gender, as an underlyingly abstract grammatical system, appears to be unproblematic in first language acquisition (Caselli et al., 1993), and not nearly as sensitive to correlated cues (in form and meaning) as in SLA (Brooks et al., 1993; Holmes and Dejean de la Batie, 1999). Likewise, learning head position, and then rapidly acquiring scrambling, appear to be unproblematic in FLA of Japanese (Murasugi and Kawamura, 2005) but problematic in SLA (Iwasaki, 2003). These are the typical observations of the ‘‘fundamental difference’’ between FLA and SLA that lead many researchers in the field to reject a role for UG in SLA (Bley-Vroman, 1989, 1990). In as much as there is no theoretical reason why UG effects should not be detectable at the earliest stage of acquisition given that clear and sufficient input is provided to learners, our data lend support to the fundamental difference hypothesis. We would suggest that SLA is best characterised as a combination of statistical learning and additional domain-general symbolic rule-learning processes.

Of course, any conclusions from the present study must be considered in the context of the conditions under which the results were obtained; that is, from a specific task (plausibility judgement on isolated sentences), with limited exposure to a semi-artificial language. This raises the issue of generalisabilty. We have shown that under such conditions, participants are sensitive primarily to the statistical structure of the input. If it is possible to demonstrate this effect in such a small-scale study then it seems unlikely that the statistical learning mechanism does not make a contribution to SLA in more naturalistic situations. But of course the failure to obtain effects of other learning mechanisms, be they UG-related or not, by no means rules these out as contributory factors in naturalistic SLA. The question then becomes under what conditions effects of these additional mechanisms can be observed. Only through a combination of experimentation in naturalistic and artificial contexts can this question be addressed.

References

Bhatt, R.M., Hancin-Bhatt, B., 2002. Structural minimality. CP and the initial state in second language acquisition. Second Language Research 18, 348–392.

Bley-Vroman, R., 1989. What is the logical problem of foreign language learning? In: Gass, S.M., Schachter, J. (Eds.), Linguistic Perspectives on Second Language Acquisition. Cambridge university press, Cambridge, pp. 41–68.
Bley-Vroman, R., 1990. The logical problem of foreign language learning. Linguistic Analysis 20 (1–2), 3–49.
Bloom, L., 1970. Language Development: Form and Functiona in Emerging Grammars. MIT Press, Cambridge, MA.
Brooks, P.J., Braine, M.D.S., Catalano, L., Brody, R., 1993. Acquisition of gender-like noun classes in an artificial language: the contribution of phonological markers to learning. Journal of Memory and Language 32, 76–95.
Brown, R., 1973. A First Language: The Early Stages. Harvard University Press, Cambridge, MA.
Caselli, M.C., Leonard, L.B., Volterra, V., Campagnoli, M.G., 1993. Toward mastery of Italian morphology: a cross-sectional study. Journal of Child Language 20, 377–393.
Chang, F., Dell, G.S., Bock, K., 2006. Becoming syntactic. Psychological Review 113, 234–272.
Chomsky, N., 1994. Bare Phrase Structure. MIT Occasional Papers in Linguistics (MITWPL), vol. 5. MIT Press, MIT, Cambridge, MA.
Chomsky, N., 1995. The Minimalist Program. MIT Press, Cambridge, MA.
Chomsky, N., Lasnik, H., 1993. The theory of principles and parameters. In: Jacobs, J., Von Stechow, A., Sternefeld, W., Vennemann, T. (Eds.), Syntax: An International Handbook of Contemporary Research. Walter de Gruyter, Berlin, pp. 506–569.
Chung, G.N., 1994. Case and its Acquisition in Korean. Unpublished Ph.D. Dissertation, The University of Texas at Austin, Austin, Texas.
Clahsen, H., Muysken, P., 1989. The UG paradox in L2 acquisition. Second Language Research 5 (1), 1–29.
De Haan, G.J., 1987. A theory-bound approach to the acquisition of verb placement in Dutch. In: De Haan, G.J., Zonneveld, W. (Eds.), Formal Parameters of Generative Grammar, vol. III. Foris, Dordrecht.
Ellis, N.C., 2002. Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. Studies in Second Language Acquisition 24, 143–188.
Ellis, N.C., 2006a. Language acquisition as rational contingency learning. Applied Linguistics 27, 1–24.
Ellis, N.C., 2006b. Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. Applied Linguistics 27, 164–194.
Elman, J.L., 1990. Finding structure in time. Cognitive Science 14, 179–211.
Elman, J.L., 1993. Learning and development in neural networks: the importance of starting small. Cognition 48, 71–99.
Fukui, N., 1995. Theory of Projection in Syntax, vol. 4. CSLI Publications, Center for the Study of Language and Information, Stanford University/Kurosio Publishers, Stanford, California/Tokyo.
Gregg, K.R., 2003. The state of emergentism in second language acquisition. Second Language Research 19, 95–128.
Grewendorf, G., Sabel, J., 1999. Scrambling in German and Japanese: adjunction versus multiple specifiers. Natural Language & Linguistic Theory 17 (11), 1–65.
Hauser, M.D., Newport, E.L., Aslin, R.N., 2001. Segmentation of the speech stream in a nonhuman primate: statistical learning in cotton-top tamarins. Cognition 78, 53–64.
Hawkins, R., Hattori, H., 2006. Interpretation of English multiple *wh*-questions by Japanese speakers: a missing uninterpretable feature account. Second Language Research 22, 269–301.
Holmes, V.M., Dejean de la Batie, B., 1999. Assignment of grammatical gender by native speakers and foreign language learners. Applied Psycholinguistics 20, 479–506.
Ishihara, S., 2001. Stress, focus, and scrambling in Japanese. MIT Working Papers in Linguistics 39, 151–185.
Iwasaki, N., 2003. L2 acquisition of Japanese: knowledge and use of case particles in SOV and OSV sentences. In: Karimi, S. (Ed.), Word Order and Scrambling. Blackwell, Oxford, pp. 273–300.
Kawamura, T., 2001. A feature-checking analysis of Japanese scrambling. Journal of Linguistics 40 (1), 45–68.
Kuribara, C., 2000. Resettability of UG Parameters in SLA: Acquisition of Functional Categories by Adult Japanese Learners of English. The University of Reading, Reading, U.K.
Kuribara, C., 2004. Structural similarity as a determinant for "misanalysis" in SLA: how Japanese learners analyse the left-periphery of English sentences. In: Otu, Y. (Ed.), The Proceedings of the Fifth Tokyo Conference on Psycholinguistcs (TCP 2004). Hituzi Shobo, Keio University, Tokyo, pp. 229–253.
Kuroda, S.-Y., 1988. Whether we agree or not: a comparative syntax of English and Japanese. Paper Presented at the Papers from the 2nd International Workshop on Japanese Syntax, Stanford, California.
Lebeaux, D., 1989. Language Acquisition and the Form of the Grammar. Unpublished Ph.D. Dissertation, University of Massachusetts at Amherst, Amherst.
Lehmann, W.P., 1973. A structural principle of language and its implications. Language 49, 47–66.
Marcus, G.F., 2000. Pabiku and Ga Ti Ga: two mechanisms infants use to learn about the world. Psychological Science 145–147.

Meisel, J., 1995. Parameters and acquisition. In: Fletcher, P., MacWhinney, B. (Eds.), The Handbook of Child Language. Blackwell, Oxford, pp. 10–35.
Miyagawa, S., 1997. Against optional scrambling. Linguistic Inquiry 28, 1–25.
Miyagawa, S., 2001. EPP, scrambling and wh-in-situ. In: Kenstowicz, M. (Ed.), Ken Hale: A Life in Language. MIT Press, Cambridge, MA, pp. 293–338.
Miyagawa, S., 2003. A-movement scrambling and options without optionality. In: Simin, K. (Ed.), Word Order and Scrambling. Blackwell, Oxford, pp. 177–201.
Miyamoto, Y., Iijima, Y., 2003. On the existence of scrambling in the grammar of Japanese elementary EFL learners. In: Foster-Cohen, S., Pekarek Doehler, S. (Eds.), EuroSLA Yearbook: The Proceedings of the Annual Conference of the European Second Language Association (Eurosla 2003), vol. 3. John Benjamins, Amsterdam/Philadelphia, pp. 7–27.
Muller, G., Sternefeld, W., 1993. Improper movement and unambiguous binding. Linguistic Inquiry 24 (3), 461–507.
Murasugi, K., Kawamura, T., 2005. On the acquisition of scrambling in Japanese. In: Sabel, J., Saito, M. (Eds.), The Free Word Order Phenomenon: Its Syntactic Sources and Diversity, vol. 69. Mouton de Gruyter, Berlin/New York, pp. 221–242.
Neeleman, A., Weerman, F., 1997. L1 and L2 word order acquisition. Language Acquisition 6 (2), 125–170.
Nemoto, N., 1999. Scrambling. In: Tsujimura, N. (Ed.), The Handbook of Japanese Linguistics. Blackwell, Oxford, pp. 121–153.
O'Reilly, R.C., Rudy, J.W., 2001. Conjunctive representations in learning and memory: principles of cortical and hippocampal function. Psychological Review 108, 311–345.
Perruchet, P., Peereman, R., 2004. The exploitation of distributional information in syllable processing. Journal of Neurolinguistics 17, 97–119.
Plunkett, K., Elman, J.L., 1997. Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations. MIT Press, Cambridge, Massachusetts.
Poeppel, D., Wexler, K., 1993. The full competence hypothesis of clause structure in early German. Language 69, 1–33.
Saffran, J.R., Newport, E.L., Aslin, R.N., 1996a. Word segmentation: the role of distributional cues. Journal of Memory and Language 35, 606–621.
Saffran, J.R., Newport, E.L., Aslin, R.N., 1996b. Statistical learning in 8-month-old infants. Science 274, 1926–1928.
Saito, M., 1985. Some Aymmetries in Japanese and Thier Theoretical Implications. Unpublished Ph.D. Dissertation, MIT, Boston, MA.
Saito, M., 2002. On the role of selection in the application of Merge. Paper Presented at The North East Linguistic Society (NELS), MIT, Cambridge, MA.
Saito, M., 2004. A derivational approach to the interpretation of scrambling chains. Lingua 113, 481–518.
Saito, M., Fukui, N., 1998. Order in phrase structure and movement. Linguistic Inquiry 29 (3), 439–474.
Schwartz, B.D., Sprouse, R.A., 1994. Word order and nominative case in non-native language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. In: Hoekstra, T., Schwartz, B.D. (Eds.), Language Acquisition Studies in Generative Grammar: Papers in Honor of Kenneth Wexler from the 1991 Glow Workshop. John Benjamins Publishing Company, Amsterdam, pp. 317–368.
Schwartz, B.D., Sprouse, R.A., 1996. L2 cognitive states and the Full Transfer/Full Access Model. Second Language Research 12 (1), 40–72.
Shanks, D.R., 1995. The Psychology of Associative Learning. Cambridge University Press, Cambridge.
Sugisaki, K., 2005. Early acquisition of basic word order: new evidence from Japanese. In: Brugos, A., Clark-Cotton, M.R., Ha, S. (Eds.), Proceedings of the 29th annual Boston University Conference on Language Development (BUCLD 29). Boston University, vol. 29, pp. 582–591.
Sun, R., Merrill, E., Peterson, T., 2001. From implicit skills to explicit knowledge: a bottom-up model of skill learning. Cognitive Science 25, 203–244.
Tada, H., 1993. A/A-Bar Partition in Derivation. Unpublished Ph.D. Dissertation, MIT, Cambridge, MA.
Timmermans, B., Cleeremans, A., 2001. Rules versus statistics in implicit learning of biconditional grammars. In: French, R.M., Sougné, J.P. (Eds.), Connectionist Models of Learning, Development and Evolution. Springer, London, pp. 185–196.
Tsimpli, I.-M., Roussou, A., 1991. Parameter-resetting in L2. University College London, London, vol. 3, pp. 149–169.
Tsimpli, I.-M., Smith, N., 1991. Second-language learning: evidence from a poliglot savant, UCL Working Papers in Linguistics. University College London, London, vol. 3, pp. 171–183.
Vainikka, A., Young-Scholten, M., 1994. Direct access to X'-theory: evidence from Korean and Turkish adult learning German. In: Schwartz, T.H.B.D. (Ed.), Language Acquisition Studies in Generative Grammar: Papers in Honer of Kenneth Wexler from the 1991 GLOW Workshops. John Benjamins, Amsterdam, pp. 265–316.

Vainikka, A., Young-Scholten, M., 1996. Gradual development of L2 phrase structure. Second Language Research 12, 7–39.

Wexler, K., 1998. Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. Lingua 106, 23–79.

Williams, J.N., 2003. Inducing abstract linguistic representations: human and connectionist learning of noun classes. In: van Hout, R., Hulk, A., Kuiken, F., Towell, R. (Eds.), The Lexicon-Syntax Interface in Second Language Acquisition. John Benjamins, Amsterdam, pp. 151–174.