Native-like attainment of dummy subjects in Dutch and the role of the L1

SONJA VAN BOXTEL, THEO BONGAERTS AND PETER-ARNO COPPEN

Abstract

In this study, we test the prediction, derived from the Critical Period Hypothesis, that a native level in L2 grammar cannot be attained by learners who start acquiring a second language after childhood. We selected 43 very advanced late learners of Dutch (native speakers of German, French and Turkish) and compared their performance on a grammar test with that of 44 native speakers of Dutch. The test consisted of an elicited imitation task and a sentence preference task. In these tasks, participants were tested on their knowledge of dummy subject constructions. These construction types are known to be very hard to acquire for L2-learners of Dutch and are hardly covered in Dutch grammars or L2 Dutch textbooks. The results show that it is possible to attain a native level of proficiency for learners who start acquiring an L2 after puberty, even for learners with a typologically distant L1.

1. Introduction

In biology, there are many instances of learning for which there is a critical period, i.e., "a time during the life span of an organism in which the organism may be affected by some exogenous influence to an extent beyond that observed at other times" (Colombo 1982: 261). During this period, typically early in life, there is a (heightened) sensitivity to stimuli that are necessary for the development of the ability concerned. After this period, there is a non-linear decline in sensitivity. If the relevant stimuli are not present during this critical period, the ability concerned will no longer (fully) develop under normal circumstances.

Although there are differences between different types of critical periods in many aspects (see, e.g., Bornstein 1989), they share certain geometric features, which distinguish them from other types of development. In general, critical periods have an onset, a peak of heightened sensitivity, an offset and a terminus with a flattening after the terminus, as in Figure 1.

IRAL 43 (2005), 355–380 0019042X/2005/043-0355

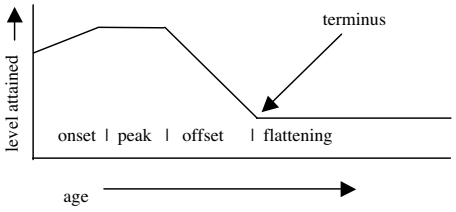

Figure 1. *Geometric features of a critical period*

For language acquisition, a Critical Period Hypothesis (henceforth CPH), has first been proposed by Penfield and Roberts (1959) and Lenneberg (1967). According to this hypothesis, learners' sensitivity to language input should decline during the offset (until the onset of puberty), due to physiological changes in the brain that are specific to language, such as loss of plasticity of the brain due to maturation. Around the age of twelve, the critical period should end. After this age there should no longer be a significant negative correlation between age of onset of language learning and the level of proficiency attained. Moreover, for learners who start acquisition after this age the level of proficiency attained should always be low and significantly different from levels attained by early learners. In the study presented in this paper, which is based on the results of Van Boxtel (2005), we will test this latter assumption by comparing the level of attainment of very proficient late learners with that of native speakers.

Researchers such as Newport (1990, 1991), Elman (1993) and Pitts Cochran, MacDonald and Parault (1999), relate the advantage of children over adults in language acquisition to differences in working memory capacity. They claim that children have an advantage over adults because, due to (initial) working memory limitations, they process smaller units than adults. This would enable children to focus more on details, like the form of morphemes, whereas adults tend not to notice these details when trying to process large units at once. This increase in working memory capacity is also a biologically determined age factor, but it is not specific to language acquisition and can be compensated for by other means, as is shown for example by Pitts Cochran et al. (1999). This means that starting with a greater working memory capacity does not necessarily have to lead to incomplete acquisition. It might, therefore, still be possible for some late learners to reach a native level of attainment in the target language.

Results and conclusions from research on the question of whether or not it is possible for late learners to acquire a native-like level of competence in a second language vary enormously. For pronunciation it has recently been shown that, at least for certain late learners, it is possible to achieve a native-like

level of attainment in a foreign or second language (see Bongaerts 1999, Moyer 1999, Bongaerts, Mennen and Van der Slik 2000, Birdsong 2003). It should be noted, however, that the learners in these studies who were indistinguishable from native speakers all spoke a first language that was typologically rather closely related to the target language.

For morphosyntax, contradictory results have been found. In several studies with very proficient late learners (e.g., Birdsong 1992, White and Genesee 1996, and Hyltenstam and Abrahamsson 2003), (quite) some participants fell within the range of native speakers for syntax. In Hyltenstam (1992), on the other hand, the scores of all very advanced late learners differed from those of native speakers, in spite of the fact that all learners started acquiring their second language in childhood. What makes the picture even more complex is that in some studies older learners seemed to have attained native-like proficiency whereas younger learners did not. McDonald (2000), for example, found that early Spanish learners (AoA ≥ 5) as well as some late Spanish learners of English (AoA > 14) performed like native speakers, whereas early Vietnamese learners of English (AoA ≤ 5) did not. They had problems with those aspects of English that differ markedly from Vietnamese, and child learners (AoA 6–10) had problems with many aspects of the test, like (most of) the late Spanish learners.

In many of these studies there were methodological problems with the constructions used and/or the selection of participants and native speaker controls (for an overview see Van Boxtel 2005). Moreover, the role of the typological distance between the source and target language(s) has not been systematically investigated in most studies. In the present study we have tried to overcome most of these problems by testing highly proficient late L2 learners with L1 backgrounds that differ in their typological distance to the target language on constructions that are known to be very hard to acquire and for which second language learners have no access to explicitly formulated rules. Learners can therefore only acquire these constructions on the basis of evidence from the input.

2. The present study

In the present study, we test the following research questions, relating to the CPH for syntax in SLA:

(1) Are there any late second language learners who fall within the native speaker range in their command of grammatical constructions that are known to be very difficult for second language learners and which can only be acquired on the basis of the input?
(2) How is the level attained in L2 grammar after the age of twelve related to the typological distance between the L1 and the L2?

(3) What are the input and background characteristics of late learners who perform within the native speaker range (if they exist)?

2.1. *Participants*

We selected 43 L2 participants who arrived in the Netherlands when they were at least twelve years old and who were very proficient in Dutch. These participants were selected through a network procedure. For each L2 participant, at least one native speaker of Dutch (usually an expert in the field of second language teaching or second language acquisition) or another L2 participant indicated that he or she spoke Dutch so well that he/she could pass for a native speaker most of the time, in all aspects of language, except perhaps for pronunciation or gender errors.

In order to systematically vary the typological distance to Dutch we tested three L1 groups that differ in their typological distance to Dutch: German, a Germanic language, which is very closely related to Dutch, French, a Romance language which is related to Dutch, but not as closely as German, and Turkish, an Altaic language, which is typologically very distant from Dutch.

The L2 group consisted of 15 native speakers of German, 15 native speakers of French and 13 native speakers of Turkish. They had arrived in the Netherlands between the ages of 12 and 35 (Mean = 22) and had lived in the Netherlands for four to fifty years (Mean = 22). The L2 participants were only selected on the basis of their proficiency in Dutch and their age of arrival. Background information was obtained through a questionnaire. Most participants were very highly educated and had a linguistic background (e.g., a degree in their native language). The Dutch control group consisted of 44 mostly highly educated native speakers of Dutch, mainly undergraduate students at Radboud University Nijmegen.

2.2. *Tasks*

We gave our participants two tasks: an oral sentence imitation task and a sentence preference task. We decided to use a sentence imitation task in addition to a sentence preference task, because an oral sentence imitation task is assumed to tap participants' implicit knowledge of grammar in a more direct way, without participants being aware of what they are doing exactly.

It has been shown that under high working memory demands, processing of sentences will be slower and less accurate (see, e.g., Just and Carpenter 1992). If a sentence is too long or complex for participants' linguistic processing and storage capacity to be retrieved in its entirety from working memory, reconstruction will take place. This may result in a "repetition" that is different from the original stimulus. If a stimulus sentence contains a grammatical feature that a participant has not acquired, this feature will unconsciously be changed

in such a way that the sentence fits the participants' own grammar (see, e.g., Bley-Vroman and Chaudron 1994). For advanced adult learners, targets (i.e., the grammatical features studied) should preferably be phonologically non-salient and occur in a non-salient position. In our test, we used dummy subjects in non-initial position, as in (1) (for a description and more examples see Section 2.3).

(1) *Meestal valt het niet mee om kaartjes voor een concert te krijgen.*
usually falls it not with to tickets for a concert get
'Usually it is not easy to get tickets for a concert.'

Because dummy subject constructions meet the criteria mentioned above, they are very suitable for an imitation task with very advanced adult learners.

Note that there are individual differences in working memory capacity (see, e.g., Just and Carpenter 1992). Because of this, people with greater working memory resources may be able to repeat sentences literally, while others with the same grammar but more limited resources may make changes in the same sentences. For this reason, only sentences that are not repeated literally can be used for analysis.

All participants did the elicited imitation task first. This task consisted of twenty sentences, which were recorded by a young female native speaker of Dutch on a DAT recorder and played to the participants. Participants were instructed to repeat the sentences literally as soon as they had heard the whole sentence. Their reactions were recorded either on cassette tape or on minidisk and analysed later. The imitation task consisted of twelve sentences containing a dummy subject (see Section 2.3) and eight "filler" items. With the exception of some of the fillers, the sentences were based on (and in most cases identical to) items in a pilot study, which we also used for the sentence preference task. The results from the native speakers on the sentence preference task showed that of the twelve dummy subject sentences six contained the dummy subject that most native speakers prefer in this context and six another dummy subject than the one preferred by most native speakers of Dutch.

In a sentence preference task (henceforth SPT), participants do not have to decide whether sentences are grammatical or not, but which sentence from a minimal pair they prefer (see Birdsong (1997), who also used this task in a study with very advanced late second language learners). The sentence preference task (henceforth SPT) in our study was presented on a computer screen and consisted of 190 items: 124 test items and 66 filler items. Each test item contained a minimal sentence pair, in which only the dummy subject (in this case *het* and *er*) was different, as in (2).

(2) a. *Het is nu gebleken dat de brand door een*
it is now turned out that the fire by an
ongeluk is ontstaan.
accident is arisen

b. *Er is nu gebleken dat de brand door een*
there is now turned out that the fire by an
ongeluk is ontstaan.
accident is arisen
'It has now turned out that the fire was caused by an accident.'

For each item in the SPT participants clicked a button on the following seven-point scale:

I strongly prefer sentence A to sentence B	○	○	○	○	○	○	○ I strongly prefer sentence B to sentence A

Participants were told that their reaction times would be measured, and that they should read the sentences carefully and then decide which sentence they preferred, without thinking too long about an item. They were also instructed to use the button in the middle when they found both sentences either equally good or equally bad. It was emphasised in the instructions that there were no good or wrong answers, and that participants should indicate which sentence they preferred, rather than what they thought was the norm in standard Dutch. Each item was presented on a separate screen (until participants clicked the 'proceed' button) and participants could not go back to previous items. Both the order of sentences within items and the order of items were randomised (the latter for each participant separately). After 85 items, participants were asked to take a five-minute break. The whole sentence preference test took between 25 and 78 minutes (Mean = 40 minutes), excluding the break. The participants' reaction times were too unreliable to use in any further analyses.

Although it has been suggested that for some items, participants could perhaps base their choice of a dummy subject on item learning of idiomatic expressions, considering the individual differences among native speakers on many items and the syntactic flexibility of dummy subject constructions in our tests, item learning alone could never result in native-like behaviour on these tests.

2.3. *Dummy subject constructions*

For our grammar tests we chose dummy subject constructions, for two reasons. First, they are known to be very hard to acquire for learners of Dutch as a second language. Second, even learners who have extensive formal training in Dutch do not have access to explicitly formulated rules for dummy subjects and can, therefore, only acquire these constructions on the basis of evidence from the input.

We use the term "dummy subject constructions" to denote constructions with a logical subject (noun phrase or embedded clause) that does not occur in the normal syntactic subject position for semantic or pragmatic reasons (e.g., heaviness). Instead, the syntactic subject position is filled by a dummy element, which can be *het* (it), or *er* (there). The position can also be empty, in which case we will say that there is an empty dummy subject *0*. There are three types of dummy subject constructions in Dutch:

(a) sentences with *er* or *0* in which the logical subject is a noun phrase (DP-type);
(b) active sentences with *er*, *het* or *0* and a sentential logical subject (AS-type);
(c) passive sentences with *er*, *het* or *0* and a sentential logical subject (PS-type).

We subdivided these three types into seven categories, based on the results of the native speakers of Dutch. Examples of each category are presented in (3)–(9) below. In these examples, the dummy subjects are represented by the symbol ◇ and logical subjects are underlined. For each category, the judgement pattern of the native speaker control group in our study (see Section 3.2) are presented to the right. In these preference judgements, a preference for 0 to *er*, for example, is indicated by the symbol $>$. When sentences with *0*, for example, are judged as either better or equally good/bad as equivalent sentences with *er*, this is indicated by the symbols $\geq$. Because *0* can never occur in initial position in main clauses (since Dutch is a verb second language), we mainly tested dummy subjects in non-initial position in our study.

(3) DP-type, DP subject general (DPg) $0 > \text{er}$

Men beseft niet altijd dat ◇ *een pinguïn een vogel*
one realises not always that ◇ a penguin a bird
is.
is
'One does not always realise that a penguin is a bird.'

(4) DP-type, non-specific DP subject + transitive predicate (DPnst) $0 \geq \text{er}$

Op televisie doen ◇ *veel mensen dingen die ze*
on television do ◇ many people things that they
normaal niet durven.
normally not dare
On television many people do things they normally would not dare to do.'

(5) DP-type, non-specific DP subject + intransitive predicate (DPnsi) er $\geq$ 0

Ik vind het vervelend dat ◇ boven een raam open staat.
I find it annoying that ◇ upstairs a window open stands
'It bothers me that there is a window open upstairs.'

(6) Active-sentential type, general (ASg) het $>$ 0 $\geq$ er

Meestal valt ◇ niet mee om kaartjes voor een concert te krijgen.
usually falls ◇ not with to tickets for a concert get
'Usually it is not easy to get tickets for a concert.'

(7) Active-sentential type, change of state (AScos) 0 / het $>$ er

Nu schiet ◇ mij ineens te binnen dat ik nog boodschappen moet doen.
now occurs ◇ me suddenly that I still shopping must do.
'Now it suddenly occurs to me that I still have to go out shopping.'

(8) Passive-sentential type, general (PSg) 0 $\geq$ er $\geq$ het

In de krant wordt ◇ beweerd dat hij dronken achter het stuur gezeten heeft.
in the newspaper is ◇ claimed that he drunk behind the wheel sat has
'It is claimed in the newspaper that he was drunk while he was driving.'

(9) Passive-sentential type, dummy object in active equivalent (PSdo) het $\geq$ 0 $\geq$ er

Door haar vrienden wordt ◇ bewonderd dat ze ook in moeilijke tijden vrolijk blijft.
by her friends is ◇ admired that she also in difficult times cheerful remains
'Her friends admire her for always remaining cheerful, even in difficult times.'

The results from the native speakers on the sentence preference task revealed that there is a considerable amount of variation on certain dummy subject constructions in Dutch, both between and within participants. There were also many predicates and categories, however, for which the judgements of the native speakers were fairly consistent. For each construction type, one or two

Table 1. *Results of the imitation task*

	L1G	L1F	L1T
No change or native-like change in all dummy subjects	8	3	1
Non-native-like change in at least one dummy subject	7	12	12

factors disturb the general pattern. Judgements for the DP-type deviate from the general pattern when the subject is non-specific and the predicate is intransitive. Those for the active sentential type (AS-type) deviate from the general pattern when the predicate expresses a change of state. Judgements for the passive sentential type (PS-type), finally, deviate from the general pattern when the equivalent active sentence contains a dummy object.

3. Analysis and results

Three native speakers of Dutch were excluded from the control group. One was excluded because on the background questionnaire he indicated that he had lived in Norway for several years before the age of twelve. The other two were excluded because they had very deviant scores in a preliminary analysis of the sentence preference task. The reason we excluded these outliers was that we wanted our criterion for native-likeness to be very strict. We also excluded a German and a Turkish participant who appeared to have lived in a Dutch environment before the age of twelve (ID61 and ID85) and one French and one Turkish participant (ID67 and ID81), because it turned out that they did not meet our proficiency criterion. The German participants were given codes ID51 to ID65. The French participants were given codes ID66 to ID80. We replaced participant ID61 and ID67 with two new participants, who were given the codes ID111 (L1G) and ID117 (L1F).

3.1. *Imitation task*

For the imitation task, we looked at differences in responses with respect to the dummy subjects between individual second language learners and native speakers. If all the dummy subjects produced by a second language learner were also produced by at least one native speaker, we considered this participant to fall within the native speaker range for the imitation task. All other participants were considered to fall outside the native speaker range for this task. The results are shown in Table 1.

Table 1 shows that there are twelve participants (eight native speakers of German, three native speakers of French and one native speaker of Turkish) who fall within the native speaker range on the dummy subject items in the imitation task (ID52-56, 59, 62, 64, 68, 70, 77 and 83).

Table 2. *Responses of sentence imitation task, all groups compared*

L1		no change	changes made in target					other changes	
		a	c	d	e	g	total	b	f
D	Total (533)	390	94	2	8	1	105	37	1
	% of (525) responses	74%	18%	0%	2%	0%	20%	7%	0%
G	Total (189)	83	50	7	4	8	69	36	2
	% of (179) responses	46%	28%	4%	2%	4%	39%	20%	1%
F	Total (204)	61	55	14	7	2	78	64	1
	% of (180) responses	34%	31%	8%	3%	2%	43%	36%	1%
T	Total (192)	26	44	21	13	9	87	78	1
	% of (155) responses	17%	28%	14%	8%	6%	56%	50%	1%

a no change in dummy subject
c predicted change in dummy subject
d change in dummy subject, while no change predicted
e other change in dummy subject than predicted
g part of the sentence with the dummy subject missing or changed substantially
b change not related to dummy subject
f large part of the sentence missing or changed substantially

If we look at the results at the group level, presented in Table 2 below, we see a strong influence of typological distance to Dutch. As the typological distance increases (from German to French to Turkish):
– the percentage of literal imitations decreases
– the percentage of unpredicted changes in the target (total minus c) increases
– the percentage of changes in other parts of the sentence (b plus f) increases.
However, the percentage of predicted changes in the target, which is 18 % for the native speaker control group, is around 30 % for all non-native speaker groups.

It should be noted that, if we had considered all participants with an unpredicted change in a dummy subject target to fall outside the native speaker range, there would still be six second language learners who would have fallen within the native speaker range (ID53, 55, 59, 68, 70 and 77). However, there would also have been nine native speakers of Dutch who would not have fallen within the native speaker range.

For the French and Turkish participants, the same items cause problems, especially item number 3. In this item of the DPnsi category, represented in (10) below, the target *er* is repeated by the native speakers (and most of the Germans) but replaced by *0* by ten French and eleven Turkish participants.

(10)	*Tot*	*nu*	*toe*	*komen*	*er*	*steeds*	*meer*	*mensen*	*naar*
	Until	now	come	there	still	more	people	to	this

dit popconcert.
pop concert
'So far, more and more people have come to this pop concert.'

It is not clear why this particular item causes more problems than the other items for the French and Turkish participants, and why this is not the case for the German participants. It should be noted that in the sentence preference task (see Section 3.2 below) the native speakers of German had a lower average score on this category as a whole than the native speakers of French and a similar average score as the native speakers of Turkish. For all other dummy subject items, only six non-native-like changes were made by five French participants and 21 non-native-like changes were made by ten Turkish participants.

3.2. *Sentence preference task*

To establish a native speaker norm for the analysis of the second language data, we used the results from the native speakers in our experiment. For both the group and the main individual analyses, we established a native speaker pattern for each dummy subject pair in a category based on the scores of the majority of the native speakers on most of the items in the category concerned. There were basically two native speaker patterns: either a clear preference for one dummy subject over the other (i.e., either a score of 1 or 2 or a score of 6 or 7[1]) or a pattern in which one dummy subject was considered better than or equal to the other dummy subject (i.e., either scores between 1 and 4 or scores between 4 and 7).

The SPT originally contained 124 test items. Of these 124 items, five were removed from the analyses because they contained errors and forty for various other reasons. Items were excluded when less than 90 % of the native speakers had the native speaker majority pattern or when their predicate behaved differently from the other predicates in the same category. For the change of state category of the Active-sentential type (AScos), there were so many items that had to be excluded on these grounds (13 out of 16), that we included the remaining three items in the total score, but excluded the category as a whole from the category analyses. There were thus 79 dummy subject items and 6 categories that were used in the analyses.

For each of these 79 items participants were given a score of 1 if they had the native speaker pattern and a score of 0 if they had another pattern. The sum of these scores for each category and in total was used to compute t-tests and ANOVAs followed by post hoc tests for the group analyses and z-scores for the individual analyses. For the individual results, we additionally computed z-scores based on differences in preference for certain dummy subject pairs between the categories within each type.

Table 3. *T-tests for each category in the sentence preference task*

	NSs		NNSs		Mean difference	t	df	Sig. (2-tailed)
	Mean	SD	Mean	SD				
DPg (10)	9.6	.7	8.9	1.2	.71	3.296	68.212	.002*
DPnst (6)	5.4	.9	4.9	1.4	.46	1.820	74.996	.073
DPnsi (5)	4.8	.4	4.0	1.1	.77	4.358	55.198	.000*
ASg (42)	40.8	1.6	38.3	5.0	2.54	3.184	50.097	.002*
AScos (3)	2.9	–	2.6	–	.28	–	-	–
PSg (5)	4.7	.7	3.7	1.4	.96	3.983	63.371	.000*
PSdo (8)	7.7	.6	6.8	1.2	.89	4.316	63.285	.000*
Total (79)	75.8	2.6	69.2	7.6	6.59	5.365	51.427	.000*

SD standard deviation
df degrees of freedom
sig. significance
* sig. ($p < .05$)

3.3. *Group results of the sentence preference task*

To see whether the near-natives together performed differently from the native speakers on the categories in the sentence preference task, we computed t-tests for each category. Levene's tests showed that there were unequal variances for all the categories in the sentence preference test ($p < .05$). T-tests revealed that the differences between the native speakers of Dutch and all non-native speakers taken together are highly significant for all categories except DPnst (transitive predicates with a non-specific noun phrase subject). The t-values (equal variances not assumed) and their significance are presented in Table 3 below. For each category the number of items is given in brackets.

For the differences between all groups separately we used ANOVA followed by Games Howell post hoc tests. We used Games Howell post hoc tests, because they can handle unequal variances and because Games-Howell is a rather liberal post hoc test. Differences between groups will therefore be revealed more easily. Levene's tests again indicated that there were unequal variances across groups for each category in the analysis ($p < .01$). Outcomes of the ANOVA were again significant for all categories, except DPnst. The F-value of the total sum of all items was also significant. The F-values and their significance are presented in Table 4.[2]

For the Games Howell post hoc tests we used a two-tailed test of significance for all group comparisons, because the second language learners were selected on the basis of their (apparent) native-like grammar and because we had no predictions as to which L1 group would perform best, although we expected that, if there were differences, the groups with a typologically relatively

Table 4. *ANOVA for each category in the SPT*

	L1 Dutch		L1 German		L1 French		L1 Turkish		Mean Square	F	Sig.
	Mean	SD	Mean	SD	Mean	SD	Mean	SD			
DPg (10)	9.6	.7	8.3	1.4	9.6	.6	8.8	1.1	7.749	8.974	.000*
DPnst (6)	5.4	.9	4.9	1.5	5.3	1.0	4.5	1.6	3.006	2.253	.088
DPnsi (5)	4.8	.4	3.9	1.3	4.1	1.1	3.9	.8	4.397	6.398	.001*
ASg (42)	40.8	1.6	40.0	2.4	39.5	3.4	34.9	7.0	117.400	10.353	.000*
AScos (3)	2.86	–	3.0	–	2.6	–	2.2	–	–	–	–
PSg (5)	4.7	.7	4.6	.7	3.8	1.2	2.6	1.5	15.884	16.769	.000*
PSdo (8)	7.7	.6	7.2	.7	6.9	1.2	6.2	1.5	8.040	9.425	.000*
Total (79)	75.8	2.6	71.9	3.8	71.8	5.5	63.2	9.7	543.944	22.071	.000*

* significant ($p < .05$)

Table 5. *Results of the Games-Howell posthoc tests for categories of the SPT*

	DPg		DPnsi		ASg		PSg		PSdo		Total	
	Md	Sig.	Md	Sig.	Md	Sig.	Md	Sig.	Md	Sig.	Md	Sig.
D-G	1.28	.021*	.82	.115	.82	.628	.08	.982	.51	.080^	3.89	.009*
D-F	.01	1.000	.68	.131	1.35	.468	.88	.069^	.77	.129	4.02	.062^
D-T	.84	.080^	.83	.023*	5.90	.045*	2.07	.002*	1.47	.018*	12.66	.002*
G-F	–1.27	.027*	–1.33	.990	.53	.960	.80	.155	.27	.880	.13	1.000
G-T	–.44	.802	.01	1.000	5.08	.105	1.99	.002*	.97	.173	8.78	.035*
F-T	.83	.109	.14	.980	4.54	.184	1.19	.133	.70	.538	8.65	.048*

NS Native speakers of Dutch
G L1 German
F L1 French
T L1 Turkish
Md Mean difference
* significant ($p < .05$; two-tailed)
^ only significant in a one-tailed test ($p < .10$)

closely related L1 (German and French) would be better than the group with a typologically very distant L1 (Turkish) group on the basis of previous studies (for example, McDonald 2000). For the DPnst category we did not consider the results of the post hoc test, because the F-value for this category was not significant. The results for all other categories are presented in Table 5, as well as the total over all categories. Results that are only significant in a one-tailed test are indicated by the symbol "^". It should be noted that such values only occur for differences between the native speakers versus the near-native groups. The mean total scores for the four groups are plotted in Figure 2.

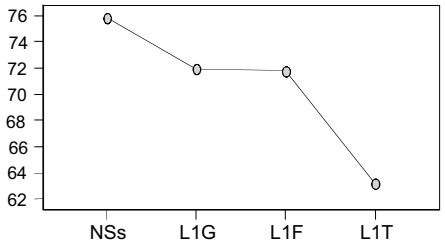

Figure 2. *Means plot for total (79 items)*

As can be seen in Table 4, the native speakers have higher scores than all three learner groups for all categories. Table 5 shows that, for the native speakers versus the L1 Turkish group, all differences are significant, except the one for DPg. For the native speakers versus the L1 French group, on the other hand, none of the differences is significant and for the native speakers versus the L1 German group, only the differences for DPg and for all items together are significant. This seems to indicate that the L1 German and L1 French group behave more like native speakers of Dutch than the L1 Turkish group. The L1 Turkish group has the lowest scores for all categories, except DPg. They seem to have most problems with the general category of the Passive-sentential (PS) type (a mean of 2.6 compared to 4.7 for the native speakers). Whereas the native speakers of German seem to have most problems with the DP-type, the native speakers of French, on the other hand, seem to have most problems with the PS-type. If all items in the analysis are considered, the L1 German and L1 French group seem to perform equally well. The results for the general category of the Active-sentential type seem to correlate best with the typological distance between Dutch and the first languages involved. In the next paragraph we will discuss what characteristics of the source languages might account for the specific problems for the three L1 groups.

3.4. *Individual results of the sentence preference task*

To determine statistically which L2 speakers fell within the native speaker range, we computed z-scores for each category in the sentence preference task (except change of state) on the basis of the pattern scores described in Section 3.2. We considered everyone with a z-score lower than -1.96 (z-scores higher than 1.96 did not occur) to fall outside the native speaker range ($\_ = .05$). This criterion has also been used by Flege, Munro and MacKay (1995) and by Bongaerts, van Summeren, Planken and Schils (1997).[3] By looking at each category separately we can see whether L2 speakers have acquired all the rules

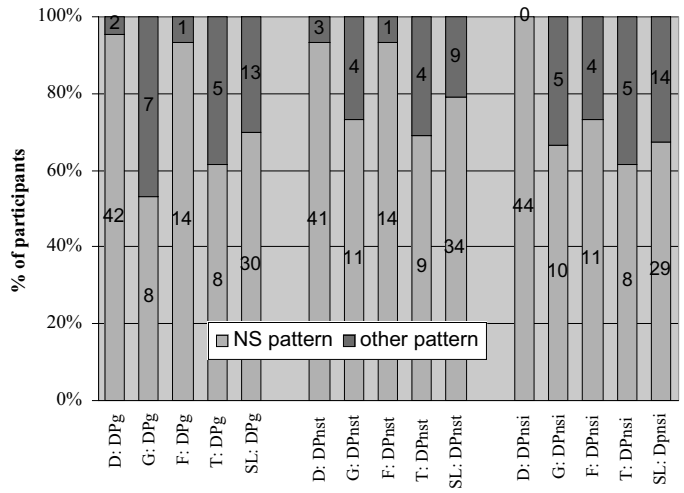

Figure 3. *Individual results for the categories of the DP-type*

for dummy subjects in our test and, if not, where exactly problems occur. It should be noted that the criterion for falling outside the native speaker range turned out to be very strict for our study. Although for each category less than 10 % (= 4) of the native speakers had a z-score lower than -1.96, there were in fact eight native speakers with such a score for at least one category. This means that it is very likely that we underrepresent the number of near-natives who behave like native speakers with respect to dummy subject constructions in Dutch. Even though such a strict criterion may not be completely fair to the second language learners, it is very suitable for falsification of the hypothesis that no late second language learner should be able to attain a native level in L2 grammar. The results are presented in Figures 3 and 4 below. The first letter under each column refers to the L1 of the participant group (Dutch, German, French, Turkish). "SL" refers to the results of all second language learners together. Z-scores lower than -1.96 are labelled "other pattern" in these figures. Figure 3 shows the results for the DP-type.

One can see in Figure 3 that, of the 43 second language speakers, 30 speakers performed within the native speaker range for DPg, 34 for DPnst and 29 for DPnsi. In the L1 French group, more participants scored within the NS range than in the L1 German and L1 Turkish group. Participants in this group who did make mistakes mainly made them in the DPnsi category, which means they prefer *0* to *er* more than native speakers of Dutch. This could be attributed to the influence of French. In French fewer predicates with a noun phrase subject can get a dummy subject (*il* or *ce*) than in Dutch. For this type (unlike

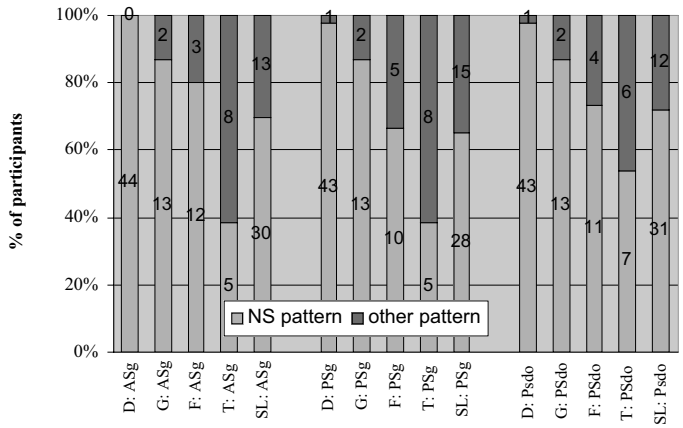

Figure 4. *Individual results for the AS- and PS-type*

for the other two types), the native speakers of German and Turkish perform about equally well, in spite of the huge differences between the two languages and their difference in typological distance to Dutch. Turkish has no dummy subjects at all, but it has a grammatical notion of specificity, which may help Turkish learners of Dutch in making a distinction between the DPnsi category and the other categories. This might explain why more native speakers of Turkish perform in a native-like way on this type than on the other types (see also Figure 4). German, on the other hand, has dummy subjects for constructions with noun phrase subjects, but only in initial position, and these are not limited to specific (or even definite) subjects. Moreover, specificity does not seem to be grammaticalised in German. This might explain why about half the native speakers of German do not have a native-like preference for *0* for the DPg category.

For the sentential types (AS and PS), the results for the different L1 groups look rather different, as can be seen in Figure 4.

Again, there are speakers in each L1 group that fall within the native speaker range for all categories. For the ASg category, 30 of the 43 second language speakers fall within the native speaker range. For the two categories of the PS-type there are as many as 28 (PSg) and 31 (PSdo) second language speakers who fall within the native speaker range.

Contrary to the results for the categories of the DP-type, only a few Germans fall outside the NS range on these types and the results for the PS-type seem to correspond well to the typological distance between the L1s and the L2. An explanation for this might be that German has dummy subjects in passive sentences, whereas French has dummy subjects, but not (normally) in passive

sentences, and Turkish does not have dummy subjects at all. In spite of this, more than 50 % of the French and Turkish participants fall within the native speaker range on the PSdo category and 67 % of the French participants fall within the native speaker range on the PSg category. Why the Turkish participants score better on the PSdo category than on the PSg category (see also the group results) is unclear. It might have to do with the fact that some Turkish participants seem to overgeneralise the preference for an overt dummy subject (*het*) to *0*. Although this explanation seems compatible with the group results, it cannot account for why the individual results for the ASg category are similar to the results for PSg rather than for PSdo. Because of the small number of participants in the L1 Turkish group, these differences are difficult to interpret.

For the ASg category, the native speakers of German and French perform very well, whereas this is one of the most difficult categories for the native speakers of Turkish. This could be due to the fact that German and French have dummy subjects in active sentences with a sentential subject (also in non-initial position), whereas Turkish does not. In French the type of predicates that can have a dummy subject is more limited than in Dutch. For German, more research is needed to see which predicates can have a dummy subject (in non-initial position) in active sentences with a sentential subject. It should be noted that, even though both German and French have dummy subjects, they do not make the same distinction as is made in Dutch for *het* and *er* and native speakers of those languages learning Dutch still have to find out which dummy subject is preferred in which construction for both the AS-type and the PS-type.

If we combine the individual results for all categories in the sentence preference task, it turns out that there are 36 native speakers and 8 second language speakers who have the native speaker pattern for each category: three from the L1 German group (ID 59, 62 and 64), four from the L1 French group (ID 68, 71, 72 and 73) and one from the L1 Turkish group (ID 83). All these participants also fall within the native speaker range on the total score over 79 items.

3.5. *Additional analyses of the sentence preference task*

To see whether the participants with a native speaker pattern for all categories also made the same distinction between the different categories of each type as the majority of the native speakers, we computed additional analyses. In these analyses we computed z-scores based on difference scores for relevant dummy subject combinations between two types of the same category.

For the DP-type we looked at the difference between DPnst and DPnsi (*er* vs *0*) to see whether all second language speakers with a native speaker pattern made a distinction between transitive and intransitive predicates, rather than having no preference for both categories. This turned out to be the case (all second language learners with a native speaker pattern for each category had a

z-score between −1.96 and 1.96). There was one native speaker, however, with a native speaker pattern for each category who had a z-score lower than −1.96.

For the change of state category of the AS-type, the data revealed that the native speakers had no clear pattern and many native speakers even had inconsistent patterns, but all native speakers had a weaker preference for *het* to *0* for this category than for the general category. We therefore computed z-scores based on the difference between the preference for *het* versus *0* for AScos and ASg. Although all native speakers had a stronger preference for *het* for the ASg category, this difference was so small for two native speakers that it resulted in a z-score lower than −1.96 and so large for two other native speakers that it resulted in a z-score higher than 1.96. The Turkish participant who had a native speaker pattern for each category also made a greater distinction than most (43) of the native speakers, resulting in a z-score of 2.2. Given the huge amount of variation between native speakers, between predicates and between items on the change of state category, it is difficult to tell what such deviations really mean.

Finally, for the PS-type we looked at the difference between the preference for *het* versus *0* for PSg and PSdo, because the native speakers had an opposite preference for these dummy subject for PSg versus PSdo, both including no preference for either dummy subject. It turned out that there was one native speaker who had the typical native speaker pattern, but who made a smaller difference than the majority of native speakers, resulting in a z-score of −2.0. One of the French participants with a native speaker pattern on all categories in the sentence preference task (ID72) had no preference for either dummy subject for both PSg and PSdo (z = -2.5).

If we combined the results for all native speakers on these three difference scores, there would be eight native speakers with at least one deviant z-score. Of these eight native speakers, there were only three who also had a deviant pattern for one of the categories of the sentence preference task. There are, therefore, a total of thirteen native speakers out of 44 (=30 %) who in some way deviate in their behaviour from the majority of the native speakers. Considering these results for these additional analyses, we will consider all L2 participants with a z-score between −1.96 and 1.96 for each category of the sentence preference task to fall within the native speaker range.

3.6. *Background characteristics*

The results discussed in the previous paragraph show that there is little difference between the native speakers of French and the native speakers of German on the sentence preference task. Both these groups perform quite well on this task. The native speakers of Turkish perform worse than the other two groups, but even in this group there is at least one participant who behaves like the

majority of native speakers on all dummy subject constructions in our test. For separate categories, between 38 % and 69 % of the second language learners in our study fall within the native speaker range.

On the imitation task, on the other hand, participants performed worse and the results corresponded much more strongly to the typological distance between the L1 and Dutch. These differences between the results for the sentence preference task and those for the imitation task can be explained as a breakdown that occurs when processing load exceeds participants' working memory capacity. The problems L2 learners have in decoding the surface form of the target language are known to be larger when the difference between the L1 and the L2 is greater. They also play a much bigger role for auditory input than for written input (see, e.g., McDonald 2000: 416–417). These decoding problems are, therefore, expected to cause more processing load for participants with a more distant L1 in a test with oral stimuli. Our results suggest, therefore, that learners with a less related L1 are able to acquire dummy subject constructions in Dutch, but that they have more processing problems, which often leads to a breakdown in the imitation task (though not for all learners). Sabourin (2003) found similar results in an ERP study on processing of second language grammar by English, Romance and German learners of Dutch. On structures for which both groups performed natively on an offline grammaticality judgement test, the German group had much more native-like ERP-patterns than the Romance group (Sabourin 2003: 152).

Let us now turn to the information from the questionnaire on participants who fell within the native speaker range on all categories of the sentence preference task. First, it is clear that the possibility of reaching a native-like command of dummy subject constructions in Dutch is not limited to learners with a typologically related L1, or to learners with an L1 that is very similar to Dutch with respect to dummy subjects. In Table 6 we compare some other background characteristics between those participants that fall within the native speaker range (within NS range) and those that do not (outside NS range). For input, we computed a new variable by multiplying the participants' number of years of residence in the Netherlands by the proportion of Dutch use in comparison to use of their L1 (self-report on a five-point scale ranging from 90 % L1 and 10 % Dutch to 10 % L1 and 90 % Dutch). This new variable purports to be a very rough estimation of the amount of input the second language learners received. For self-reported language skills we report the results for the skill with the largest absolute difference between the two groups for each language. For the L1 the differences were so small for all skills that we did not include them in the table.

To see whether the differences reported in Table 6 were significant or not, we computed t-tests in which we compared the within NS range group to the outside NS range group. Age at the time of testing is very similar for both

Table 6. *Comparison of some background characteristics*

Variables	Within NS range	Others	Total L2 speakers
Sex (m/f)	1/7	19/16	20/23
Age	45 (33–55)	44 (23–73)	44 (23–73)
Age of arrival (AoA)	23 (12–35)	21 (12–35)	22 (12–35)
Length of residence (LoR)	20 (10–42)	22 (4–50)	21 (4–50)
Proportion use L1-Dutch[a]	70%/30%–10%/90%	770%/30%–10%/90%	770%/30%–10%/90%
Input	14 (3–37)	13 (2–39)	13 (2–39)
Dutch lessons (yes/no)	6/2	25/10	31/12
Duration Dutch lessons[a]	2 weeks–3 years	3 lessons–10 years	3 lessons–10 years
Level of education[a]	VWO-PhD	MBO-PhD	MBO-PhD
Speaking Dutch	4.5 (4–5)	4.1 (3–5)	4.2 (3–5)
Listening Dutch	4.6 (4–5)	4.3 (3–5)	4.4 (3–5)
Writing Dutch	4.0 (2–5)	3.4 (2–5)	3.5 (2–5)
Reading Dutch	4.8 (4–5)	4.6 (3–5)	4.6 (3–5)
Grammar Dutch	3.8 (2–5)	3.6 (1–5)	3.6 (1–5)
Speaking other language[b]	3.6 (3–5)	2.9 (1–5)	3.1 (1–5)
Listening other language[b]	4.4 (3–5)	3.3 (1–5)	3.5 (1–5)
Writing other language[b]	3.6 (3–5)	2.9 (1–5)	3.1 (1–5)
Reading other language[b]	4.4 (4–5)	3.5 (1–5)	3.7 (1–5)
Grammar other language[b]	3.5 (2–5)	3.2 (1–5)	3.2 (1–5)

Input Length of residence (in years) * the proportion of Dutch spoken in relation to the L1
a For these variables only the range is presented
VWO secondary education preparing for university
MBO intermediate professional education
b this variable is presented for the participants' best other language and only for participants who had answered this question (n=8 for within NS range; n=32 for other L2 speakers and n=40 for total L2 speakers)

groups. Interestingly, the participants who fall within the native speaker range have a slightly higher mean age of arrival (AoA), although this difference is not significant and the ranges for both groups are the same. They also have hardly had more input and the difference in input and LoR is not significant. It should be noted, though, that the real amount of input is hard to determine, because it is hard to estimate as it can change over the years, and because the variable is based in part on a relatively rough measure (a five-point scale for proportion of Dutch used relative to the L1). There may also have been qualitative differences in the input. However, even under favourable conditions, three years of "full input" (30 % use of Dutch during nine years LoR (for one of

the French participants)) seems to be a rather small amount to reach native-like attainment in such a difficult area of grammar.

The self-reported skills in Dutch are on average higher for the group that falls within the native speaker range. These differences are significant for speaking and writing ($p < .05$ (one-tailed)), but surprisingly not for self-reported proficiency in grammar. The self-reported skills in the participants' best other language are also on average higher for the group that falls within the native speaker range. These differences are significant for listening and reading ($p < .05$ (one-tailed)), but again not for self-reported proficiency in grammar in these languages. This latter result is difficult to interpret, because we do not have enough information about differences in input for the two groups for these languages. We do know, however, that most L2 participants learned these languages in school and did not receive substantial amounts of input in these languages. Although this information is too limited to draw any conclusions from, it seems to suggest, especially in combination with the results on input, that a large amount of input is insufficient for late learners to reach a native level in L2 grammar, and that something like language aptitude plays a role as well.

None of the t-tests for skills in the native language were significant. Although self-report on a five-point scale is of course a very rough measure of L1 proficiency, these results seem to suggest that one can reach a very high level in L2 grammar without it effecting the normal use of the L1.

If we look at variables from the questionnaire that are more difficult to quantify, perhaps the most striking result is that all participants that fell within the native speaker range had either studied a language or worked in an environment where language plays an important role. Of these eight participants, three were translators, one was a teacher of her L1 (French), one was a linguist, one had worked at Maison Descartes (an institute in Amsterdam with the literary and academic mission to promote French culture and stimulate French-Dutch dialogue), one had worked in a German department at a Dutch university and one had studied French. This suggests that something like metalinguistic awareness might play an important role. It should be noted, though, that in the other group there were also 23 participants who had some kind of special background with respect to language and two of them were translators. The eight participants within the native speaker range were also very highly educated. Seven of them had an academic degree and the other one had had pre-university education. Even though most participants in our study had a high level of education (only five participants had a level below higher professional education), the relation between level of education and NS range is significant ($t = -2.95$ (equal variances not assumed); $p = .01$ (two-tailed)). Whereas linguistic background seems to play a role, the data with respect to attending Dutch lessons seem to indicate that, for acquiring dummy subject constructions, it does not matter

much whether participants receive specific instruction in Dutch or not. More empirical evidence is necessary to shed light on the role of education and metalinguistic awareness in reaching a native level in L2 grammar. It is conceivable that metalinguistic awareness makes people more focussed on form and on phonologically less salient grammatical elements such as dummy subjects, which would enable them to process information that normally goes unheeded in late L2 acquisition.

There were many more women than men in the within NS range group (1 man and 7 women), whereas there were more men in the outside NS range group (19 men and 16 women). This result, striking though it may be, is difficult to interpret, however, because the proportion of men and women was very different for the different L1 groups: in the L1 German group there were seven men and eight women, in the L1 French group there were three men and twelve women, and in the L1 Turkish group there were ten men and three women. The fact that women perform better than men may, therefore, be an artifact of the fact that the L1 French group is better than the L1 Turkish group.

Finally, we will look at the situations in which participants use Dutch and their native language. Of the eight participants within the NS range, there are five who speak Dutch at home, there is one who speaks Dutch everywhere except at home and two who speak Dutch at work and with Dutch friends (one of them also speaks Dutch with relatives). The use of the native language in this group ranges from only with relatives (outside home) to at work, at home and with relatives. In the group that falls outside the native speaker range, the use of Dutch ranges from use with relatives (outside home) and friends to "everywhere and with everyone in the Netherlands". The use of the native language in this group varies from "only when there is no other solution" to with relatives (outside home) and friends and either at home or at work. There are no people in either group who speak their native language both at work and at home, except when they talk about emotional issues (at home).

4. Conclusion

According to the Critical Period Hypothesis (CPH), levels attained after the presumed terminus of the critical period (around age 12 for L2 syntax) should always be low and significantly lower than levels attained by learners who started to acquire a second language before the alleged terminus (see, e.g., Bialystok and Miller 1999). Native-like attainment by late learners for constructions that are difficult to acquire, that are different in the L1 and for which no explicitly formulated rules are available, therefore, constitutes counterevidence against the CPH for syntax in SLA. In the study presented in this paper, we provided such evidence by showing that there are late second language learners who fall within the native speaker range in their command of dummy

subject constructions in Dutch, even if we use very strict criteria for falling within the native speaker range. In order to dismiss this evidence and prove that there is a critical period for syntax in SLA, an account would be necessary of how late learners with a strongly reduced sensitivity to the input due to having passed the critical period could possibly acquire such grammatical constructions. So far, such an account has not been put forward.

With respect to the role of the L1, our results clearly show that it is possible to reach a native level in L2 syntax for learners with language backgrounds with different degrees of typological distance to the L2, ranging from a very close typological relationship (German and Dutch) to a very distant typological relationship (Turkish and Dutch). Since Turkish does not have any dummy arguments, our results also show that it is possible for late learners to acquire a grammatical construction not present in their native language.

Although we question the existence of a critical period for syntax in SLA, we acknowledge that biological and other factors usually confounded with age play an important role in second language acquisition. An important biological factor usually confounded with age is entrenchment of L1 structures. Because of this entrenchment, adult learners may be less attentive to certain aspects of the L2 input. An important non-biological factor confounded with age is the nature of the input. Differences in prosodic features, stress, utterance length, etc. of the input directed to young children versus the input directed to adults may result in differences in attention to certain characteristics of the input. However, we do believe that the empirical evidence available shows that the general decline in the ability to acquire L2 syntax is not absolute and deterministic, as is the case for orientation specificity in the visual cortex in cats or imprinting in ducklings, for example (see, e.g., Baxter 1966; Ratner and Hoffman 1974). It has been shown by Pallier et al. (2003), for example, that, when L2 learners are completely deprived of input in the L1 (in this case Korean children adopted by French speaking parents), the levels attained in the L2 are much more native-like than is normally the case in L2 acquisition.

It could be that other biologically determined factors related to age constrain the acquisition of L2 syntax, for example maturation of working memory. It seems plausible that elements such as dummy subjects, which are phonologically non-salient and do not really contribute to the meaning of the sentence, are overlooked by late learners, who can process large units at once because of their larger working memory capacity. Although this is a biologically determined age factor, it is not specific to language and not absolute. Details in the form that in general are not attended to by late learners might be acquired if special attention is paid to them. It seems plausible that second language learners with a linguistic background or high language aptitude are capable of paying special attention to elements like dummy subjects, which are ignored by most L2 learners. The results with respect to the background characteristics of

the learners who score within the NS range seem compatible with this hypothesis. Unfortunately, however, our study cannot test this prediction, because we did not investigate the relation between language aptitude and sensitivity to the input. Education may also play a role. Informal contacts with the participants in this study led to the impression that there were differences between participants with respect to their attitude towards prescriptive language standards in general, due to differences in education. It would be interesting to investigate in future studies whether differences in language aptitude and education can lead to differences in attention to the exact form of the input.

Rather than providing more evidence for the (non-)existence of a critical period, future research should be aimed at trying to find out what factor(s) contribute(s) most to language learning success and what are the best ways to influence learners' sensitivity to non-salient aspects of the input.

Radboud University Nijmegen
<S.vanBoxtel@let.ru.nl>

Acknowledgements

We would like to thank Roeland van Hout and Frans van der Slik (Radboud University Nijmegen, Department of Linguistics) for their help with and discussion on the statistic analyses we used. We would also like to thank the native speakers of and experts on German, French and Turkish who provided us with insights about these languages and the audiences of "Attaining the ultimate", the BUCLD, SLRF, EUROSLA, DGfS, LIN, NOTT and the NBG for their useful questions and comments on some of the results and conclusions presented here.

Notes

1. There was one exception to this pattern: for the category of the Passive Sentential type with a dummy object in the active equivalent (PSdo) the NS pattern for *het* vs. *er* was 1–3.
2. Because of the unequal variances across groups we also computed Welch and Brown-Forsythe. For these analyses, too, all categories except DPnst were significant at the .05-level.
3. These researchers use $2 < z < -2$ for reasons of convenience.

References

Baxter, Bruce L. (1966). Effect of visual deprivation during postnatal maturation of the electroencephalogram of the cat. *Experimental Neurology* 14: 224–237.

Bialystok, Ellen and Barry Miller (1999). The problem of age in second language acquisition: Influences from language, structure and task. *Bilingualism: Language and Cognition* 2: 127–145.

Birdsong, David (1992). Ultimate attainment in second language acquisition. *Language* 68: 706–755.

— (1997). Intransitivity and SE in French: Aspects of late learnability. Paper presented at the Boston University Conference on Language Development, November 1997.

— (2003). Authenticité de prononciation en français L2 chez des apprenants tardifs Anglophones: Analyses segmentales et globales. *AILE* 18: 17–36.

Bley-Vroman, Robert and Craig Chaudron (1994). Elicited imitation as a measure of second-language competence. In *Research Methodology in Second-Language Acquisition*, Elaine Tarone, Susan Gass and Andrew Cohen (eds.), 245–261. Hillsdale, NJ: Lawrence Erlbaum.

Bongaerts, Theo (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In *Second Language Acquisition and the Critical Period Hypothesis,* David Birdsong (ed.), 133–159. Mahwah, NJ: Lawrence Erlbaum.

Bongaerts, Theo, Susan Mennen and Frans van der Slik (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica* 54: 298–308.

Bongaerts, Theo, Chantal van Summeren, Brigitte Planken and Erik Schils (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19: 447–465.

Bornstein, Marc (1989). Sensitive periods in development: Structural characteristics and causal interpretations. *Psychological Bulletin* 105: 179–197.

Colombo, John (1982). The critical period concept: Research, methodology, and theoretical issues. *Psychological Bulletin* 91: 260–275.

Elman, Jeffrey (1993). Learning and development in neural networks: The importance of starting small. *Cognition* 48: 71–99.

Flege, James E., Murray J. Munro and Ian R.A. MacKay (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97: 3125–3134.

Hyltenstam, Kenneth (1992). Non-native features of near-native speakers. On the ultimate attainment of childhood L2 learners. In *Cognitive Processing in Bilinguals,* Richard J. Harris (ed.), 351–368. Amsterdam: Elsevier Science Publishers.

Hyltenstam, Kenneth and Niclas Abrahamsson (2003). Age of onset and ultimate attainment in near-native speakers of Swedish. In *Multilingualism in Global and Local Perspectives. Papers from the 8th Nordic Conference on Bilingualism*, November 1–3, 2001, Stockholm Rinkeby, Kari Fraurud and Kenneth Hyltenstam (eds.), 319–340. Stockholm: Centre for Research on Bilingualism, Stockholm University, and Rinkeby Institute of Multilingual Research.

Just, Marcel A. and Carpenter, Patricia A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99: 122–149.

Lenneberg, Eric (1967). *Biological Foundations of Language*. New York: John Wiley.

McDonald, Janet L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics* 21: 395–423.

Moyer, Alene (1999). Ultimate attainment in L2 phonology. The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition* 21: 81–108.

Newport, Elissa (1990). Maturational constraints on language learning. *Cognitive Science* 14: 11–28.

— (1991). Contrasting conceptions of the critical period for language. In *Epigenesis of Mind: Essays of Biology and Cognition,* Susan Carey and Rochel Gelman (eds.), 111–130. Hillsdale, NJ: Lawrence Erlbaum.

Pallier, Christophe, Stanislas Dehaene, Jean-Baptiste Poline, Denis LeBihan, Anne-Marie Argentini, Emmanuel Dupoux and Jacques Mehler (2003). Brain imaging of language plasticity in adopted adults: Can a second language replace the first? *Cerebral Cortex* 13: 155–161.

Penfield, Wilder and Lomar Roberts (1959). *Speech and Brain Mechanisms.* Princeton, NJ: Princeton University Press.

Pitts Cochran, Barbara, Janet L. McDonald and Susan J. Parault (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language* 41: 30–58.

Ratner, Alan, and Howard Hoffman (1974). Evidence for a critical period in khaki campbell ducklings (anas platyrhynchos domesticus). *Animal Behavior* 22: 249–255.

Sabourin, Laura L. (2003). *Grammatical Gender and Second Language Processing: An ERP study.* [Groningen Dissertations in Linguistics 42]. Groningen: University of Groningen.

Van Boxtel, Sonja (2005). *Can the late bird catch the worm? Ultimate attainment in L2 syntax.* (LOT Dissertations 109.) Doctoral dissertation, Radboud University Nijmegen. Utrecht: LOT.

White, Lydia and Fred Genesee (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research* 12: 233–265.